A brief exploration of S-curves: Forecasting global energy transition pathways

Benjamin Pim Wagenvoort

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hillary 2026

Supervised by: Dr. François Lafond, Prof. J. Doyne Farmer, and Dr. Matthew Ives

Abstract

A fast transition from fossil fuels to renewables is important to minimise further global warming, yet energy forecasts have systematically underestimated the technologies most critical to decarbonisation. This thesis asks: *How can S-curves be used to better understand and forecast technological change?* It comprises three papers that identify biases in standard S-curve estimation, develop a Bayesian forecasting framework grounded in empirical diffusion regularities, and apply the framework to assess the probability of energy transition scenarios in the IPCC AR6 report.

The first paper characterises the downward bias in S-curve asymptote estimates under standard fitting techniques. The bias can persist even when 30% of the curve is observed, and estimates exhibit high variance. We develop a parametric bootstrap debiasing method that leverages approximately unbiased growth and noise parameter estimates to correct the bias. Applied to electric vehicle and solar PV adoption data, the debiased estimates indicate continued sustained growth trajectories.

The second paper compiles 120 historical growth trajectories spanning centuries of technology adoption and shows that diffusion is systematically asymmetric. The Bertalanffy–Richards curve with $\beta \approx 2/3$ provides superior out-of-sample fit, explaining nearly 95% of variance across technologies. These cross-technological regularities enable a technology-agnostic Bayesian forecasting framework whose distributional forecasts are well-calibrated. Applied to renewables, the framework forecasts that

global solar PV will reach a median of roughly 80 PWh per year by 2050 (90% prediction interval 13–470 PWh), comparable to all the useful energy the world consumes today, surpassing nearly all IPCC AR6 scenarios and the International Energy Agency (IEA) net-zero by 2050 scenario. The wind forecast (median $\sim$ 5 PWh per year, 90% prediction interval 4–17 PWh) falls below many AR6 targets.

The third paper exploits a complementarity between empirically grounded forecasts and integrated assessment model (IAM) scenarios: scenarios explore what policy *could* achieve, while S-curve forecasts assess what historical diffusion trends *suggest*. The AR6 ensemble over-represents scenarios with low solar growth relative to our forecasts. Under trajectories consistent with observed diffusion patterns, warming at net-zero is likely limited to below 2°C, with rapid electrification and a significant decline in fossil fuel share by 2050. The median net-zero year is 2068, with a 10% probability of reaching net-zero before 2050.

Taken together, the thesis demonstrates that technologies exhibit some remarkable regularities, and one can use S-curves to forecast them. Combining such forecasts with appropriate statistical methods provides a powerful empirical benchmark against which scenarios can be transparently evaluated.

Declaration

This thesis has not been submitted, either partially or in full, for another qualification at this University, or for a qualification at any other institution.

The work presented in this thesis is entirely my own, except where otherwise stated in figure captions or listed below.

Paper contributions:

1. **Chapter 3 (Paper 1):** *Why S-curves underestimate the future: Systematic bias in diffusion forecasting.* This is a single-author paper. All research, analysis, and writing were performed by myself. For helpful comments, insightful feedback, guidance on statistical methods, and stimulating discussions I thank François Lafond. I thank J. Doyne Farmer for useful discussions and feedback on earlier drafts.

2. **Chapter 4 (Paper 2):** *Universality and predictability of technology diffusion. Conceptualization:* B.W., F.L., J.D.F. *Methodology:* B.W., F.L., J.D., J.D.F. *Data Curation:* B.W. *Investigation:* B.W. *Software:* B.W., J.D. *Visualization:* B.W. *Validation:* B.W. *Writing – original draft:* B.W., F.L., J.D.F. *Writing – reviewing and editing:* B.W., F.L., J.D., J.D.F. *Supervision:* F.L., J.D.F.

3. **Chapter 5 (Paper 3):** *Empirical solar and wind forecasts expose a solar-growth shortfall in AR6 mitigation scenarios. Conceptualization:* B.W. *Method-*

ology: B.W. *Data Curation:* B.W. *Investigation:* B.W., M.I. *Software:* B.W. *Visualization:* B.W. *Validation:* B.W. *Writing – original draft:* B.W., M.I. *Writing – reviewing and editing:* B.W., M.I. *Supervision:* M.I.

Where: B.W. = Benjamin Wagenvoort, F.L. = François Lafond, J.D.F. = J. Doyne Farmer, J.D. = Joel Dyer, M.I. = Matthew Ives.

Data and code:

1. All data used in this thesis are from publicly available sources, except data sourced from ReThinkX. Specifics are documented in each chapter's data description sections and supplementary materials.

2. All code was written by myself unless otherwise stated.

Use of generative AI:

During the preparation of this work the author used generative AI to assist with structuring and editing sentences. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the thesis.

Acknowledgements

I write this at the end of my DPhil, but I want to express thanks to those who have supported me from the start.

First, I want to thank my initial supervisor, Dr. Peter Barbrook Johnson, for taking me on as a student and giving me the freedom to explore my interests. I want to express deep gratitude to my current supervisors, Dr. François Lafond, Prof. J. Doyne Farmer, and Dr. Matthew Ives, for their invaluable guidance, support, and mentorship throughout this research journey. While we had some disagreements, and sometimes heated debates, I am grateful for the intellectual rigor and critical feedback that you provided, which helped shape this thesis into its final form, and ultimately develop me into the researcher I am today.

I want to thank the Institute for New Economic Thinking (INET), for providing a stimulating academic environment and access to resources that were crucial for my research. I want to thank Augustinus Fonden, Knud Højgaard Fonden, and Cultuurfonds, for their financial support, which made it possible for me to pursue this research. I want to thank the University of Oxford for providing a stimulating academic environment and access to resources that were crucial for my research.

I am grateful for Dorothy Nicholas, the effective captain of the ship, that is the chaotic, complexity group. The countless Thursday cinnamon sugar donuts are part of the reason I delayed my thesis hand in just a little longer.

I want to thank my teachers and friends at my high school, Herlufsholm Skole, where I developed a love for learning and a fascination with the world. Here, I learned that one should be curious, ask questions, and work hard, but also always remember to have fun along the way.

I also want to thank my undergraduate tutor, Dr. Anco Lankrijer, for his support and encouragement during my undergraduate studies, which laid the foundation for my interest in research and ultimately led me to pursue a DPhil. I want to thank my friends during my undergraduate studies, Joella and Nick, for their friendship and support during those formative years.

During the DPhil I had my ups and downs, my productive, and certainly unproductive days. I want to thank all my friends at INET. In particular, I express my deepest gratitude to Dr. Sam Wiese, Dr. Valentina Seminova, Dr. Julian Winkler, Dr. Aymeric Vie, Dr. Emilien Ravigne, Dr. Fernanda Senra de Moura, and Karolina Bassa for keeping me sane, helping me through the tough times, listening to my countless talks, rants and thoughts on S-curves, and for being my friends. I want to give a special thanks to Rohit Sahasrabuddhe and Adam Ferris for spending some late nights in the office with me, ordering food into the office, and for being great friends, on whom I relied heavily on to keep me going during the last few months of writing.

I want to thank my housemates, Laurence Peinturier and Samuel Howard, for being great company and making it enjoyable to come home. Although sometimes we managed to not see each other for days, pancake evenings and codenames nights were a great way to catch up and have fun. I want to thank my first friends in Oxford, John Kainer, Arunima Cheruvathoor and Tobias Hurst. You made me want

to stay in Oxford longer, made me feel at home, and like I belonged. My interest in research was highlighted by measuring Jowett walk with Arunima using a ruler, and more importantly a bottle of pink rosé.

People always dedicate their accomplishments to someone. I will do the same; whether it is an accomplishment remains to be seen, but for now I shall treat it as such. I want to dedicate this thesis to my parents, Bart Wagenvoort and Erna de Greeff, my four brothers, their wives, partners, and my niece, Jochem Wagenvoort, Laura Wagenvoort, Anna Wagenvoort, Jonathan Wagenvoort, Esther Hulzebosch, Ruben Wagenvoort, and Daniel Wagenvoort. While others keep the acknowledgements short, I think an extra sentence or two in a near 400 page thesis will not hurt. Bart Wagenvoort, Papa, thank you for your love, support, and for always believing in me. Thank you for teaching me to be curious, and to remember to smile; Wie lacht die wint. Laugh and smile through the hard times, because you are loved. Erna de Greeff, Mama, thank you for your love, thank you for being my biggest fan. Thank you to all my four brothers, I look up to all of you, literally and figuratively. Jochem, thank you for being a role model. You are the reason I try to work so hard, and the reason I am so ambitious. Jonathan, thank you for being my best friend, and always coming to visit when I needed it. Thank you Ruben for being the brother that ensured fun was had. Lastly, thank you Daniel, without you and the maths book you gave me 'Zo cool is Wiskunde', I would never have been able to do the maths in this thesis. Laura and Esther, thank you for being the sisters that I never had, and for always being there for me. Anna, maybe one day you will read this, and I can convince you to do research, to the potential chagrin of your parents.

Someone left out of the list above is someone who needs a special mention, Dr.

Felice Wallner. We met on our first day in Oxford, and are now leaving together on our last day in Oxford. You have been a cornerstone of my DPhil experience, and I am grateful for your friendship, support, your weirdness, and your endless room for me in your life. I am grateful for all the moments we shared, and I am excited for all the moments we will share in the future.

That leaves me with a final comment. I want to thank you, the reader, for taking the time to read this thesis. I hope that you find value and possibly even inspiration in this work.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
ABM	Agent-based model
AI	Artificial intelligence
AIC	Akaike information criterion
AIM	Asia-Pacific Integrated Model
AR5	Fifth Assessment Report (IPCC)
AR6	Sixth Assessment Report (IPCC)
AR7	Seventh Assessment Report (IPCC)
B-R	Bertalanffy–Richards (curve / process)
BECCS	Bioenergy with carbon capture and storage
BIC	Bayesian information criterion
BMA	Bayesian model averaging
CCS	Carbon capture and storage
CDF	Cumulative distribution function
CDR	Carbon dioxide removal
CF	Capacity factor
CGE	Computable general equilibrium
CHAT	Cross-country Historical Adoption of Technology dataset
CI	Confidence interval
CO_2	Carbon dioxide
COP28	28th Conference of the Parties
DGP	Data-generating process
EJ	Exajoule
ETL	Endogenous technological learning
EV	Electric vehicle
FE	Final energy
FP	Fisher–Pry (transformation)
FTT	Future Technology Transformations
GCAM	Global Change Analysis Model
GDP	Gross domestic product
GFP	Generalised Fisher–Pry (transformation)
GHG	Greenhouse gas
$GtCO_2$	Gigatonnes of carbon dioxide
GW	Gigawatt
HMC	Hamiltonian Monte Carlo
IAM	Integrated assessment model

Abbreviation	Definition
ICT	Information and communications technology
IEA	International Energy Agency
IMAGE	Integrated Model to Assess the Global Environment
IMP	Illustrative Mitigation Pathway
IPCC	Intergovernmental Panel on Climate Change
IQR	Interquartile range
IRENA	International Renewable Energy Agency
KS	Kolmogorov–Smirnov (statistic / test)
MA	Moving average
MAD	Mean absolute deviation
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MCMC	Markov chain Monte Carlo
MedALE	Median absolute log-error
MESSAGE	Model for Energy Supply Strategy Alternatives (MESSAGEix)
MLE	Maximum likelihood estimation
MW	Megawatt
MWh	Megawatt-hour
NLLS	Non-linear least squares
NZE	Net-Zero Emissions (IEA scenario)
OLS	Ordinary least squares
OWID	Our World in Data
PI	Prediction interval
PIT	Probability integral transform
POLES	Prospective Outlook on Long-term Energy Systems
PV	Photovoltaic
PWh	Petawatt-hour
REMIND	REgional Model of INvestments and Development
SE	Secondary energy
SI	Supplementary information
SSP	Shared Socioeconomic Pathway
T&D	Transmission and distribution (electricity grid)
TIAM	TIMES Integrated Assessment Model
TIMES	The Integrated MARKAL–EFOM System
TW	Terawatt
TWh	Terawatt-hour
VRE	Variable renewable energy
WGIII	Working Group III (IPCC)
WITCH	World Induced Technical Change Hybrid model

Chapter 1

Introduction

1.1 Background

Diffusion patterns in society Ideas, practices, and innovations diffuse through societies in characteristic patterns. Across diverse domains, from disease spread in epidemiology [1, 2] and biological growth to tumor growth [3], product adoption in marketing [4], innovation diffusion in sociology [5, 6], technology adoption [7, 8], and scientific literature growth [9, 10], processes often follow sigmoid-shaped trajectories known as Sigmoidal curves or *S-curves*.

The mathematical description of S-curves dates to the 19th century, originally developed for phenomena unrelated to technology. Gompertz [11] introduced the Gompertz function in 1825 to model human mortality, while Verhulst [12] developed the Logistic function in 1838 to describe population growth. Later extensions, including the Richards function [13] and the Bass diffusion model [4], increased flexibility to capture varying adoption rates and external influences [14].

S-curves in technological change Technology diffusion represents a particularly important application of the S-curve framework. In early diffusion, adoption is limited but growth is rapid. Adoption increases as the technology becomes cheaper,

better understood, and more widely available. Eventually growth tapers off, and the technology reaches saturation. This happens as the remaining potential adopters shrink, due to technology competition and constraints, and other factors.

Throughout history, innovations from monasteries and railroads to mobile phones and the internet have all exhibited characteristic S-shaped adoption patterns [7, 15, 16]. These trajectories reflect the non-linear dynamics of technological change [15], including feedback loops, emergent properties, and phase transitions [17].

Estimation and forecasting challenges. Despite excellent fits to historical data, S-curve forecasts have frequently failed to predict future diffusion accurately. Traditional fitting methods yield highly variable forecasts and exhibit systematic biases, especially when fitted to early-stage data [18, 19]. Different models (logistic vs. Gompertz) can produce vastly different predictions, with little consensus on model selection [20]. Historical forecasts have often been overly pessimistic, systematically underestimating adoption speeds [21–24], while forecasters have frequently relied on expert judgment despite evidence that statistical models outperform expert elicitation [23, 25].

Renewable energy transitions and the need for better forecasting These methodological challenges are particularly consequential for energy policy. The Paris Agreement requires rapid deployment of renewable electricity to limit warming to 1.5-2°C [26, 27]. Yet S-curve forecasts for renewable energy have been notably inaccurate, frequently underestimating renewable adoption rates [28] and overestimating cost [29]. Inaccurate forecasts lead to insufficient policy support, inadequate infrastructure planning, and missed opportunities [24, 30]. Baumgartner et al. [31] also highlight

the need for improved robustness analysis in the modeling of S-curve forecasts.

Probabilistic scenario assessment Using forecasts of individual technologies, one can offer a probabilistic interpretation of energy transition scenarios. The IPCC uses integrated assessment models to generate energy system scenarios [27]. Critics and the IPCC authors themselves [27] argue these scenarios lack probabilistic assessments of likelihood [32–34], representing "projections" or possible futures under a set of assumptions, rather than being grounded in empirical evidence. By grounding forecasts in historical patterns and quantifying uncertainty probabilistically, we can assess which transition scenarios are consistent with empirical evidence.

Thesis research question Improving S-curve forecasting methods is critical for informing policy and for strengthening understanding of technological change more broadly. This thesis asks: *How can S-curves be used to better understand and forecast technological change?* By investigating S-curve diffusion patterns in the context of technological change and energy transitions, the thesis speaks to long-running debates on socio-technical transitions, including how futures are anticipated and governed through models and scenarios. This thesis develops and tests probabilistic, empirically grounded S-curve forecasting methods and uses them to evaluate the likelihood of alternate transition pathways. In doing so, it strengthens assessments of transition dynamics and provides a more robust basis for decision-making when using scenarios to inform policy.

1.2 Research Context

This thesis addresses the overarching research question through three interconnected research questions, each addressed in a separate research article.

1.2.1 Research question 1 – Why have S-curves failed to accurately predict diffusion patterns?

Motivation. Traditional S-curve forecasting relies on fitting models to historical data using non-linear least squares (NLLS) regression. However, this approach yields highly variable and unreliable forecasts, particularly for early-stage diffusion data [18, 31]. Many renewable energy forecasts have been systematically pessimistic, underestimating actual deployment [28, 29]. Understanding why these forecasts fail is critical for improving technology adoption predictions and informing long-term planning.

Main contributions. Chapter 3 investigates the causes of S-curve forecast and parameter estimation bias. We demonstrate how the proportion of the diffusion process observed, the number of data points, and noise levels all affect estimation bias and forecast accuracy. We quantify this bias across simulated diffusion trajectories, showing it can persist even when more than 25% of the curve is observed. We present a debiasing method that, under certain conditions, reduces median parameter estimate bias, improving forecast accuracy and reliability. Applying debiased methods to electric vehicle and solar PV adoption data indicates these technologies are on continued sustained growth trajectories to 2030.

1.2.2 Research question 2 – How universal and predictable is technological growth?

Motivation. Building on the first question, we intend to improve S-curve forecasting by identifying commonalities across technologies. If each technology follows a unique growth pattern dependent on idiosyncratic factors, such as firm competence, policy support, and geopolitics, then forecasting becomes inherently difficult and requires careful consideration for each individual technology of interest. However, technologies may share common diffusion dynamics driven by underlying social, economic, and network effects [5], which could lead to common patterns or shapes across technologies [35–37]. Identifying such commonalities could provide a better understanding of technological change and enable more reliable, technology-agnostic forecasting approaches.

Main contributions. Chapter 4 presents novel insights into common patterns of technological growth. We compile an extensive database of 120 historical S-curve growth trajectories, ranging from 12th century monasteries to 21st century mobile phones. We analyze commonalities across diverse technologies spanning time, geographies, and sectors, and uncover that technology adoption exhibits universal characteristics. These cross-technological regularities allow for a technology-agnostic modeling approach and cross-sectional validation. We develop a Bayesian inference framework that generates well-calibrated distributional forecasts. Applying this to solar PV and wind useful energy generation, we conclude that rapid growth is likely to continue in coming decades, with solar and wind supplying the majority of global energy by 2050.

1.2.3 Research question 3 – What are likely future energy transitions?

Motivation. The IPCC Assessment Reports present ensembles of energy transition scenarios from Integrated Assessment Models (IAMs). While these scenarios explore various pathways to climate goals, they do not provide probabilistic forecasts [38, 39]. Policymakers need quantitative tools to assess which scenarios are plausible versus which rely on unrealistic assumptions, and to calculate the likelihood of achieving net-zero by specific target years conditional on empirical technology diffusion patterns. Using results and methods from Research Question 2, we derive a probability measure for transition scenarios, providing a more robust basis for decision-making and scenario evaluation.

Main contributions. Chapter 5 provides a methodology to calculate approximate probabilities for an ensemble of transition scenarios. Using distributional forecasts from Chapter 4, we compare IPCC AR6 scenarios for solar PV and wind electricity share against empirically grounded forecasts. We construct likelihood scores enabling probabilistic interpretation of the scenario ensemble. Our analysis demonstrates that AR6 scenarios over-represent low-growth scenarios for solar PV compared to empirical forecasts. Accounting for likely scenarios, we find that low-warming, high-renewables, low Carbon, Capture and Storage (CCS) scenarios are more probable than suggested by the scenario ensemble average. This highlights a need to recalibrate scenario assumptions to include more rapid technology growth (in particular solar).

1.3 Thesis Structure, Methods, and Scope

This thesis comprises seven chapters that integrate literature review, three empirical papers addressing the research questions, and synthesis. Table 1.1 summarizes the thesis structure, and Figure 1.1 illustrates how the papers build upon each other and the literature review, positioned along an S-curve trajectory reflecting the diffusion processes central to this thesis.

Ch.	Component	Link	Purpose & Key Contributions
1	Introduction	–	Motivates the problem and sets up the three linked research questions.
2	Literature review	–	Reviews S-curve theory, Bayesian forecasting methods, and scenario/transition-modelling foundations.
3	Paper 1	RQ1	1. Diagnose and quantify estimation/forecast bias 2. Propose debiasing, applied to EV & solar PV 3. Highlight need for improved S-curve forecasting methods
4	Paper 2	RQ2	1. Cross-technology database + empirical regularities 2. Bayesian distributional forecasting framework 3. Produce probabilistic forecasts for solar PV and wind (used in Paper 3)
5	Paper 3	RQ3	1. Uses Paper 2 distributions to evaluate AR6 scenario pathways 2. Builds likelihood scores for scenario evaluation
6	Discussion	–	Integrates contributions; discusses policy implications, limitations, and future research.
7	Conclusion	–	Summarises the key findings and contributions of the thesis.

Table 1.1: Thesis structure and how the papers connect.

Chapter overview. Chapter 1 motivates the research and introduces the three interconnected research questions. Chapter 2 reviews literature on S-curves, Bayesian forecasting methodology, and energy transition modeling. Chapters 3, 4, and 5 present the three empirical papers addressing Research Questions 1, 2, and 3 respectively,

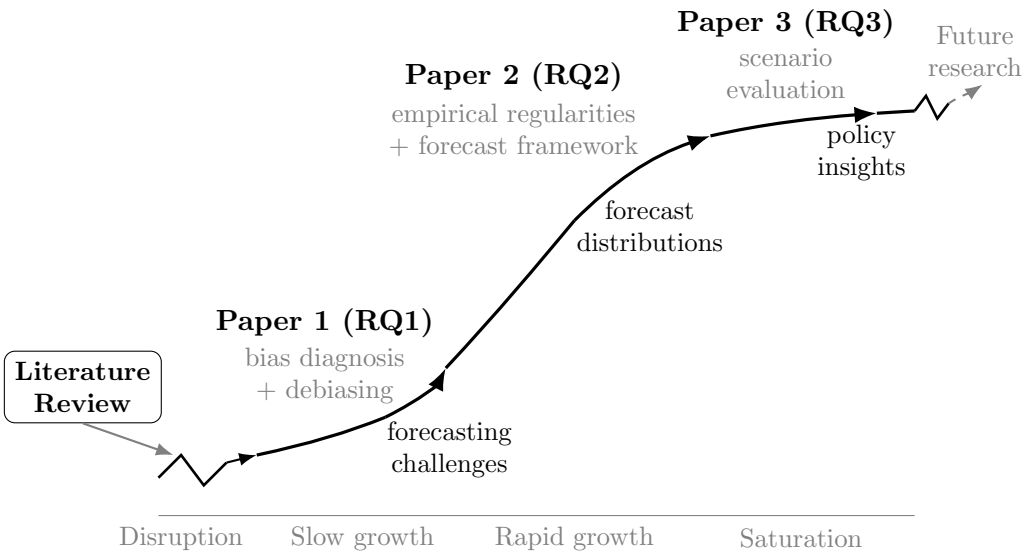


Figure 1.1: Thesis flow and how the papers connect, illustrated along an S-curve trajectory.

with each building on insights from the preceding work. Chapter 6 synthesizes the findings, discusses policy implications and limitations, and proposes future research directions. Chapter 7 summarizes the key contributions and concludes the thesis.

Terminology: forecast, prediction, projection, and scenario. The thesis uses the terms *forecast*, *prediction*, *projection*, and *scenario* in distinct, IPCC-consistent senses [27]. A *prediction* is a best estimate of the actual future evolution of a quantity. A *forecast* is a probabilistic prediction: a distribution over future outcomes derived from data and a statistical model. The Bayesian forecasts in Chapters 3–4 are conditional on observed data and a prior, and should therefore be read as *conditional* probabilistic statements rather than unconditional predictions of what will happen. A *projection* is the potential future evolution of a quantity computed with a model under specified assumptions; projections are explicitly conditional on those assumptions and do not carry attached probabilities. A *scenario* is an internally consistent narrative about how the future may unfold, typically operationalised through a set of model projections under coherent assumptions about drivers (e.g. population, GDP, policy).

Scenarios are used by the IPCC and IEA to explore plausible futures; they are not assigned likelihoods within those frameworks. Chapters 3 and 4 produce *forecasts* in the probabilistic sense above; Chapter 5 compares these forecasts against the IPCC AR6 *scenario* ensemble, which is itself a collection of model *projections*.

Methodological approach. The thesis employs three complementary methods. All three papers use S-curve models and statistical inference. Paper 1 uses simulations to generate synthetic S-curve trajectories with known parameters, enabling controlled diagnosis of estimation bias under varying conditions. Paper 2 combines cross-technology empirical analysis with Bayesian inference, to identify regularities and make well-calibrated forecasts. Paper 3 applies likelihood-based scenario evaluation, comparing modelled transition scenarios against the empirically grounded forecast distributions from paper 2.

Thesis scope. The scope of the thesis is focused on understanding and improving S-curve forecasting methods, understanding technological change, and evaluating energy transition scenarios. The empirical applications (forecasts) emphasise renewable energy generation technologies (solar PV, wind) and, secondarily, electric vehicles, reflecting the centrality of renewable electricity to energy transitions. Other relevant applications exist, such as AI. The analysis operates primarily at global or national aggregation levels and does not explicitly model energy system feedbacks (e.g., grid integration constraints, supply-chain bottlenecks, or policy endogeneity). As a result, forecasts and scenario likelihoods should be interpreted as *empirical plausibility conditional on historical diffusion patterns* rather than as full-system or political feasibility claims.

Chapter 2

Literature Review

2.1 S-Curves: Theory, Applications, and Forecasting

S-curves are non-linear functions that describe the diffusion, adoption, or growth trajectory of a phenomenon over time [40, 41]. These curves exhibit three distinct phases: an initial slow growth phase, a rapid growth phase, and a saturation phase where growth levels off [42, 43]. The initial phase represents early adoption, where only a few individuals or entities adopt. This is followed by a rapid growth phase, where adoption accelerates as more individuals become aware and start adopting. The saturation phase occurs when the majority of the population has adopted, leading to a slowdown in growth as the market becomes saturated [5, 41]. A fourth phase, the decline or death phase, can be observed in some diffusion processes, where the phenomenon loses relevance and is substituted by a new superior innovation [37], or in cases of an epidemic (e.g., COVID-19), where the number of infections eventually declines. The S-curve description typically focuses on the first three phases and ignores this decline phase.

The first descriptions of S-curves date back to the 19th century [11, 12]. Gompertz

[11] introduced the Gompertz function in 1825 to model human mortality rates, and shortly after, Pierre-François Verhulst [12] developed the logistic function in 1838 to describe population growth. Since then, researchers have widely used S-curves to model various diffusion processes across different fields [4, 5, 7, 44]. In 1957, Griliches fitted S-curves to the diffusion of hybrid corn across US states [8]. In 1972, Crane identified logistic S-curve patterns across a range of agricultural innovation diffusion [45].

Fisher and Pry [37] used logistic curves to forecast substitution of new technologies for old ones, applying it to energy. Bhargava [46] extended their model in an attempt to better model energy substitution. Fouquet [47] shows that multi-century energy transitions (coal, oil, gas) follow S-curves spanning 50-100 years, and Comin and Hobijn [48] find S-curves universal across 15 technologies and 25 countries over two centuries. Grubler et al. [7, 15, 16, 49] provide extensive analysis of S-curve patterns in energy systems, transport modes, and industrial processes.

This empirical universality has motivated the extensive application of S-curve models as a fundamental description of technology diffusion dynamics. Geroski [50] provides a comprehensive review of S-curve models for technology forecasting, and Kucharavy and De Guio [44] review their application to innovation diffusion. Beyond technology, S-curves have found extensive application in fields ranging from biology [51] to scientific literature growth [9, 10], the science of science [52], and technology policy [53].

2.1.1 Mathematical Foundations

S-curves are a phenomenological pattern: many diffusion processes exhibit an S-shaped trajectory when viewed at the aggregate level. However, technological change unfolds in complex, path-dependent systems, which can limit the reliability of smooth S-curve extrapolations. External shocks such as policy shift, feedback processes, and competition can produce departures from an idealized S-shaped trajectory [54–57]. This motivates statistical approaches that quantify uncertainty rather than relying on a single best-fit curve.

In the remainder of this section, we introduce a generalized S-curve formulation that nests commonly used special cases such as the logistic and Gompertz curves, and we discuss how these models are fit to empirical diffusion data. We then briefly outline candidate mechanisms that can give rise to aggregate S-shaped dynamics, providing interpretation without assuming a single universal causal explanation.

The generalized S-curve model. To accommodate the complexity of real-world diffusion processes, researchers have added additional parameters to the basic S-curve models [14, 58, 59]. Most S-curve models can be derived from a single general differential equation [60]. Tsoularis and Wallace [14] present a differential equation that describes most S-curve models as

$$\frac{dQ_t}{dt} = kQ_t^{\alpha}\left[1 - \left(\frac{Q_t}{L}\right)^{\beta}\right]^{\gamma} \tag{2.1}$$

where Q_t is the diffusion level, L is the asymptotic stock for $t \to \infty$, k is the exponential growth rate for $t \to -\infty$, and α, β, γ determine the shape of the S-curve, such as the length of the approximate exponential growth phase and the position of

the inflection point (point of maximum growth).

The Bertalanffy-Richards curve. When $\alpha = \gamma = 1$, eq. 2.1 reduces to the Bertalanffy-Richards (B-R) curve [61–63], [1]

$$\frac{dQ_t}{dt} = kQ_t \left[1 - \left(\frac{Q_t}{L}\right)^{\beta}\right] \tag{2.2}$$

with analytical solution:

$$Q(t) = \frac{L}{(1 + e^{-\beta k(t-t_0)})^{1/\beta}}. \tag{2.3}$$

Originally developed to model biological organism growth (particularly fish [65] and other animals [64]), this curve has since been applied to technology diffusion and energy transitions [66]. The parameter β determines the symmetry and inflection point position. When $\beta > 1$, the inflection point occurs after 50% adoption; when $\beta < 1$, it occurs before 50% adoption.

Logistic curve - the symmetric and exponential S-curve. Setting $\beta = 1$ in eq. 2.3 reveals the symmetric S-curve

$$Q(t) = \frac{L}{1 + e^{-k(t-t_0)}} \tag{2.4}$$

where t_0 defines the inflection point. This is known as the logistic curve [50, 67]. The two defining characteristics are its symmetry around the inflection point and its approximate exponential growth behavior for early levels of diffusion where $t \ll t_0$.

[1] A generalized S-curve function of the Bertalanffy function [64] and a simplification of the Richards curve [13].

Researchers have widely applied the logistic curve across diverse fields. Applications include technology adoption [8, 36, 49, 68, 69], patent data [70], population dynamics [71], cell population growth [72], epidemic spreading [2, 73, 74], biology [51, 75], the science of science [52], and numerous other domains [4, 58, 59, 76, 77]. Ayres [78] claimed that the logistic curve may be considered a quasi-law of technological change [79].

Gompertz curve - the asymmetric and superexponential S-curve. Despite the popularity of the logistic curve, others have suggested alternative forms that may provide better representation of empirical data [7, 20]. The Gompertz curve addresses the desire for asymmetry in diffusion processes. It can be derived from eq. 2.2 by dividing by β and taking the limit $\beta \to 0$, yielding

$$Q(t) = Le^{-e^{-k(t-t_0)}} \tag{2.5}$$

The Gompertz curve is characterized by its asymmetry around the inflection point and its superexponential growth behavior for early levels of diffusion where $t \ll t_0$. Researchers have widely used it to model epidemics [80], cell biology [81], population dynamics [59], technology adoption [37, 46, 69, 78, 82], tumour growth [3], bacterial growth [83], and more [11, 84–90].

2.1.2 S-curves in empirical observations

Empirical evidence for early-stage growth behaviors. The logistic and Gompertz curves exhibit fundamentally different growth behaviors. The logistic curve has approximate exponential growth for early diffusion ($t \ll t_0$), meaning $\log Q_t$

grows linearly with constant slope k. The Gompertz curve exhibits growth where the log-growth rate $\frac{d \log Q_t}{dt}$ decays exponentially rather than remaining constant.

Nagy et al. [23] provide systematic empirical evidence that energy technologies exhibit approximate exponential growth during early diffusion stages, supporting the logistic model's growth assumption. Subsequent work by Baumgartner et al. [31] and Way et al. [29] confirms this pattern across diverse energy technologies. Comin et al. [91] also find that many technologies exhibit exponential growth during early adoption phases, preceded by a period of faster-than-exponential growth.

Despite these different theoretical growth assumptions, both logistic and Gompertz curves can fit empirical data well [41, 69, 92, 93], while yielding fundamentally different growth patterns; Logistic is exponential and symmetric, while Gompertz prescribes exponentially decaying growth and an asymmetric adoption curve.

Empirical model comparison studies. Numerous empirical studies have compared logistic and Gompertz models across applications with mixed results. Sudtasan et al. [94] conclude that Gompertz best fits telecommunications diffusion in Thailand. Vieira et al. [95] and Dhar and Dutta [96] similarly favor Gompertz for certain growth processes. Conversely, Akin et al. [97] find the logistic equation better models bacterial growth, while Gompertz excels at tumor growth [85]. Satoh et al. [93] show that US mobile phones and tractors were better modeled using logistic, while cars in the Netherlands and computer-aided software engineering jobs fit Gompertz better. Nagula [98] finds Gompertz provides better forecasts for fuel-cell technologies. Bakher et al. [92] find that while logistic and Gompertz fit similarly well, the Bass model fits online shopping adoption in Australia better than either. These mixed results highlight that model choice matters but no single model dominates across

contexts.

The theoretical underpinnings also remain contested. Comin et al. [91] find that the logistic curve generally fails to fully capture technology diffusion dynamics at early stages, while the Gompertz curve fails to capture exponential growth despite better representing asymmetric patterns. Epidemics similarly exhibit asymmetric patterns [74, 80], though both logistic and Gompertz have been applied to epidemic modeling. These challenges motivate principled approaches to model selection and averaging (Section 2.2).

This raises a deeper question about whether exponential growth patterns must inevitably continue. Ayres [99] argues that while historically many technologies have exhibited exponential growth, this progression need not continue indefinitely—societal resources, expectations, and values that enable exponential growth may shift, causing technologies to deviate from exponential patterns even at relatively early stages [100, 101]. For instance, a societal shift towards sustainable consumption and minimalism could reduce energy demand growth, slowing the current exponential trajectory [102]. Unexpected external shocks, such as technological breakthroughs, policy interventions, or tipping points, can also cause sudden changes in diffusion trajectories [103–106]. Recent examples include the policy of *flattening the curve* during the COVID-19 pandemic [107] and the rapid adoption of remote work technologies [108].

Why symmetry matters for forecasting. The choice between symmetric and asymmetric S-curves can fundamentally shape forecasting outcomes. Symmetric S-curves assume the growth and decline phases mirror each other, where rapid growth leading to the inflection point is followed by an equally rapid deceleration toward saturation [10]. Asymmetric S-curves, in contrast, allow for different dynamics on

either side of the inflection point. This distinction becomes critical when extrapolating from partial data: if we observe only the early growth phase, the assumption of symmetry versus asymmetry dramatically affects where we place the inflection point and ultimate saturation level [20].

Evidence for asymmetry. Grubler [7] examined historical energy transitions and found that asymmetric models often fit better than symmetric logistic curves, with slower initial growth followed by more rapid acceleration better capturing coal, oil, and natural gas adoption in the 19th and 20th centuries. Meade and Islam [20] demonstrate that capturing the appropriate degree of symmetry matters more than the specific functional form, though the optimal degree may vary across technologies and contexts. While multiple S-curve formulations may fit historical data well, this makes it difficult to determine which model truly describes the underlying diffusion dynamics. This creates fundamental forecasting challenges, as different models can yield similar historical fits but divergent long-term predictions.

2.1.3 The B-R Curve: A Flexible Compromise

Addressing the symmetry question with flexibility. The B-R curve, given in eq. 2.3, offers a practical solution to both the symmetry debate and the ambiguity of model selection problem identified by Meade and Islam [20]. The shape parameter β controls the degree of asymmetry and encompasses both logistic ($\beta = 1$) and Gompertz (dividing by β and taking limit $\beta \to 0$) as special cases. This allows the data to determine the appropriate asymmetry rather than requiring the modeler to choose a priori. This flexibility makes the B-R curve attractive for forecasting applications where the true underlying diffusion mechanism is uncertain [14]. Wang et

al. [109] applied the Richards curve (generalized B-R curve) to COVID-19 forecasting, while Ye et al. [66] use the B-R curve for energy transition forecasts with Bayesian regularization to prevent overfitting.

Model selection and overfitting. The B-R curve's flexibility comes at a cost. Adding the β parameter increases model complexity, which can lead to overfitting when data are limited, producing unreliable parameter estimates and poor out-of-sample forecasts [110]. Occam's razor, the principle of preferring simpler explanations when there is lack of evidence for complexity, has rigorous statistical justification through the bias-variance tradeoff [111]. Because all models are approximations of reality [112], simpler approximations often prove more useful in out-of-sample prediction. Complex models fit training data well (low bias) but are sensitive to noise and overfit (high variance), while simple models are more robust (low variance) but may miss important patterns (higher bias). For S-curve forecasting with limited data, this favors parsimonious logistic, Gompertz, or B-R (with fixed (a-)ssymmetry β) curves over flexible alternatives unless data justify additional complexity.

2.1.4 Forecasting Challenges and Renewable Energy Applications

Estimation bias with partial data. Forecasting S-curves has proven challenging across many domains. Van den Bulte and Lilien [18] identified systematic bias in S-curve parameter estimation when only partial diffusion data are available, while Baumgartner et al. [31] show that non-linear least squares (NLLS) S-curve fitting produces highly variable and unreliable forecasts with limited data. Kung et

al. [22] demonstrate these challenges in practice, finding that S-curve forecasts of COVID-19 deaths systematically underestimated actual mortality during 2020, with fitted models predicting premature saturation and failing to anticipate subsequent waves. These estimation challenges have important implications for energy transition forecasting and policy planning (discussed in Section 2.3).

Renewable energy S-curve forecasting applications. Recent efforts have focused on forecasting growth paths of solar, wind, heat pumps, and EVs using S-curve models [66, 69, 113, 114]. These studies reach different conclusions about renewable growth trajectories. Such discrepancies may arise from methodological choices including model selection, prior specification, data aggregation levels, and validation approaches, which affect forecast outcomes even within rigorous probabilistic frameworks (discussed in detail in Section 2.2).

2.1.5 Diffusion Mechanisms

S-curves may emerge from social and economic diffusion processes where individual adoption decisions aggregate into characteristic patterns. Several theoretical frameworks offer explanations for how adoption spreads through populations and why the resulting curves exhibit S-shaped dynamics.

Laggard-innovator patterns. Rogers [40] categorizes populations into innovators (2.5%), early adopters (13.5%), early majority (34%), late majority (34%), and laggards (16%). This sequential adoption process generates the S-curve shape. The transition from early adopters to early majority occurs around 16% penetration, which is a critical threshold that Moore [115] calls the *chasm* separating visionary

from pragmatic adopters. This innovator-laggard pattern extends geographically. Granovetter [116] formalized how heterogeneous adoption thresholds aggregate into cascades: once early adopters with low thresholds adopt, they influence marginal adopters, triggering an avalanche effect. This explains sudden S-curve acceleration at the inflection point [117]. Comin et al. [118] analyze 115 technologies and find that follower countries catch up to technology leaders faster over time, with the adoption gap half-life declining from 35 years (pre-1925 technologies) to 12 years (post-1950). This suggests diffusion may occur more rapidly for later adopters, and that adoption itself may be accelerating. This is discussed further in section 2.3.

Bass diffusion model. Bass [4] formalizes adoption dynamics by distinguishing between two mechanisms, *innovation* and *imitation*. Innovation represents external influence (advertising, media), while imitation captures word-of-mouth and social contagion. The Bass model describes the adoption through a differential equation, $\frac{dN}{dt} = (p + q\frac{N}{M})(M - N)$, where N is cumulative adopters and M is market potential. As adoption increases, the imitation term dominates, accelerating growth through social reinforcement. Eventually saturation occurs as the pool of potential adopters diminishes. Bass's [4] diffusion model generates asymmetric curves. The Bass model predicts time-to-peak adoption and has been widely applied in marketing and forecasting [41].

Complex contagion in social networks. Other work extends the Bass model's simple contagion framework to account for network structure and multiple reinforcement requirements. While disease spread involves simple contagion (single exposure can transmit), technology adoption often requires complex contagion where

individuals need exposure to multiple adopters before committing [119]. As such, the underlying network topology can alter the resulting diffusion pathway, where early growth occurs rapidly within densely-connected clusters through highly-connected hubs, but slows when reaching less-connected network periphery [120, 121]. Young [122] demonstrates that heterogeneous agent networks with social thresholds generate adoption curves that decelerate before 50% penetration.

Network effects and increasing returns. Arthur [123] demonstrated that technologies with increasing returns experience positive feedback that accelerates S-curve growth once critical mass is reached. Network effects arise when technology value increases with adoption, creating self-reinforcing dynamics [124]. During early exponential growth, technologies benefit from learning-by-doing cost reductions and economies of scale [29, 30], enabling rapid growth. As markets mature, however, the favorable opportunities become exhausted and the transition from increasing to diminishing returns slows further growth.

Mechanistic understanding: observation versus explanation. While these frameworks offer plausible conceptual mechanisms for S-curve emergence, formal derivation of specific functional forms (such as the asymmetric B-R curve with particular inflection points) from first principles remains an open challenge. This thesis observes empirical S-curve patterns, characterizes regularities across technologies, and develops well-calibrated probabilistic forecasting methods based on these regularities. We do not provide mechanistic explanations for *why* technologies follow specific S-curve shapes. That is, deriving the observed functional forms from underlying social, economic, or network processes. Developing such mechanistic

explanations, and testing whether network models, biological scaling analogies, or economic transition theories can formally predict observed diffusion parameters, represents an important avenue for future research (see Section 6).

2.2 Fitting and forecasting S-curves

Building on the S-curve theory and model selection challenges discussed in Section 2.1, this section examines statistical methods for fitting S-curves and generating forecasts. Forecasting technology diffusion trajectories requires balancing multiple objectives: producing accurate point predictions, quantifying uncertainty, avoiding overfitting with limited data, and accounting for model uncertainty when multiple functional forms fit historical data well.

In the empirical applications considered later in this thesis, limited data are often a central constraint. Many technologies are observed only for a short portion of their diffusion trajectory. Moreover, early-stage observations are precisely where forecasts are most sensitive to parameter uncertainty and model choice, but this is also where forecasts are the most useful. This motivates an emphasis on methods that regularize estimation and quantify uncertainty under sparse or noisy information.

We review statistical and econometric approaches with emphasis on uncertainty quantification, model selection, and fundamental limitations in forecasting practice. We discuss tools that can be applied to S-curve modelling, focusing on expert judgment, non-linear least squares (NLLS), and Bayesian inference, followed by model selection, computational implementation, and forecast validation techniques.

2.2.1 Forecasting Techniques Applied to S-curves

2.2.1.1 Expert Forecasting vs. Statistical Models

Technology forecasts often rely on expert elicitation when statistical models are unavailable [125], particularly for long-term energy system planning where historical data may be limited [126]. Yet, experts systematically underestimate uncertainty, a phenomenon known as overconfidence bias [127]. When asked to provide 90% confidence intervals, experts' intervals contain the true value far less than 90% of the time, indicating severe miscalibration [128]. Anchoring effects may bias expert judgments toward current trends or recent experiences [129]. Experts may struggle to envision exponential growth or disruptive change, relying instead on linear extrapolations.

Tetlock [25] tracked expert predictions and found that experts often perform worse than simple statistical models. Moreover, Engelberg et al. [130] find that professional economic forecasters' probability distributions are inconsistent with their own point predictions. Morgan and Henrion [125] emphasize that infrastructure decisions require assessing risks under deep uncertainty, and expert-based deterministic forecasts prove inadequate for such planning.

Examples from technology forecasting. Nagy et al. [23] compared expert technology cost forecasts to simple trend extrapolation for 62 technologies, finding that statistical models outperformed experts for 48 of them. Meng et al. [131] compare expert elicitation to model-based probabilistic forecasts for energy transition technologies, finding that data-driven approaches consistently outperform expert judgment. These systematic biases in expert forecasting motivate probabilistic

forecasts with explicit uncertainty quantification rather than subjective expert ranges [125].

2.2.1.2 Non-Linear Least Squares Estimation

The traditional approach and its implementation. Non-linear least squares (NLLS) serves as the traditional workhorse for fitting S-curves to technology adoption data [132]. The approach minimizes the sum of squared residuals between observed data and model fit by iteratively adjusting parameters $\theta = \{L, k, t_0\}$ until convergence. While computationally straightforward and widely implemented in statistical software, NLLS estimation faces significant challenges when applied to S-curve forecasting [18].

Limitations of NLLS in S-curve forecasting. First, Baumgartner et al. [31] show that NLLS fits to empirical S-curve data yield varied and unreliable forecasts, sensitive to small changes in training data. Second, limited early-stage data often weakly identifies S-curve parameters, which may lead to overfitting, high forecast variance, and biased parameter estimates [18, 19, 133]. Third, NLLS provides point estimates without inherent uncertainty quantification. Standard errors derived assume asymptotic normality, which may be violated with small samples and/or highly non-linear functions [134].

2.2.1.3 Bayesian Inference for Probabilistic Forecasting

Foundations and Bayes' theorem. Bayesian methods yield full posterior distributions over parameters rather than point estimates, with the foundation being

Bayes' theorem

$$p(\theta|Y_t) = \frac{p(Y_t|\theta)p(\theta)}{p(Y_t)} \tag{2.6}$$

where θ are S-curve parameters, Y_t is observed data, $p(\theta)$ is the prior distribution encoding beliefs before seeing data, $p(Y_t|\theta)$ is the likelihood function, and $p(\theta|Y_t)$ is the posterior distribution representing updated beliefs after observing data.

Prior specification Priors encode our beliefs about plausible parameter values before observing data. Well-chosen priors can dramatically improve forecast accuracy by preventing the model from exploring unrealistic parameter regions [135]. Prior distributions range from non-informative to strongly informative depending on the level of prior knowledge [135, 136]. The choice of prior distribution substantially impacts Bayesian forecasts, particularly when training data are limited [136]. Other priors draw from empirical data through empirical Bayes or hierarchical modeling [137, 138].

Traditional priors: non-informative to informative. Prior distributions span a spectrum of informativeness. Non-informative (or uninformative) priors aim to let the data dominate inference but can be problematic for S-curve forecasting: with limited early-stage data, they may allow for implausible parameter combinations and may yield improper posteriors for non-linear models [136]. Often, uniform distributions or flat priors are considered non-informative. Yet, any prior implicitly encodes information; for example, a uniform prior over a wide range may still inform the inference procedure by assigning zero probability to values outside that range.

Weakly informative priors strike a pragmatic balance, providing some regularization to parameter estimates without being overly restrictive [135]. Common examples

include normal distributions with large variances or half-Cauchy distributions for scale parameters [139]. Gelman [139] recommends weakly informative priors allowing a wide range but avoiding extreme values that could destabilize estimation. For S-curves, this may involve setting a wide prior on growth rates k greater than zero, that still limits unrealistically high values.

At the far end of the spectrum, informative priors encode substantive domain knowledge such as resource constraints, policy targets, or physical limits. Ye et al. [66] use informative priors based on policy constraints when forecasting renewable energy capacity, constraining saturation levels using government targets and resource assessments. Van De Schoot et al. [140] apply Bayesian estimation to small datasets in psychology, and demonstrate that well-informed priors can substantially improve parameter estimation when observations are limited. Expert elicitation provides another route to inform priors [141, 142], though cognitive biases including overconfidence and anchoring pose significant risks [25, 125, 143]. This may lead to overconfident posteriors. For example, if experts believe a technology's saturation level L is to be within the policy target range, an informative prior centered on that range may unduly constrain forecasts, especially if data and prior conflict. When using informative priors, transparency and sensitivity analysis are essential to avoid letting subjective beliefs overwhelm empirical evidence.

Empirical and hierarchical priors: data-driven approaches. Empirical Bayes approaches use data from multiple related processes to inform prior distributions, bridging frequentist and Bayesian perspectives [137, 144].

Hierarchical priors extend empirical Bayes by modeling parameters as drawn from a common distribution, itself estimated from the data [138]. Rather than treat-

ing each unit (technology, region, individual) as entirely independent, hierarchical models assume parameters share statistical structure. For example, growth rates $k_1, k_2, \ldots, k_n$ might all be drawn from a normal distribution $\mathcal{N}(\mu_k, \sigma_k^2)$, where we estimate the hyperparameters μ_k and σ_k^2 alongside the individual parameters. This two-level structure enables "borrowing strength" across units, such that those with abundant data inform the hyperprior distribution, which then provides regularization for units with sparse data [138]. Applied to S-curve forecasting, this means technologies with extensive adoption histories inform the distribution of plausible growth rates, saturation levels, and inflection points. For instance, if 100 technologies exhibit growth rates distributed according to some pattern, this distributional knowledge helps constrain forecasts for a 101st technology even when few observations are available.

A pragmatic variant combines empirical Bayes with weakly informative elements. This approach generates more accurate forecasts by combining frequentist and Bayesian perspectives [137], with data-adapted priors reducing sensitivity and improving robustness when data are limited [144]. For example, this could allow setting a weakly informative prior on growth rates k centered around the empirical mean growth rate observed across similar technologies, with a variance reflecting observed heterogeneity and potential expected future uncertainty, due to structural breaks or market changes. However, it makes double use of the data-once for prior specification, again for posterior estimation—which can bias results [145] and underestimate uncertainty by not accounting for hyperparameter uncertainty [137].

Likelihood function. Upon observing the data, Bayesian inference updates prior beliefs by calculating the likelihood function $p(Y_t|\theta)$, which quantifies how probable

the observed data are given specific parameter values [136]. We often construct this assuming normally distributed errors, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, where ϵ_t represents observation noise and $\mathcal{N}(0, \sigma^2)$ denotes a normal distribution with mean 0 and variance σ^2. Across iterations, the likelihood evaluates how well different parameter combinations explain the observed adoption data, converging towards the posterior distribution. In cases of finite data and sampling, the posterior reflects both data evidence and prior beliefs, balancing fit and uncertainty [136].

Prior sensitivity analysis. Prior sensitivity analysis tests how forecasting conclusions change with different prior specifications [146, 147]. If posterior distributions and resulting forecasts change dramatically when priors are varied within reasonable bounds, this indicates the data provide insufficient information to overcome prior uncertainty. In such cases, acknowledging forecast uncertainty is preferable to overstating precision.

Potential advantages of Bayesian S-curve forecasting. The Bayesian approach offers several advantages, particularly when data are scarce [148], as is the case for some S-curve forecasts. The posterior distribution directly quantifies uncertainty. For example, rather than reporting a single growth rate k, we obtain a full probability distribution, reflecting both data limitations and structural uncertainty [149]. Prior distributions allow us to incorporate domain and empirical knowledge [137], stabilizing estimates when data are limited [135].

The Bayesian framework accommodates sequential updating. As new adoption data arrive, the previous posterior can become the new prior, enabling computationally efficient forecast updates without complete refitting [150]. This may be relevant

for operational forecasting systems that must update regularly as new observations accumulate, though we do not implement sequential updating in this thesis (see Section 6.4 for discussion of this potential extension). Finally, Bayesian methods often outperform NLLS with small samples, where informative priors prevent overfitting by regularizing parameter estimates [136].

2.2.2 Model Selection

Different S-curve specifications can fit the same diffusion history nearly equally well while implying very different forecasts. In this thesis, we therefore treat model selection as a predictive question: candidate models are compared using a combination of in-sample fit diagnostics and, crucially, out-of-sample performance assessed via hindcasting (Section 2.2.3). The resulting selection procedure is implemented in the empirical chapters (e.g., Chapter 4). This subsection reviews some other common model comparison tools.

Information criteria balance goodness-of-fit against complexity [110, 151, 152]. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) both select models based on in-sample predictive accuracy, where AIC penalizes complexity less than BIC, which may lead to different model choices [110]. Watanabe-Akaike Information Criterion (WAIC) extends AIC to Bayesian models, accounting for the entire posterior distribution [151], and handles model complexity more robustly. Leave-One-Out Cross-Validation (LOO-CV) estimates out-of-sample predictive performance by iteratively leaving out each data point, fitting the model to the remaining data, and evaluating predictive accuracy on the held-out point [152]. Such methods provide quantitative tools for model comparison, but they do not account for model

uncertainty.

Bayesian model averaging (BMA) [153, 154] offers an alternative by averaging predictions across models weighted by posterior probabilities, allowing uncertainty quantification over models [153]. Implementing BMA introduces additional complexity and inherits prior sensitivity issues. Most applications select a single functional form based on information criteria or domain knowledge, accepting model uncertainty as an irreducible forecast error source. We do not implement BMA in this thesis, but it represents a promising avenue for future research to better account for model uncertainty in S-curve forecasting. We discuss this further in the discussion and limitations (Section 6.3).

2.2.3 Forecast Validation

Because S-curves can achieve excellent in-sample fits even while their extrapolations are poor, forecast evaluation in this thesis emphasizes out-of-sample validation. This subsection introduces a validation framework used to compare candidate specifications and to assess both point accuracy and probabilistic calibration. We first describe hindcasting and rolling-origin evaluation, then diagnostics for distributional forecasts, followed by point-forecast accuracy metrics and cross-sectional validation. These tools are applied to the Bayesian S-curve forecasts in Chapter 4.

Hindcasting and out-of-sample validation. While model selection criteria identify which model to use, validation assesses the absolute quality of the chosen model's forecasts. Validation must extend beyond in-sample goodness-of-fit [31]. R-squared can mislead because S-curves' flexible functional form naturally fits well, masking poor forecast performance. Hindcasting provides critical reality checks.

A common approach is to truncate the historical data, generate forecasts, and then compare predictions to subsequently observed outcomes. This simulates real forecasting conditions, revealing model performance when extrapolating beyond training data [155]. Rolling-origin evaluation [156] further assesses stability by repeatedly retraining and forecasting as new data arrive, mimicking operational forecasting. These methods expose overfitting and assess forecast robustness.

Probabilistic forecast assessment. Probabilistic forecasts require calibration assessment beyond point accuracy [157]. Well-calibrated distributional forecasts are such that, for example, 90% prediction intervals should contain truth 90% of the time [158]. Calibration plots visualize whether credible intervals achieve nominal coverage across forecast horizons. Quantile-quantile (q-q) plots compare realized and predicted quantiles, with deviations from the 45-degree line indicating miscalibration [159]. The probability integral transform (PIT) provides a complementary diagnostic. If forecasts are well-calibrated, the PIT should be uniformly distributed [160]. Convex q-q plots or U-shaped PIT histograms indicate overconfident (too narrow) forecasts, while concave q-q plots or inverse-U-shaped PIT histograms indicate underconfident (too wide) forecasts. The Kolmogorov-Smirnov test [161] can be used to test differences between forecast and observed distributions; the L1 norm provides a complementary distance metric [157]. Together, these diagnose whether a model's predictive distributions are well-calibrated, and offer additional heuristics for model comparison. We employ these methods to assess the calibration and compare models of our Bayesian S-curve forecasts in Chapter 4.

Point forecast accuracy. For point forecasts, mean/median absolute percentage error (MAPE, MdAPE), or mean/median absolute log error (MALE, MedALE) provide interpretable accuracy metrics that are robust to outliers [162]. MAPE and MdAPE are particularly useful for comparing forecast performance across different diffusion levels and forecast horizons, revealing whether accuracy degrades as predictions extend further into the future.

Cross-sectional validation. When data behave similarly across sections (e.g., technologies, regions), cross-sectional validation exploits similar patterns across technologies [163, 164]. By treating technologies as exchangeable draws from a common distribution, one increases effective sample size, enabling model validation. In such cases, forecast performance on historical data can be used to inform calibration for forecasts of new technologies. This approach relies on the assumption that technologies share similar diffusion dynamics, described by the same underlying model.

2.2.4 Computational Methods

Computing Bayesian posteriors typically requires numerical methods, as analytical solutions exist only for special cases. Here we discuss several appropriate numerical methods.

Markov Chain Monte Carlo fundamentals. Markov Chain Monte Carlo (MCMC) algorithms generate samples from the posterior distribution when direct computation is intractable [165]. The fundamental idea is to construct a Markov chain whose stationary distribution equals the target posterior, where after sufficient

iterations, samples from the chain approximate samples from the posterior.

Sampling algorithms. The Metropolis-Hastings algorithm [166, 167] is the classical MCMC sampling procedure. At each iteration, it proposes a new parameter value, then accepts or rejects based on the ratio of posterior densities. The optimal acceptance probability is often considered to be around 23.4% for high-dimensional spaces [168, 169]. The acceptance rule ensures the chain converges to the correct posterior distribution, though proposal choice affects convergence speed. Random walk proposals centered on current parameter values are common, though more sophisticated adaptive schemes can improve efficiency [170]. More sophisticated sampling algorithms, such as Hamiltonian Monte Carlo [171] and its adaptive variant No-U-Turn Sampling [172],use gradient information to explore complex posterior geometries more efficiently, particularly for high-dimensional parameter spaces with strong correlations. When limited to few parameters as in standard S-curve models, the simpler Metropolis-Hastings algorithm typically suffices.

Convergence diagnostics. Convergence diagnostics verify that MCMC chains have reached the stationary distribution. Running multiple chains initially provides robust convergence assessment [173], though single chains may suffice for production forecasting once convergence properties are established. Effective sample size (ESS) quantifies how many independent samples the autocorrelated MCMC chain provides after accounting for thinning [136]. Visual diagnostics like trace plots and autocorrelation plots complement these numerical metrics. Practical convergence checks include verifying that longer sampling runs or higher thinning factors do not substantially change posterior estimates or out-of-sample forecast accuracy. Without proper

convergence, posterior samples may not represent the true posterior distribution, leading to unreliable forecasts.

Computational complexity and practical considerations. MCMC requires thousands of likelihood evaluations. For complex models with hierarchical structures across multiple technologies or large datasets, computation can become expensive. Parallel chains enable both faster computation and convergence diagnostics. For typical S-curve forecasting applications, MCMC remains feasible on standard computing hardware, requiring minutes to hours rather than days. Alternative approaches such as variational inference [174] offer approximate Bayesian inference via optimization rather than sampling, trading accuracy for computational speed.

2.2.5 Applications to Energy and Climate

These computational tools enable practical applications across energy and climate policy domains.

Probabilistic forecasts for policy evaluation. Probabilistic forecasting methods enable quantitative policy evaluation under uncertainty through explicit probability statements. Wilson et al. [175] applied sequential Bayesian updating with informative priors to forecast coal electricity decline, estimating a 70% probability that UK coal generation falls below 5% by 2025 conditional on maintaining the carbon price floor. UK coal electricity generation decreased from 40% in 2012 to under 2% in 2020, consistent with the forecast [176]. This demonstrates how probabilistic forecasts can inform policy decisions and provide accountability for policy effectiveness.

Raftery et al. [177] illustrate uncertainty propagation through integrated model-

ing chains. Their Bayesian probabilistic projections of population and GDP feed into emission scenarios and climate models, estimating $> 2°C$ warming has 90% probability under current policies. These analyses demonstrate the value of probabilistic forecasting in quantifying uncertainty for assessing policy effectiveness and climate risks.

Forecast sensitivity to methodological choices. Renewable energy forecasting illustrates how methodological choices critically affect conclusions even when analyzing identical phenomena. Cherp et al. [69] use historical growth rates to forecast future energy growth. Ayres [78] notes that extrapolation of historical growth rates to inform growth rates of future innovations is misleading, and likely leads to underestimation of future growth. Further, Cherp et al. [113] generate probabilistic forecasts via bootstrapping from (altered) S-curve fits. Ye et al. [66] apply Bayesian inference with informative priors determined by policy targets, similarly forecasting constrained growth despite the different probabilistic framework. By contrast, Zielonka et al. [114] use Bayesian methods and predict increased yet heterogeneous growth in Europe, with renewable forecasts often encompassing or surpassing the green deal targets.

These juxtaposed conclusions for the same technologies demonstrate that prior specification, model selection, and estimation methods can dominate forecast outcomes even within rigorous probabilistic frameworks. A related concern, is that several studies introduce *ad hoc* constraints. Examples include imposing cost floors or exogenous caps on maximum growth rates for solar and wind, which can strongly shape projected diffusion. We examine the implications of such constraints in Chapter 5. This sensitivity underscores the importance of transparent methodology, sensitivity

analysis, and validation against out-of-sample data rather than relying solely on in-sample fit or theory-driven constraints.

2.2.6 Challenges and Limitations

Data quality and non-stationarity. Technology adoption exhibits structural breaks, regime shifts, measurement error, heteroscedasticity, and autocorrelation that may violate standard econometric forecasting assumptions [91, 178, 179]. Residual diagnostics and appropriate transformations may detect and mitigate some issues, though some violations may persist.

Model misspecification and limits of extrapolation. When the true process lies outside the candidate model space, all models will be misspecified and produce poor forecasts. Testing multiple models and reporting only the best fit inflates apparent accuracy while masking true uncertainty (specification search bias) [180]. Posterior predictive checks provide valuable diagnostics [181], though they cannot detect unknown uncertainty outside the model space. Beyond model class limitations, rare high-impact events—policy shocks, disruptions, innovations—can radically alter trajectories [182, 183]. COVID-19 rendered 2019 energy forecasts obsolete [184]. While predictable regularities exist [7, 69], forecasts must quantify uncertainty including structural break possibilities, with decision-makers employing robust decision-making approaches that perform well across multiple scenarios [185].

Prior specification and sensitivity. With limited data, priors substantially affect posteriors, raising concerns that subjective choices may drive forecasts more than empirical evidence [135]. Prior sensitivity analysis where one varies hyperparameters

to assess whether posteriors remain stable provides essential diagnostics [142]. Best practice uses weakly informative priors that regularize toward plausible values while allowing data to override initial beliefs [135]. Prior sensitivity analysis tests how forecasting conclusions change with different prior specifications [146, 147]. If posterior distributions and resulting forecasts change dramatically when priors are varied within reasonable bounds, this indicates the data provide insufficient information to overcome prior uncertainty.

2.3 Energy Scenarios and Transition Pathways

Having established S-curve theory and forecasting methods in the previous sections, we now examine how these approaches apply to energy transition modeling and policy. Energy transitions occur over multiple decades and involve complex interactions between technologies, economics, policy, and social behavior [186]. To navigate this uncertainty, policymakers rely on energy scenarios—internally consistent narratives describing potential future energy system evolution [187]. Integrated Assessment Models (IAMs) formalize these scenarios mathematically, projecting energy demand, supply, emissions, and costs under different assumptions [188]. The Paris Agreement requires limiting warming to 1.5-2°C [26], necessitating rapid renewable deployment and deep decarbonization by mid-century. IAMs provide the primary analytical framework for assessing whether such targets remain achievable and at what cost.

2.3.1 Integrated Assessment Models and IPCC Scenarios

Integrated Assessment Models (IAMs). The most prominent global IAMs include AIM (Asian-Pacific integrated assessment) [189], GCAM (Global Change

Analysis Model focusing on technology competition) [190], IMAGE (spatially explicit model linking energy, land use, and climate) [191], MESSAGE (energy systems optimization with detailed technology representation) [192], and REMIND (recursive dynamic optimization coupling economy and climate) [193]. Other widely used models include TIMES (technology-rich bottom-up optimization models) [194], and WITCH (endogenous technical change with climate policy) [195]. These IAMs integrate energy systems, economics, land use, and climate physics into unified frameworks, enabling exploration of technology portfolios, policy instruments, and transition pathways.

Shared Socioeconomic Pathways for climate scenarios. The IPCC's Sixth Assessment Report (AR6) [27] uses scenarios that are generated by IAMs to inform readers about the costs, benefits and trade-offs associated with varying levels of decarbonization ambition. The IAMs utilize a consistent range of Shared Socioeconomic Pathways (SSPs) for these scenarios, which are designed to span a wide range of plausible futures [196, 197]. They include SSP1 (sustainability), SSP2 (middle of the road), SSP3 (regional rivalry), SSP4 (inequality) and SSP5 (rapid growth). Researchers apply these SSPs to generate future energy and mitigation scenarios to explore different transition pathways in a wide array of socio-economic contexts.

Illustrative Mitigation Pathways in IPCC AR6. The IPCC AR6 Working Group III (WGIII) report selected five Illustrative Mitigation Pathways (IMPs) from the AR6 scenario database to represent the range of possible transition approaches [39, 198]. IMP-Ren emphasizes rapid renewable deployment, IMP-LD (Low Energy Demand) achieves 1.5°C through efficiency and behavioral change without large-scale

carbon removal [68], IMP-SP (Shifting Pathways) combines demand-side lifestyle changes with rapid supply-side transitions, IMP-Neg relies heavily on carbon dioxide removal, and IMP-GS (Gradual Strengthening) features delayed mitigation with policies intensifying slowly over time.

2.3.2 Historical Transition Speeds

To contextualize and evaluate these IAM projections used by the IPCC, and to assess whether Paris Agreement timelines are feasible, we examine the speed of historical energy and technological transitions.

Speed of transitions. Primary energy transitions have historically required centuries [199]. Marchetti and Nakićenović [200] document that primary energy source substitutions follow logistic patterns with time constants around 100 years, with coal peaking around 1920 and oil in the 1960s. Actual transitions deviated from pure logistic substitution patterns. For example, oil's share of UK primary energy did not reach near 0 by 2000 as expected by the model, but still accounts for almost 30% of UK primary energy[201].[2] Fouquet [47] documents UK wood-to-coal taking several hundred years (16th-18th centuries), and gas-to-electric lighting requiring 70 years (1880-1960) after an initial 70 years from invention to impact (1810-1880). Geels found the transition from sailing ships to steamships spanned over 100 years (1780–1900) [199]. Sovacool [202] finds aggregate transitions require a minimum of 40 years for single systems, potentially centuries for complete transformation. Gross et al. [203] find successful energy technologies require 15-70 years (median 35) from prototype to impact.

[2]This limitation applies equally to the S-curve forecasting approach employed in this thesis, as discussed in Section 6.3.

Socio-technical transitions could, however, be accelerating [16, 204]. Grubler [205] finds coal phase-out (47-67yr) proceeded faster than coal adoption (96-160yr), and Fouquet [47] documents UK's coal-to-gas transition faster than wood-to-coal. However, substitution times vary dramatically by technology. Fischer-Pry [37] show 58 years for synthetic fibers (10%-90% share) yet 10.5 years for steel furnaces. Grubler [7] finds automobiles replaced horses in 12 years (10%-90%), while air transport took 30 years to reach 50% share despite diesel/electric locomotives replacing steam in 12 years [206]. Furthermore, Sovacool [202] documents ten rapid transitions achieving 1-25% penetration in 1-16 years: Indonesian LPG (3yr), Brazilian flex-fuel vehicles (1yr), France nuclear (11yr, 1974-1982), Netherlands gas (10yr), Ontario coal phase-out (11yr).

Renewables: modular diffusion at unprecedented speed. Renewable electricity growth is historically unprecedented. Cherp et al. [207] document Germany's rise from 0% to 30% (1990-2015), though still slower than Paris-compatible 70-90% by 2040 [208]. Solar PV deployment patterns (2010-2020) resemble mobile phones (15yr to 50% penetration, 1995-2010) and internet (20yr, 1995-2015) [209] more than traditional primary energy transitions (40-100yr). Solar reached 1% global electricity share in 2010, then accelerated to 3% by 2020 [29] and 7% by 2024 [210]. Wind took 30 years to reach 1% (1980-2010), then accelerated to 6% by 2020 and 8% by 2024 [210]. Wilson et al. [211] argue that modularity determines adoption speed—modular renewables such as solar and batteries enable mass production and rapid learning compared to bulky technologies such as nuclear and CCS. Whether this unprecedented speed can be sustained to achieve Paris targets, however, remains an open question.

2.3.3 Limitations of Scenarios

Despite their central role in climate policy, IAMs exhibit persistent shortcomings. Recent comprehensive reviews identify systematic biases that may mislead policy decisions.

Systematic underestimation of renewable deployment and costs. Pfenninger et al. [32] and Ringkjøb et al. [33] highlight insufficient uncertainty quantification, lack of empirical validation and backtesting, and limited transparency in IAM frameworks, arguing that energy systems models *cannot be properly validated* and *should be seen as a source of possible storylines rather than of fundamental truth.* Krumm et al. [212] find that social aspects enter mainly as exogenous assumptions rather than endogenously modelled dynamics. DeCarolis et al. [213] emphasize that researchers rarely backtest IAMs against historical data despite these models informing trillion-dollar policy decisions, while Kachirayil et al. [214] find that flexibility and robustness challenges remain underrepresented.

Beyond these methodological limitations, IAM scenarios have been found to systematically underestimate renewable deployment. Creutzig et al. [28] document this pattern across many energy scenarios, with solar PV forecasts often $10-20\times$ too pessimistic [29]. Ives et al. [24] and Way et al. [29] provide further empirical evidence that renewable deployment has consistently outpaced IAM projections across multiple technologies and regions.

Multiple factors contribute to this systematic bias. Pfeiffer et al. [215] identify expert judgment anchored on historical fossil fuel trajectories (systematic biases in expert forecasting are discussed in Section 2.2). Wilson et al. [175] provide empirical evidence that MESSAGE and REMIND scenarios require longer durations

to reach given growth extents than historical technology diffusion patterns suggest, attributing this conservatism to parametric growth rate constraints. Wilson and Grubler [216] demonstrate that IAM scenarios emphasize supply-side technologies while underrepresenting efficiency improvements. Historical R&D investment allocated only 9% to efficiency (1974-2008), yet scenario analysis projects efficiency delivering 59% of cumulative emission reductions (2000-2100). Gilbert and Sovacool [217] identify persistent bias against renewables in US energy models, attributing this to institutional inertia and model assumptions favoring incumbent technologies. This mismatch between past investment and projected impact suggests IAMs systematically undervalue efficiency improvements, which could lead to underestimation of transition speeds and overestimation of costs.

These biases have theoretical foundations. Arthur [123] demonstrates that under increasing returns to adoption, early advantages of existing technologies create path dependence that can exclude superior alternatives. Beyond technical factors, Süsser et al. [218] and Sovacool et al. [219] show that energy models sometimes serve political agendas, with models selectively designed or interpreted to justify predetermined outcomes rather than objectively informing policy.

Conservative cost assumptions compound these biases by failing to capture empirical learning dynamics [23]. Applying technology cost declines with cumulative production, or Wright's Law, [220], have provided empirical learning rates that have proved remarkably consistent across renewable energy technologies [30]. Nemet [221] reports solar PV learning rates of 17-26%, later narrowed by Farmer and Lafond [30] and Way et al. [29] to 20-24%. Wiser et al. find wind learning rates around 10-15% [222], and Ziegler and Trancik [223] report battery learning rates of 19-21%. Yet

many IAM scenarios implement cost floor assumptions contradicting this evidence [224]. Krey et al. [225] find that nearly all major IAMs assume capital costs for solar PV plants close to or higher than 1 €/W in 2050, despite 2019 costs already averaging 0.995 €/W. However, Gritsevskyi and Nakićenović [226] demonstrate that technological learning itself is fundamentally uncertain, finding 53 different technology dynamics with similar overall costs but fundamentally different energy system structures, where near-term investment decisions can create lock-in effects that matter more than later decisions for determining long-term trajectories.

These systematic biases have profound economic implications. Way et al. [29] show that learning-curve-based projections predict net-zero transitions will likely save $12 trillion compared to fossil baselines, contradicting the traditional IAM view exemplified by Nordhaus's DICE model [227] that mitigation is costly and economically optimal warming is around 2.5-3°C. Victoria et al. [224] argue that solar PV is already cost-competitive and further innovation with suitable policy support could accelerate cost declines, enabling further rapid growth. Sivaram [228] discusses that solar PV could be a main energy provider, but only under suitable policy support and avoid a similar fate to nuclear power.

Perfect foresight versus bounded rationality. Most optimization-based IAMs assume perfect foresight, where decision-makers know future technology costs, resource availability, and climate damages when making current investments [193, 229]. This contrasts with reality where investors face deep uncertainty about future policies, technology breakthroughs, and geopolitical shocks [230], while bounded rationality, satisficing behavior, and organizational routines dominate actual decision-making [231]. Scamman et al. [232] critique that perfect foresight models lack endogenous

feedback loops and sequential decision-making that characterize real-world energy transitions. Some IAMs such as MESSAGE offer myopic alternatives where one optimizes 5-10 years ahead, producing patterns closer to empirical deployment but potentially missing long-term optimality [229].

2.3.3.1 Deterministic Framing and Missing Uncertainty

Lack of probabilistic assessment. AR6 scenarios inform international climate negotiations, national climate targets, and energy investments [39]. Scenarios explore *what if* pathways conditional on assumptions [38], not probability distributions over futures. The spread across scenarios reflects structural uncertainty regarding model assumptions rather than probabilistic likelihood [233].

Likelihood in the IPCC scenario framework. A central distinction is between (i) the likelihood of a *climate outcome conditional on a given pathway* and (ii) the likelihood that society follows a particular socioeconomic pathway in the first place. In the IPCC framework, IAM scenarios are used as internally consistent *conditional* explorations rather than forecasts, and AR6 does not assign probabilities to specific SSP–mitigation pathways [234, 235]. Probabilistic language in AR6 therefore typically refers to uncertainty in assessed outcomes (e.g., temperature implications) conditional on emissions trajectories, rather than to the likelihood of the trajectories themselves [235]. Moreover, the AR6 scenarios database constitutes an ensemble of opportunity, reflecting research activity and modelling conventions. Thus, its distribution should not be interpreted as a probability distribution over socioeconomic futures [236].

This deterministic framing has important implications. Trutnevyte et al. [34]

survey more than 150 scenario studies and find that over 90% present deterministic pathways without probability distributions. Through ex-post analysis of the UK electricity system (1990–2014), they demonstrate that cost-optimal scenarios deviate substantially from realised transitions, with cumulative costs differing by 9–23% depending on assumptions. This suggests that cost-optimization, while computationally tractable, may systematically misrepresent actual transition dynamics and costs, potentially misleading policy decisions.

On the other hand, the vetted scenarios presented in IPCC AR6 represent a smaller and more policy-relevant subset of the full scenario ensemble. Recent work by Kazlou et al.[237] narrows the AR6 database to a feasible set that limits warming while relying on more realistic levels of carbon dioxide removal (CDR). Similarly, Hausfather and Peters [238] evaluate the accuracy of past emission scenarios, finding that while some trajectories have tracked observations reasonably well, others have systematically overestimated emissions. These efforts improve scenario appraisal, but they stop short of providing formally calibrated probability distributions over technology deployment pathways.

Attempts to attach likelihood to scenario ensembles and to benchmark scenarios against empirical forecasts. A growing literature has sought to make scenario ensembles more decision-relevant without claiming to predict socioeconomic futures. Ho et al. [239] elicit *subjective* probabilities over emissions outcomes from experts rather than assuming that published scenarios are equally likely. Beath et al. [240] develop a scenario-ensemble weighting method to correct for biases, such as dominance of particular models or intercomparison projects. This attempts to produce summaries that are more balanced and transparent for decision makers

[241].

A complementary line of research benchmarks normative scenarios and policy targets against empirically grounded diffusion forecasts. Zielonka [242] and later, Zielonka Trutnevyte [114], use probabilistic S-curve projections and compare these to net-zero pathways or required deployment levels to estimate the feasibility of meeting scenario-consistent targets under continuation of observed diffusion dynamics. Uidhir et al. [243] combines diffusion modelling with energy-system and policy simulation frameworks to evaluate whether policy targets imply diffusion rates beyond historical precedents, providing an empirical feasibility screen for scenario narratives.

Together, these approaches illustrate practical routes toward a more probabilistic scenario assessment by quantifying how consistent particular pathways are with empirical diffusion regularities and forecast uncertainty. This thesis contributes to this emerging literature by developing probabilistically well-calibrated S-curve forecasts and applying them to assess the empirical plausibility/feasibility of scenario-consistent deployment trajectories.

Multiple sources of uncertainty compound over time. Technology learning rates for solar PV vary year-to-year [29], carbon prices differ substantially per tonne CO_2 by 2050 across scenarios [197], and behavioural parameters in technology adoption [244] represent additional layers of uncertainty. Harmsen et al. [245] show that incorporating technology cost uncertainty expands 2050 emission ranges by 50–100% compared to deterministic scenarios.

2.3.4 Policy Implications of Scenario Limitations

IAM scenarios directly inform policy decisions worth trillions of dollars [213], and depending on modelling choices, investing in renewables or fossil fuels may be economically beneficial [29, 227]. Paltsev [246] argues that if history is any guide, energy scenarios overestimate the extent to which the future will look like the recent past[3], yet policymakers continue to rely on these projections for long-term infrastructure and investment decisions. Systematic biases and lack of uncertainty quantification create risks for long-term planning under deep uncertainty [185].

Paris feasibility despite modeling constraints. Despite these limitations, recent research suggests that Paris Agreement targets of 1.5-2°C remain technically feasible, in part because IAMs systematically underestimate renewable deployment speeds. Rockström et al. [247] propose a carbon law of halving emissions every decade, requiring doubling of zero-carbon energy shares every 5-7 years. Grubler et al. [68] demonstrate that ambitious efficiency improvements and demand-side changes can reduce final energy demand by 40% by 2050 despite population and economic growth, achieving deep decarbonization without speculative carbon removal technologies, though this requires unprecedented societal transformation. Rogelj et al. [208] find that 1.5°C pathways limiting overshoot require emissions reductions starting now (as of 2018).

Infrastructure planning under uncertainty. Energy infrastructure has 30-50 year lifetimes, requiring planners to commit to designs based on uncertain future conditions. When scenarios systematically underestimate renewable deployment,

[3]This critique extends to any forecasting method relying on historical patterns, including the approaches employed in this thesis (discussed further in Section 6)

infrastructure planning may lag reality, which may lead to grid bottlenecks, transmission constraints, and storage inadequacies [248], further limiting the ability to accelerate renewable deployment. Conservative renewable projections justify continued fossil fuel investment, risking stranded assets if transitions accelerate faster than assumed [215]. Janipour et al. [249] document how such investments create carbon lock-in through technological incompatibility, sunk costs, and institutional inertia, making subsequent transitions to low-carbon alternatives more difficult and costly. Without probabilistic bounds, planners face a binary choice; design infrastructure for either a high or low renewable future, with enormous costs if wrong due to path dependence and lock-in effects [123].

Carbon removal reliance and transition risk. Most 1.5°C pathways in IPCC AR6 rely heavily on large-scale carbon dioxide removal (CDR) in the second half of the century to compensate for near-term overshoot [208, 250]. CDR technologies like BECCS and direct air capture remain speculative at scale, with unclear learning curves [251], vast land requirements competing with food production and biodiversity [252], and potential moral hazard effects that weaken near-term mitigation ambition [253]. Fuss et al. [254] characterize this as "betting on negative emissions", which is a high-risk strategy if CDR fails to materialize or proves more expensive than anticipated. If conservative renewable deployment assumptions prove wrong and solar/wind accelerate faster than projected, less reliance on speculative CDR may be needed, while eventually being economically cheaper if learning rates persist [30].

2.3.5 Alternative Modeling Approaches

Agent-based models. Agent-based models (ABMs) offer an alternative approach where heterogeneous actors (consumers, firms, governments) make decisions based on local information, bounded rationality, and social interactions [255]. In contrast to IAMs' representative agents and optimization, ABMs model heterogeneous preferences (e.g., early adopters vs. laggards), social networks (observational learning), and path dependence (network effects and infrastructure lock-in). Researchers have applied ABMs to residential solar adoption [256], technology diffusion [244], and policy evaluation, excelling when heterogeneity matters qualitatively and distributional impacts are important [219]. Krumm et al. [212] find that only agent-based models adequately integrate heterogeneity of actors and public ownership. Recent reviews emphasize ABMs' strength in capturing social aspects and actor heterogeneity [257, 258]. However, ABMs face challenges in validation, parameterization, data requirements, and computational scalability.

Network substitution models. A related alternative embeds S-curve dynamics directly into macroeconomic frameworks. The Future Technology Transformations (FTT) model couples logistic substitution equations with the E3ME macroeconomic model [259]. Nijsse et al. [260] apply it to show solar PV has reached self-sustaining momentum, likely to continue without additional policy support. Unlike IAMs that optimise energy mixes exogenously, FTT models generate technology trajectories endogenously. These may better capture behaviour of the full system, at the cost of producing accurate forecasts for individual technologies; and are not systematically backtested against historical diffusion data. We return to this trade-off in Sections 6.3 and 6.4.

Chapter 3

Why S-curves underestimate the future

This chapter is taken from: Wagenvoort, B. (2026). Why S-curves underestimate the future: Systematic bias in diffusion forecasting [261].

S-curves are used to model a wide range of diffusion processes from innovation adoption to disease spread. Through a simulation study, we demonstrate that asymptote and inflection point parameter estimates exhibit systematic downward bias when only a small proportion of the S-curve is observed. This bias occurs across different functional forms (logistic, Gompertz), noise models (Fisher-Pry, additive, multiplicative), and estimation procedures (OLS on Fisher-Pry transform, NLLS). These biases lead to inconsistent point forecasts that consistently underestimate long-term adoption, offering a statistical reinterpretation of Amara's law. We develop a debiasing method using parametric bootstrapping that leverages approximately unbiased growth and noise parameter estimates. The method requires knowledge of the data-generating process and scale-invariant noise structure, limiting its applicability. Applications to solar PV capacity and electric vehicle adoption suggest these technologies may sustain higher growth rates than conventional forecasts predict, though real-world complexities (policy changes, technological disruptions, market saturation) introduce uncertainties not captured by our simplified model. The results highlight the importance of accounting for estimation bias when applying conventional fitting methods to S-curve data.

Keywords: S-curves; Estimation bias; Amara's law; Diffusion forecasting

S-curves have modeled diffusion processes for over two centuries [11, 12, 44]. Despite their widespread use, S-curve forecasts frequently underestimate long-term adoption. This pattern appears in renewable energy [28, 29], electric vehicles [184, 262], and COVID-19 case projections [21, 263]. Standard estimation methods like non-linear least squares (NLLS) produce highly variable forecasts when applied to early-stage diffusion data [31].

We investigate the statistical origins of these systematic underestimates and develop a parametric bootstrap debiasing method. While our approach shows promise in controlled simulations, its real-world applicability is constrained by strong assumptions about data-generating processes. Applications to solar PV and electric vehicle adoption illustrate how accounting for estimation bias can shift technology feasibility assessments, though we emphasize the preliminary nature of these projections.

S-curve forecasts and their limitations. S-curves describe three phases: initiation, exponential growth, and saturation [43]. Applications span technology adoption [7, 36, 49, 82], infrastructure dynamics [4, 14, 20], and biological systems [59]. In epidemiology, sophisticated S-curve specifications improved COVID-19 trajectory forecasts [263], while simpler models underestimated death counts and post-peak dynamics [21].

In technology diffusion, S-curve forecasts have led to conflicting conclusions. Some analyses suggest rapid energy transitions are unlikely [69] or infeasible [66], while others document persistent underestimation of renewable energy adoption [24, 28, 29]. Baumgartner et al. [31] show that NLLS regression applied to different segments of solar PV data produces highly variable forecasts, highlighting fundamental estimation challenges.

Electric vehicles present similar forecasting difficulties. Despite doubling sales share every two years [264], projections disagree on whether growth will continue [265, 266]. Historical EV adoption has been consistently underestimated [184, 262], though whether this pattern will persist remains uncertain [267].

Statistical origins of systematic bias. Standard S-curve estimation at early diffusion stages produces systematic parameter bias. We demonstrate through simulation that asymptote and inflection point estimates are consistently downward biased, providing a quantitative explanation for why S-curve forecasts underestimate long-term adoption—a phenomenon commonly attributed to Amara's law [268]. Our work builds on prior evidence that diffusion model parameters are biased [18], extending this by characterizing bias as a function of diffusion progress and developing practical correction procedures.

We develop a parametric bootstrap debiasing procedure that exploits two empirical regularities: (1) growth and noise parameters remain approximately unbiased across diffusion stages, and (2) the Fisher-Pry noise model exhibits scale-invariance. The method improves point forecasts in controlled settings but requires strong assumptions that limit real-world applicability. We illustrate the approach with solar PV and electric vehicle data, showing how correcting for estimation bias affects technology feasibility assessments. These applications are exploratory, serving to demonstrate methodology rather than provide authoritative forecasts.

Our findings contribute to understanding why energy transition scenarios consistently underestimate renewable adoption [24, 29] and why NLLS produces unreliable forecasts from early-stage data [31]. However, we emphasize that real-world diffusion involves complexities—policy interventions, technological breakthroughs, market dynamics—that our simplified model cannot capture.

3.1 Theoretical background and model framework

3.1.1 S-curve models and diffusion dynamics

Most S-curves derive from a general differential equation [14, 60]:

$$\frac{dQ(t)}{dt} = kQ(t)^{\alpha}\left[1 - \left(\frac{Q(t)}{L}\right)^{\beta}\right]^{\gamma} \tag{3.1}$$

where $Q(t)$ represents adoption at time t, L denotes the asymptotic saturation level, k measures the intrinsic growth rate, and α, β, γ control functional form.

Symmetric versus asymmetric formulations. When $\alpha = \gamma = 1$, equation 3.1 reduces to the Bertalanffy-Richards (B-R) curve:

$$\frac{dQ(t)}{dt} = kQ(t)\left[1 - \left(\frac{Q(t)}{L}\right)^{\beta}\right] \tag{3.2}$$

with analytical solution

$$Q(t) = \frac{L}{(1 + e^{-\beta k(t-t_0)})^{1/\beta}} \tag{3.3}$$

where t_0 defines the inflection point and β determines curve symmetry [20]. Setting $\beta = 1$ yields the symmetric logistic curve:[1]

$$Q(t) = \frac{L}{1 + e^{-k(t-t_0)}}. \tag{3.4}$$

The logistic form has been applied to technology adoption [36], population dynamics [71], cell growth [72], and epidemic spreading [4, 73].

Alternative functional forms may better represent certain empirical patterns [7, 20]. The asymmetric Gompertz curve [11] is widely used in epidemiology [80], cell biology [81],

[1] See appendix 3.6.1 for derivation.

and technology adoption [69]. See appendix 3.8.1 for details.

Functional form comparison. Figure 3.1 shows B-R curves for different β values alongside the Gompertz curve. Without loss of generality, we focus on the symmetric

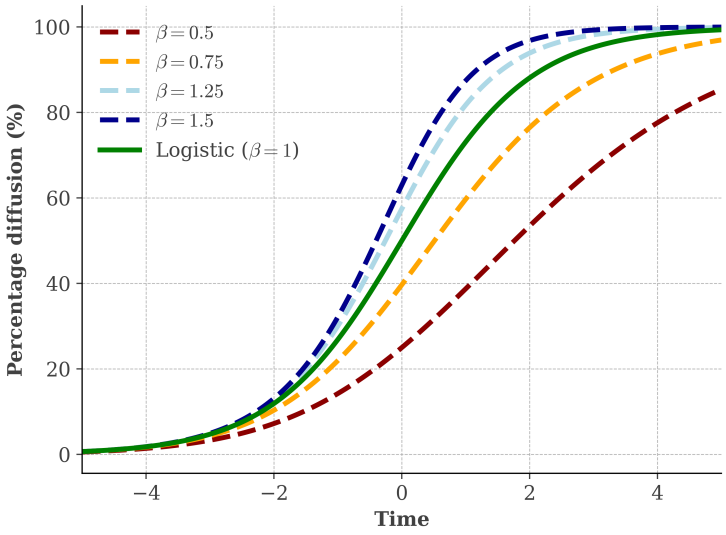


Figure 3.1: Bertalanffy-Richards S-curves with $L = 1$, $k = 1$, $t_0 = 0$ for varying β. All curves exhibit similar exponential behavior at early stages. Inflection points occur beyond 50% diffusion when $\beta > 1$ and before 50% when $\beta < 1$.

logistic curve. All findings extend to asymmetric forms such as the Gompertz curve (appendix 3.8.1).

Fisher-Pry transformation and parameter estimation. Fisher and Pry [37] introduced a linearizing transformation for S-curve estimation. We consider additive Gaussian noise under the Fisher-Pry (FP) transform of the logistic curve:

$$Y_t = \frac{L}{1 + e^{-k(t-t_0)-\sigma\epsilon_t}} \tag{3.5}$$

where $\epsilon_t \sim N(0, 1)$ and σ is the shock intensity (noise parameter). This noise model exhibits scale-invariance—the proportional error $\sigma\epsilon_t$ remains constant across diffusion stages—which proves essential for our debiasing method (Section 3). While we acknowledge that this specification makes strong assumptions about real-world noise structures, it

allows analytical tractability while capturing qualitatively important features observed in alternative noise models (appendices 3.8.3, 3.8.4).

Applying the Fisher-Pry substitution $f_t := -\log(L/Y_t - 1)$ yields:

$$f_t = k(t - t_0) + \sigma\epsilon_t. \tag{3.6}$$

This linearized form enables parameter estimation via ordinary least squares (OLS) regression once L is known. However, when L is unknown—as in forecasting applications—we must jointly estimate all parameters $\theta = \{L, k, t_0, \sigma\}$ by minimizing:

$$\hat{\theta} = \arg\min_{L,k,t_0,\sigma} \sum_{t \in T} [f_t(L) - k(t - t_0)]^2 \tag{3.7}$$

where T denotes the observation window and $f_t(L) = -\log(L/Y_t - 1)$.[2] This optimization is well-specified for the noise model in equation 3.5, yet as we demonstrate in Section 4, it produces systematically biased parameter estimates when only early-stage diffusion data are observed. At very early diffusion stages, the optimization landscape exhibits a flat ridge where multiple parameter combinations produce nearly identical objective function values (appendix 3.6.5), contributing to high parameter uncertainty.

3.1.2 Theoretical framework for parameter bias

Understanding why bias occurs requires examining the interplay between observable data and parameter identification. When fitting S-curves to early-stage data, the optimization in equation 3.7 faces a fundamental challenge: multiple parameter combinations can produce similar fit quality over limited observation windows, yet these combinations imply dramatically different long-term forecasts.

[2] See appendix 3.6.2 for analytical expressions.

The identification problem. At early diffusion stages, only the initial exponential growth phase is visible. Data cannot distinguish between a small asymptote with fast growth versus a large asymptote with slow growth—both produce similar curvature over short horizons. This identification problem becomes more severe as observation windows shrink. The asymptote L determines the eventual saturation level but affects near-term fit quality weakly, leading to large uncertainty.

In contrast, the growth rate k is identified primarily from the slope of the linearized Fisher-Pry transform (eq. 3.6), which early-stage data can estimate reliably. Similarly, the noise parameter σ is determined by residual variance, which accumulates quickly even with limited observations. This differential identification explains why k and σ remain approximately unbiased while L and t_0 exhibit substantial bias (Results, Section 4.1).

Parameter coupling in joint estimation. The Fisher-Pry transformation $f_t(L) = -\log(L/Y_t - 1)$ creates nonlinear coupling among parameters. When the optimization algorithm explores parameter space, underestimating L makes the denominator $L/Y_t - 1$ smaller, which increases f_t for given data Y_t. To maintain fit quality, the optimization compensates by adjusting k upward and t_0 downward, preserving the linear relationship $f_t \approx k(t - t_0)$ over the observed window.

This coupling means that while individual parameter estimates may be biased, their joint behavior produces reasonable in-sample fit. The bias manifests primarily in out-of-sample forecasts, where underestimated asymptotes lead to systematic underestimation of long-term adoption. Appendix 3.6.3 provides analytical characterization of these parameter relationships.

Scale-invariance and its implications. The Fisher-Pry noise model (eq. 3.5) possesses a crucial scale-invariance property: multiplying both L and Y_t by a constant factor leaves the Fisher-Pry transform $f_t(L) = -\log(L/Y_t - 1)$ unchanged. This implies

that bias patterns depend on the relative position within the S-curve (measured by $Q(t)/L$) rather than absolute values of L or Y_t.

This scale-invariance enables our debiasing methodology (Methods, Section 3.2). We can characterize bias using simulations with arbitrary asymptote values, provided we match the growth rate k, noise level σ, and observation window. This property fails for additive noise models $Y_t = Q(t) + \sigma\epsilon_t$, where absolute scale matters, explaining why our debiasing method cannot be straightforwardly extended to such specifications.

3.2 Methods

3.2.1 Simulation design

S-curves model diffusion trajectories well when fitted after the inflection point [269]. Van den Bulte and Lilien [18] documented that parameter estimates from S-curve models can exhibit systematic bias when applied to limited early-stage data. To characterize this bias systematically and develop correction procedures, we simulate diffusion processes with known parameters and estimate from finite samples. We generate synthetic time series following the Fisher-Pry data-generating process (eq. 3.6), which has additive Gaussian noise in the logit transformed space,

$$f_t = k(t - t_0) + \sigma\epsilon_t, \quad \epsilon_t \sim N(0, 1) \tag{3.8}$$

where the observed data in levels is $Y_t = L/(1+\exp(-f_t))$. We use true parameters $L = 10^4$, $k = 0.3$ year^{-1}, $t_0 = 0$, noise level $\sigma = 0.1$, and sample size $m = 50$ observations at annual frequency from $t = -25$ to $t = 24$. These baseline parameters represent moderate-speed diffusion over a 50-year observation window (see appendix 3.7.1 for complete configuration details). To examine real-world data addition scenarios where new observations arrive

sequentially (e.g., annual updates), we also analyze bias patterns when keeping the starting point constant while adding data forward in time (appendix 3.7.4), finding similar bias patterns.

For each replication, we estimate parameters via the optimization in equation 3.7. We analyze bias as a function of diffusion progress by computing estimates across observation windows ending at different diffusion levels: $Q(t_m)/L \in [0.01, 0.95]$. Since the logistic function depends only on normalized time $(t - t_0)$ and diffusion proportion Y/L, bias patterns are determined by the observation window relative to the inflection point rather than absolute parameter values (appendix 3.6.4). This property allows characterizing bias using arbitrary (L, t_0) values. We conduct 10,000 replications per window to precisely estimate bias and variance.

Section 4 presents three main sets of results: (1) core bias patterns under well-specified Fisher-Pry estimation with the baseline parameters above (Section 4.1), (2) model misspecification when non-linear least squares (NLLS) is applied to Fisher-Pry data (Section 4.2), and (3) robustness of bias patterns across alternative noise structures—Fisher-Pry, additive, and multiplicative—demonstrating that systematic bias arises from S-curve geometry rather than specific model assumptions (Section 4.3). Additional robustness checks appear in the appendix: sensitivity to noise levels ($\sigma \in \{0.05, 0.10, 0.20\}$, appendix 3.11.1), sample sizes ($m \in \{20, 50, 100\}$), and alternative functional forms (Gompertz, appendix 3.8.1), with comprehensive NLLS misspecification analysis (appendix 3.9.1). We use $k = 0.3$ as a representative moderate growth rate; while varying k primarily changes which diffusion level is observed at fixed times, additional k-dependence may exist through curvature and discretization effects (appendix 3.7.2).

3.2.2 Debiasing algorithm

The systematic bias documented in Section 4 follows predictable patterns that enable correction. We develop a parametric bootstrap debiasing procedure that exploits two key empirical regularities: (1) growth and noise parameters remain approximately unbiased across diffusion stages, and (2) bias in the asymptote depends primarily on the observation window rather than specific parameter values. Algorithm 1 in appendix 3.10.1 provides the complete pseudocode; appendix 3.10.1.2 details the simulation configuration; appendix 3.10.3 contains full mathematical proofs.

Our debiasing procedure operates in four stages. Given observed data $\{Y_t\}_{t\in T}$, we first estimate $\hat{\theta}^{(b)} = \{\hat{L}^{(b)}, \hat{k}^{(b)}, \hat{t}_0^{(b)}, \hat{\sigma}^{(b)}\}$ using equation 3.7. We then use parametric bootstrapping [270] to generate surrogate data with $k^{(s)} = \hat{k}^{(b)}$ and $\sigma^{(s)} = \hat{\sigma}^{(b)}$ across different observation windows $d \in \mathcal{D}$. Notably, we can use arbitrary fixed surrogate parameters for the asymptote and inflection point, i.e., $(L^{(s)}, t_0^{(s)}) = (10^4, 0)$,[3] as the Fisher-Pry noise model is scale-invariant.

For each window d and simulation $i \in \{1, \ldots, N\}$, we generate surrogate data in the Fisher-Pry transform space:

$$f_t^{(s,i,d)} = k^{(s)}(t - t_0^{(s)}) + \sigma^{(s)}\epsilon_t^{(s,i,d)}, \quad \epsilon_t^{(s,i,d)} \sim N(0,1) \tag{3.9}$$

with corresponding levels $Y_t^{(s,i,d)} = L^{(s)}/(1 + \exp(-f_t^{(s,i,d)}))$. We then estimate $\hat{\theta}^{(s,i,d)}$ via equation 3.7. We compute the mean bias factor for each window:

$$\beta_L(d) = \mathbb{E}_i\left[\frac{\hat{L}^{(s,i,d)}}{L^{(s)}}\right] \tag{3.10}$$

and the mean inflection point estimate $\bar{t}^{(s,d)} = \mathbb{E}_i[\hat{t}_0^{(s,i,d)}]$.

Third, we identify which observation window the original data corresponds to by

[3] See appendix 3.10.3 for mathematical justification of why fixed surrogate parameters work.

computing the relative time to estimated inflection point: $t^{(b)} = t_m - \hat{t}_0^{(b)}$, where t_m denotes the final observation time. We match this to the surrogate window that minimizes $|t^{(b)} - \bar{t}^{(s,d)}|$, denoted d^*. This matching exploits the systematic relationship between $t^{(b)}$ and the observation window.

Finally, we debias the asymptote estimate as $\hat{L} = \hat{L}^{(b)}/\beta_L(d^*)$ and re-estimate remaining parameters by OLS on the Fisher-Pry transform $f_t(\hat{L}) = k(t - t_0) + \sigma\epsilon_t$, conditional on the debiased $\hat{L}$. This secondary estimation step ensures consistency among all parameters and leverages the conditional unbiasedness of k and t_0 given a correct asymptote estimate.

Theoretical foundations. Three properties underpin the debiasing procedure's validity. We provide intuitive explanations here; appendix 3.10.3 contains complete mathematical derivations.

Property 1: Conditional unbiasedness of k and σ. Given correct asymptote L, OLS on the Fisher-Pry transform (eq. 3.6) yields approximately unbiased estimates: $\mathbb{E}[\hat{k}|L] = k$, $\mathbb{E}[\hat{\sigma}|L] = \sigma$. Joint estimation via equation 3.7 couples parameters—when $\hat{L}$ is biased downward, optimization compensates by adjusting $\hat{k}$ and $\hat{t}_0$ (appendix 3.6.3). However, empirical evidence (Section 4.1) shows $\hat{k}$ and $\hat{\sigma}$ exhibit minimal bias across diffusion stages because they are determined primarily by slope and residual variance, which early-stage data estimate reliably despite asymptote uncertainty.

Property 2: Scale-invariance of Fisher-Pry noise. The bias factor $\beta_L(d) = \mathbb{E}[\hat{L}/L]$ at window d depends only on (k, σ, d), not on specific (L, t_0) values. This allows characterizing bias via surrogate simulations with arbitrary $(L^{(s)}, t_0^{(s)})$ provided we use estimated $(\hat{k}^{(b)}, \hat{\sigma}^{(b)})$. The Fisher-Pry transform's ratio structure L/Y_t creates scale-invariance in L, while t_0 shifts only translate time without changing relative S-curve positioning when matching windows via $t^{(b)} = t_m - \hat{t}_0^{(b)}$. Additive noise models lack this property, preventing straightforward method extension.

Property 3: Window identification via inflection point. The inflection point estimate $\hat{t}_0$ exhibits systematic bias depending on observed S-curve proportion. Computing $t^{(b)} = t_m - \hat{t}_0^{(b)}$ provides a sufficient statistic for bias regime identification without knowing true t_0. At early diffusion (5%), $\hat{t}_0 \ll t_0$ yields large positive $t^{(b)}$; as more curve is observed, $\hat{t}_0 \to t_0$ and $t^{(b)} \to 0$ (data ending near inflection) or negative (past inflection). Matching $t^{(b)}$ to surrogate-based $\bar{t}^{(s,d)}$ identifies window d^* and corresponding correction factor $\beta_L(d^*)$, proving robust even with noisy estimates.

3.2.3 Assumptions and limitations

The debiasing procedure requires approximately unbiased estimates of growth (k) and noise (σ) parameters. While this assumption holds for moderate noise levels and sufficient data (see Results, Section 4.1), it fails when high noise combines with sparse data. Combinations like $\sigma = 0.2$ with $m = 20$ produce non-negligible bias in k and σ (see Results, Section 4.3). Higher noise levels degrade initial parameter estimates (appendix 3.7.2), which propagate through the debiasing procedure.

When optimization converges near parameter space boundaries, the procedure may incorrectly adjust estimates away from true values. Performance can degrade at higher noise levels (appendix 3.11.2.2).

The Fisher-Pry noise model's scale-invariance property (appendix 3.10.3) is essential but restrictive. Alternative noise structures, such as additive noise $Y_t = Q(t) + \sigma\epsilon_t$, lack scale-invariance, creating a circularity: generating appropriate surrogate data requires knowing the true L we aim to estimate. Moreover, additive noise produces extreme variance at early diffusion (appendix 3.8.3), where median estimates remain downward biased while outliers drive mean estimates upward. Since debiasing uses the mean bias factor $\beta_L = \mathbb{E}[\hat{L}^{(s)}/L^{(s)}]$, these outliers may cause overcorrection, worsening rather than improving estimates.

3.3 Results

S-curve parameters are interdependent, such that estimation errors propagate across parameters.[4] For example, underestimating the asymptote typically induces underestimation of the inflection point. This section demonstrates how asymptote and inflection point bias depends primarily on the proportion of the S-curve observed, while growth and noise parameter uncertainty decreases primarily with the number of data points. Through simulations, we show that asymptote and inflection point parameters exhibit systematic downward bias at early diffusion stages, while growth and noise parameters remain approximately unbiased—a finding that persists across functional forms (logistic, Gompertz), noise specifications (Fisher-Pry, additive, multiplicative), and estimation procedures (well-specified OLS on Fisher-Pry transformation, misspecified NLLS).

3.3.1 Bias patterns: well-specified OLS versus misspecified NLLS

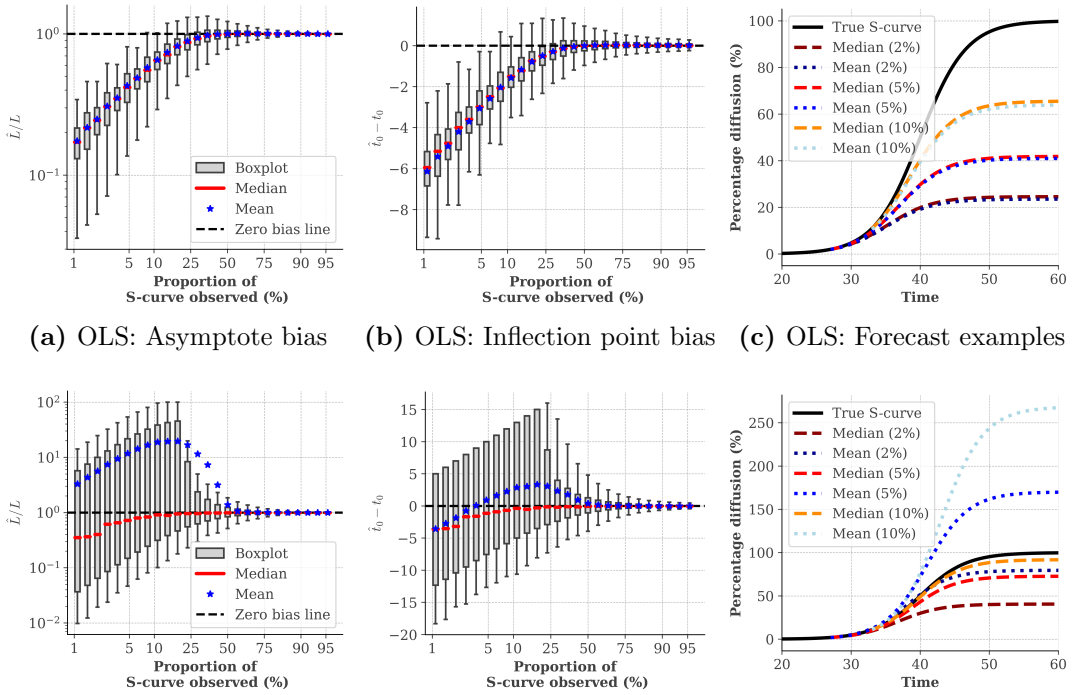
We establish bias patterns by comparing well-specified Fisher-Pry estimation (OLS on logit transformation) with misspecified non-linear least squares (NLLS) [134] applied to Fisher-Pry noise data. Using 10,000 simulations with $m = 50$ data points and true parameters $L = 10^4$, $k = 0.3$, $t_0 = 0$, $\sigma = 0.1$, Figure 3.2 shows parameter bias and forecast implications.

Both methods produce downward bias in asymptote and inflection point parameters at early diffusion. At 5% diffusion, median $\hat{L} \approx 0.4L$ for OLS and $\hat{L} \approx 0.7L$ NLLS (Figures 3.2a, 3.2d), and median $\hat{t}_0 \approx t_0 - 3$ time-units (Figure 3.2b) and $\hat{t}_0 \approx t_0 - 1$ (Figure 3.2e). Bias persists until 30-40% diffusion is observed. NLLS exhibits extreme variance. OLS produces IQR $[0.2L, 0.55L]$ while NLLS shows IQR $[0.2L, 20L]$. This model misspecification

[4] See appendix 3.6.3 for theoretical analysis of parameter relationships.

makes NLLS unreliable for early-stage forecasting despite similar median bias. NLLS also performs poorly when applied to other noise structures—for misspecification on multiplicative noise data, see appendix 3.9.1.

In contrast, growth parameter k and noise parameter σ remain approximately unbiased across all diffusion levels under well-specified estimation.[5] This differential behavior is fundamental: k and σ are identifiable from local curve properties (steepness and residual variance), while L and t_0 depend on observing the asymptotic regime.



(a) OLS: Asymptote bias **(b)** OLS: Inflection point bias **(c)** OLS: Forecast examples

(d) NLLS: Asymptote bias **(e)** NLLS: Inflection point bias **(f)** NLLS: forecast examples

Figure 3.2: Parameter bias and forecast implications: well-specified OLS (Fisher-Pry, top row) versus misspecified NLLS (bottom row) applied to Fisher-Pry noise data. 10,000 simulations per condition, $m = 50$, $L = 10^4$, $k = 0.3$, $t_0 = 0$, $\sigma = 0.1$. Red: median; black: zero bias; blue: mean. **(a,d)** Bias in asymptote estimates. Both produce biased median at 5% forecast origin, but NLLS exhibits IQR $[0.20L, 20L]$ versus OLS on FP transform $[0.20L, 0.55L]$—over 100-fold wider variance. **(b,e)** Inflection point: Both exhibit bias in median but NLLS spans $[-7, 10]$ versus OLS $[-4, -2]$. **(c,f)** Forecasts: OLS shows systematic underestimation (median and mean similar); NLLS mean forecasts severely overestimate due to occasional extreme parameter values.

At 5% diffusion, true t_0 occurs 10 time-units after the last observation. Median estimate

[5] Sensitivity to noise levels (varied via $\sigma \in \{0.05, 0.10, 0.20\}$ for fixed $k = 0.3$) and sample sizes ($m \in \{20, 50, 100\}$) appears in appendix 3.11.1, showing bias patterns persist at higher noise levels with increased variance, and variance decreases with more data points as expected. Appendix 3.11.2 presents detailed box plots showing how bias distributions change with σ.

$\mathbb{P}_{50}[\hat{t}_0] \approx t_0 - 3$. places the inflection point 7 units ahead instead of 10, implying it occurs at $\approx 32\%$ diffusion instead of 50%. Combined with $\mathbb{P}_{50}[\hat{L}] \approx 0.4L$, the 3-year underestimate of inflection timing translates to nearly 3 years less exponential growth, producing forecasts that track reality only to $\approx 20\%$ diffusion before diverging (Figure 3.2c, red and blue lines (middle set)). This provides statistical basis for Amara's law [268]. Systematic estimation bias causes forecasts to underestimate long-term diffusion when fitting early-stage data.

The extreme variance under NLLS manifests in forecasts (Figure 3.2f). Table 3.1 quantifies this effect, comparing variance and bias under OLS and NLLS for the L asymptote.

3.3.2 Quantifying NLLS misspecification effects

Table 3.1 quantifies the variance under NLLS misspecification. At 5% diffusion, OLS and NLLS produce median $\hat{L} \approx 0.4L$ and $\hat{L} \approx 0.7L$. Notably, the NLLS median estimate is slightly less biased than OLS at this diffusion level, but the normalized standard deviation is more than 50-fold higher (0.35 vs 19.80).

Method	**Median $\hat{L}/L$**	**IQR $\hat{L}/L$**	**IQR Width**
Well-specified OLS (FP)	0.4	[0.20, 0.55]	0.35
Misspecified NLLS	0.7	[0.20, 20]	19.80
Variance ratio	-	—	$> 50\times$

Table 3.1: Asymptote estimates at 5% diffusion: well-specified OLS versus misspecified NLLS. Both underestimate in median; NLLS overestimates in mean. NLLS over 50-fold wider interquartile range due to model misspecification (visible in Figures 3.2a vs 3.2d). NLLS assumes homoskedastic errors in levels but Fisher-Pry data have homoskedastic errors in logit space—this misspecification produces extreme variance while preserving median bias.

The over 50-fold variance ratio has critical forecasting implications (Figures 3.2c vs 3.2f). Under well-specified OLS, median and mean forecasts behave similarly (both underestimate due to systematic parameter bias). Under NLLS, median forecasts underestimate while mean forecasts severely overestimate because occasional extreme parameter estimates (e.g., $\hat{L} = 50L$) dominate the average. This offers an explanation for Baumgartner et al.'s [31]

finding that NLLS applied to different segments of solar PV data produces highly variable forecasts, and often underestimates long-term diffusion (median is downward biased at early diffusion).

3.3.3 Robustness across noise models and functional forms

Bias patterns persist across alternative noise specifications. Figure 3.3 shows additive and multiplicative noise models under well-specified estimation produce downward bias in L and t_0 comparable to Fisher-Pry results (Figure 3.2), persisting to 30-40% diffusion.[6] This robustness indicates the bias arises from fundamental S-curve geometry and observational truncation rather than specific noise structure. Growth and noise parameters remain approximately unbiased across all specifications.[7]

Alternative functional forms show similar patterns. Gompertz curves (asymmetric S-curves with inflection at 37% rather than 50% of asymptote) exhibit a similar bias relationship. The debiasing procedure can be extended to Gompertz curves with Fisher-Pry noise (appendix 3.8.1, 3.10.4).

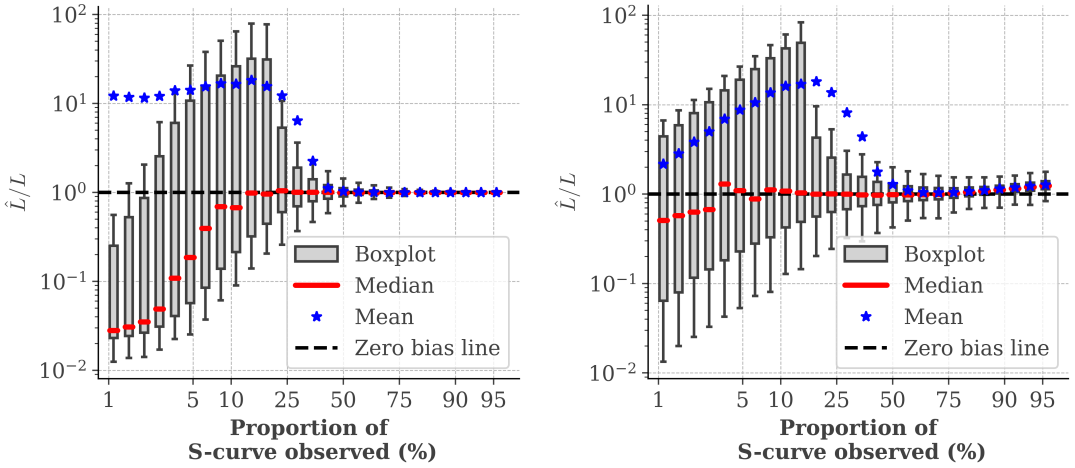
3.3.4 Debiasing procedure performance

We develop a parametric bootstrap debiasing method exploiting two properties: (1) growth and noise parameters remain approximately unbiased at early diffusion, and (2) Fisher-Pry noise exhibits scale-invariance under the logit transform (appendix 3.10.3). The procedure simulates surrogate datasets using initial $\hat{k}$ and $\hat{\sigma}$ estimates, calculates empirical bias across diffusion levels, then corrects original estimates.

Figure 3.4 shows debiasing performance across diffusion levels. All parameters achieve

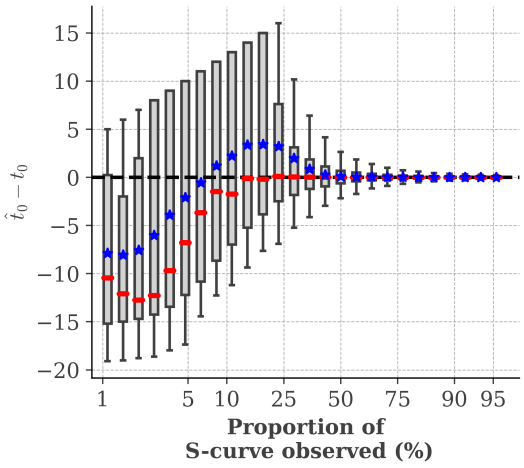
[6] Additive noise: $Y(t) = L/(1+e^{-k(t-t_0)}) + \sigma\epsilon_t$ estimated via NLLS with $\sigma = 100$. Multiplicative noise: $Y(t) = Le^{\sigma\epsilon_t}/(1+e^{-k(t-t_0)})$ estimated via NLLS on $\log Y(t)$ with $\sigma = 0.25$. See appendix 3.8 for detailed noise scaling explanation. Full model specifications and growth/noise parameter bias analysis in appendices 3.8.3-3.8.4.

[7] The debiasing procedure can be extended to multiplicative noise models but not to additive noise models, due to scale invariance requirements (appendix 3.10.4).

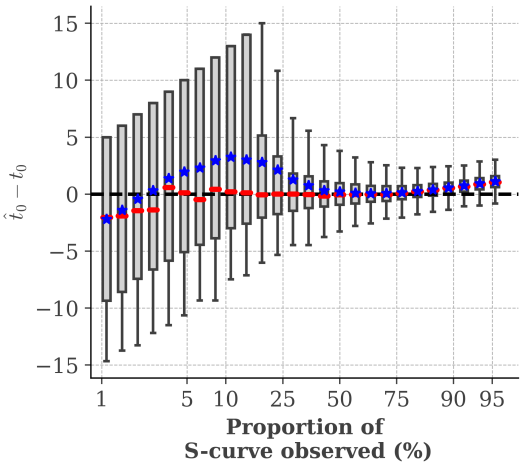


(a) Asymptote bias (additive)

(b) Asymptote bias (multiplicative)



(c) Inflection point bias (additive)

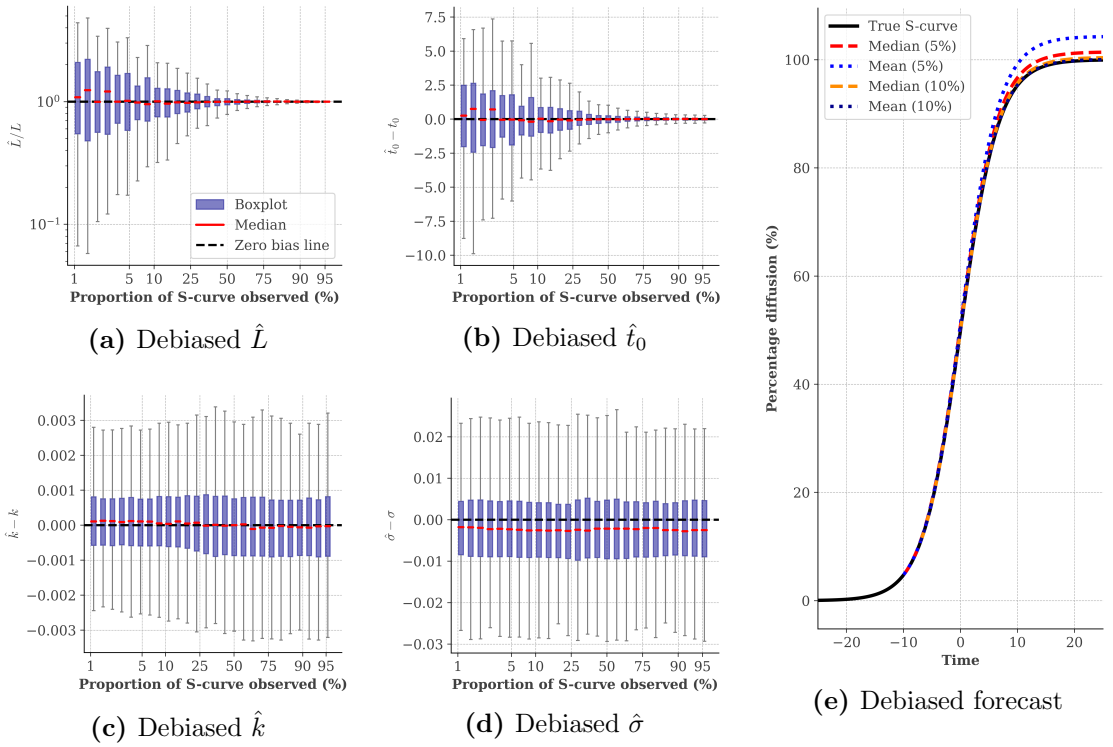


(d) Inflection point bias (multiplicative)

Figure 3.3: Median parameter bias under well-specified estimation for additive ($\sigma = 100$) and multiplicative ($\sigma = 0.25$) noise models. Both produce substantial downward bias in **(a-b)** asymptote and **(c-d)** inflection point at early diffusion stages, with magnitude comparable to Fisher-Pry results ($\sigma = 0.1$, Figure 3.2). Each line represents median of 10,000 simulations with $m = 50$ data points, $L = 10^4$, $k = 0.3$, $t_0 = 40$. Noise parameters calibrated to produce comparable noise magnitude at 10% diffusion (see appendix 3.8). Growth and noise parameters remain approximately unbiased across all specifications. See appendices 3.8.3-3.8.4 for comprehensive analysis including variance, all parameters, and NLLS performance.

approximate median unbiasedness, though asymptote and inflection point bias varies slightly at very early diffusion (1-5%). At 5% diffusion, debiased median $\mathbb{P}_{50}[\hat{L}] \approx L$ versus $0.4L$ before debiasing, with IQR widening from approximately $[0.2L, 0.55L]$ to approximately $[0.6L, 1.25L]$. Median and mean forecasts track the true S-curve well (Figure

3.4e), after debiasing.[8] The method requires correct specification of the noise-generating process and scale-invariant noise structure, limiting applicability to cases where these assumptions hold.



(a) Debiased $\hat{L}$ **(b)** Debiased $\hat{t}_0$

(c) Debiased $\hat{k}$ **(d)** Debiased $\hat{\sigma}$ **(e)** Debiased forecast

Figure 3.4: Parameter estimates after debiasing procedure. 1,000 simulations, $m = 50$ data points, $L = 10^4$, $k = 0.3$, $t_0 = 0$, $\sigma = 0.1$. **(a-b)** Asymptote and inflection point achieve median unbiasedness across all diffusion levels, with increased variance at early stages. **(c-d)** Growth and noise parameters maintain approximate unbiasedness. **(e)** Median (red) and mean (blue) forecasts track true S-curve (black) well.

3.3.5 Applications to renewable energy and electric vehicles

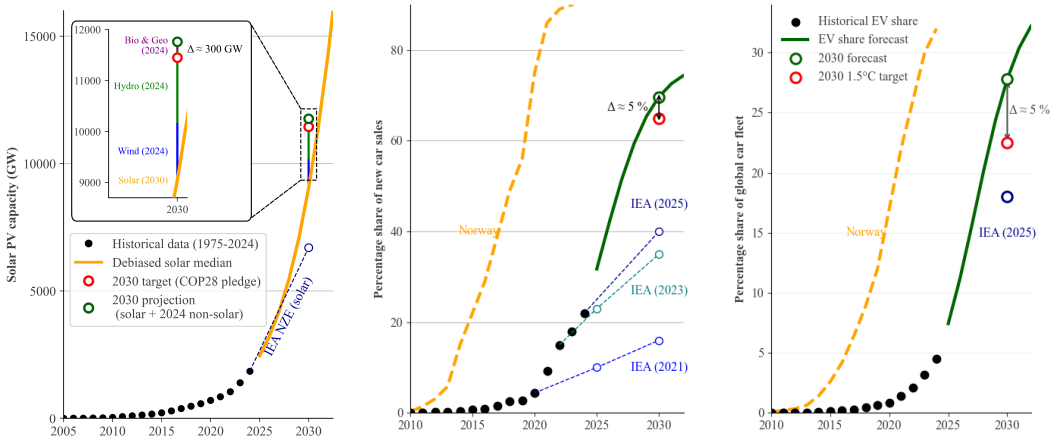
Renewable energy forecasts have consistently underestimated actual deployment [24, 28, 29]. Recent projections indicate that the COP28 pledge of tripling renewable capacity by 2030 is not achievable [66]. The International Energy Agency (IEA) projected 6,699 GW of solar PV capacity by 2030 in its 2023 Net Zero Scenario [271] (approximately 3-4 times 2024

[8] Performance degrades at higher noise levels ($\sigma = 0.2$ with $k = 0.3$) due to increased variance in initial $\hat{k}$ and $\hat{\sigma}$ estimates. See appendix 3.11.2.2 for detailed analysis. Appendix 3.10.2 demonstrates that when true k and σ are known (rather than estimated), the procedure achieves improved debiasing even at $\sigma = 0.2$ across all diffusion levels. This establishes that the limitation stems from estimation uncertainty in these parameters, not fundamental method failure.

levels). Similarly, the IEA EV projections have required repeated upward revisions, with their 2023 forecast of 23% global sales share in 2024 achieved one year early [262, 264, 265].

EV adoption has accelerated rapidly, though projections disagree on sustainability [266, 272]. Challenges include charging infrastructure gaps, battery supply constraints, and supply chain bottlenecks [273, 274]. Standard projections suggest EV share of sales may fall short of IEA's 2030 net-zero target (65% of new sales) [265], while recent analyses project 62-80% by 2030 [267, 272].

We apply our debiasing method to solar PV capacity and EV adoption data (Figure 3.5), illustrating how correcting for estimation bias affects technology feasibility assessments. These applications demonstrate the methodology rather than provide definitive forecasts.



(a) Solar PV capacity **(b)** EV share of new sales **(c)** EV share of fleet

Figure 3.5: Debiased projections for solar PV and electric vehicles. **(a)** Solar PV capacity could reach over 9,000 GW by 2030 (5× 2024 levels). Combined with other renewables at 2024 capacity (2,726 GW), the total would exceed the COP28 tripling target (11,450 GW) by 311 GW. Data: Solar capacity 1975-2024 [30, 275, 276], Wind-capacity [277], geothermal [278], bio-energy [279] and hydropower capacity [280]. **(b)** EV new sales could reach ∼70% by 2030, exceeding IEA net-zero target (65%). **(c)** EV fleet could reach >25% by 2030, versus IEA 22%. Norway's trajectory demonstrates feasibility but required extensive policy support. Data: 2010-2024 [264].

Table 3.2 compares debiased projections with IEA scenarios and historical precedents. Solar PV projections using 1975-2024 data suggest capacity could quintuple by 2030, supporting the successful achievement of the COP28 tripling targets [281] when combined with other renewables. EV projections using 2010-2024 data suggest sales share could reach

70% by 2030, exceeding net-zero scenarios. Norway's rapid EV adoption (22% to 80% sales share, 2015-2021)[264] demonstrates feasibility under strong policy support, though conditions may not generalize globally. Similarly, the proportion of EV in the total fleet could surpass 25% by 2030, exceeding IEA's 22% target [265]. The debiased projection suggests a more optimistic outlook for renewable energy and EV adoption than standard estimates, with EV projections aligning with recent analyses [267, 272], following Norway's precedent [264].

Technology	2024 Level	Debiased 2030	IEA Target	Assessment
Renewable Capacity (GW)				
Solar PV	1,852	9,035	6,699	Exceed by 2,336
Other renewables[b]	2,726	2,726	—	—
Total renewables	**4,578**	**11,761**	**11,450**[a]	**Exceeds by 311**
Electric Vehicles				
Sales share (%)	~ 20	$\sim$70	65[c]	Exceeds net-zero
Fleet share (%)	~ 5	>25	22[d]	Above scenarios

Table 3.2: Debiased 2030 projections compared to IEA scenarios. [a]COP28 target: triple 2022 renewables (11,450 GW total) [281]. [b]Wind (1,133 GW), hydropower (1,427 GW), bioenergy (151 GW), geothermal (15 GW) held at 2024 levels—conservative assumption as wind and others may also grow. [c]Net Zero Scenario target [265]. [d] Net Zero (22%) [265]. Norway achieved 80% sales share (2021) from 22% (2015), demonstrating feasibility under strong policy support. Data sources: solar [30, 275, 276], wind [277], hydropower [280], geothermal [278], bioenergy [279], EV [264].

The empirical data present substantial challenges for validation. Solar PV uses 50 annual observations (1975-2024) with estimated noise $\hat{\sigma} \approx 0.52$. Sensitivity analysis (appendix 3.11.1) shows that at $\sigma = 0.5$ bias patterns and variance patterns change. In this regime, appendix 3.11.2.2 shows debiasing performance degrades (but generally still reduces bias) at low diffusion levels despite adequate sample size for higher noise levels. Hence, these projections carry considerable uncertainty. Nonetheless, these projections provide useful guidance for a 6-year forecast horizon (2024-2030) as is our interest here, indicating solar is on a continued rapid growth trajectory for the near future. Furthermore, it highlights the risk of underestimation using S-curve models and the need for more robust forecasting methods to mitigate this risk.

EV sales share and fleet share use only 15 observations (2010-2024) with $\hat{\sigma} \approx 0.42$ and $\hat{\sigma} \approx 0.36$, respectively. The combination of limited sample size ($m = 15$ for EV) and high noise levels ($\hat{\sigma} \approx 0.4 - 0.5$) makes it difficult to determine the true validity of the debiasing method for these applications. Furrthmore, substantial uncertainty from policy shifts, technological disruptions, supply chain constraints, grid integration challenges, and validity of Fisher-Pry noise assumptions for real data may limit accuracy. These projections illustrate methodology under favorable trend continuation.

3.4 Discussion

Our findings reveal a fundamental statistical mechanism underlying persistent pessimism in technology forecasting. The systematic underestimation of long-term adoption in S-curves may emerge from inherent properties of nonlinear parameter estimation when applied to incomplete diffusion trajectories. This has important implications for how we interpret historical forecasting failures and structure future projection efforts.

Statistical reinterpretation of Amara's law. Standard S-curve estimation at early diffusion produces forecasts that slightly overestimate near-term adoption while substantially underestimating long-term potential, matching Amara's law [268]. This pattern emerges from the statistical properties of nonlinear estimation: upward-biased growth estimates combined with downward-biased asymptote and inflection point estimates. The effect persists across well-specified and misspecified models, suggesting forecasting failures stem partly from fundamental estimation challenges rather than solely inadequate models. These statistical biases may help explain historical scientific pessimism about long-run technology adoption [282], complementing cognitive explanations of Amara's law. This helps explain persistent underestimation of renewable energy and EV adoption [24, 29, 184].

Implications for technology governance and policy. The documented bias has concrete implications for technology governance and investment. Way et al. [29] demonstrate that IEA forecasts have consistently underestimated solar and wind adoption, leading to policy recommendations favoring slower transitions. Similarly, Gilbert et al. [217] show that US projections were consistently pessimistic for wind electricity adoption. Our analysis provides a methodological explanation for these systematic errors: standard S-curve estimation produces downward-biased forecasts when applied to early-stage S-curve diffusion processes, such as renewable technologies. When such forecasts inform long-lived infrastructure decisions (power plants, transmission lines, refining capacity), initial underestimation creates path dependencies that slow adoption rates [283]. Thus, accounting for estimation bias is beneficial when evaluating transition scenarios. This matters for climate policy adequacy assessments, where the difference between pessimistic and bias-corrected projections affects conclusions about whether current policies suffice to meet climate targets.

Limitations and applicability. These results should *inform* rather than *determine* policy and investment decisions. The debiasing method requires three conditions: (1) approximately unbiased estimates of k and σ, (2) scale-invariant noise structure (such as additive noise under the FP transform), and (3) data following S-curve dynamics without structural breaks. The method may fail to fully eliminate bias when high noise is combined with sparse data. For example, this occurs for combinations like $\sigma = 0.2$ with $m = 50$ (appendix 3.11.2.2).

The scale-invariance property (appendix 3.10.3) of the Fisher-Pry noise model is essential for the debiasing procedure, but restrictive. Alternative specifications like additive noise $Y_t = Q(t) + \sigma\epsilon_t$ lack scale-invariance, making the debiasing method circular. In such a case, generating appropriate surrogate data requires knowing the true L we aim to

estimate. Moreover, additive noise produces extreme variance at early diffusion (appendix 3.8.3). While we examine multiple noise specifications (appendices 3.8.1, 3.8.3, 3.8.4), non-Gaussian distributions may exhibit different patterns. Real-world applications assume data-generating processes resemble our simulated model and involve policy interventions, supply chain constraints, and market dynamics absent from this framework. *Our applications demonstrate methodology rather than provide definitive forecasts.* When these conditions hold, noise remains moderate, and there are sufficient data-points, accounting for estimation bias can meaningfully improve feasibility assessments.

Future research directions. Several extensions would strengthen this work. Fully characterizing debiasing feasibility as a function of noise level (σ), sample size (m), and growth parameter (k) would establish practical applicability boundaries. Developing robust diagnostics to identify when debiasing assumptions fail would enable practitioners to assess applicability to specific datasets. Exploring bias patterns under non-Gaussian noise distributions and extending the framework to non-scale-invariant models would broaden practical utility. Investigating how external factors, such as policy interventions or market dynamics, interact with statistical bias would better connect our simplified framework to real-world complexities. Comparing debiasing performance against alternative forecasting methods (Bayesian approaches, ensemble models, machine learning) would contextualize its relative strengths, while systematic application to historical datasets where true saturation levels are known would validate real-world performance.

More broadly, improving S-curve forecasting requires addressing both statistical and structural challenges. Statistical improvements should focus on more robust forecast techniques for S-curves [31]. Structural improvements include incorporating economic constraints, infrastructure bottlenecks, or modeling policy shifts, which could help forecast technological diffusion in the real world, where the curve may deviate from the idealized

S-curve. Neither approach alone suffices; progress requires integrating statistical rigor with domain expertise about specific technologies and markets.

3.5 Conclusions

This paper demonstrates that systematic bias in S-curve estimations and projections is not an artifact of model misspecification but a fundamental property of nonlinear parameter estimation applied to incomplete diffusion trajectories. Asymptote and inflection point parameters are consistently underestimated across functional forms and noise specifications, creating forecasts that underestimate long-term adoption. Our analysis offers a statistical reinterpretation of Amara's law, offering an explanation for historically pessimistic technology projections.

These findings have broad implications for technology forecasting and policy analysis. Any application fitting sigmoid curves to early-stage diffusion, ranging from energy transitions, AI adoption, epidemic spreading, faces this systematic bias. For energy transitions specifically, such underestimation may have consequences for infrastructure planning, investment allocation, and climate policy adequacy assessments. The bias provides a statistical explanation for the persistent underestimation of renewable energy adoption documented in prior studies [29, 217].

We develop a debiasing methodology that corrects this bias under specific conditions. The method performs well when noise structures are well-characterized and data follow idealized diffusion dynamics, but degrades with high noise. Future research should clarify applicability boundaries, refine forecasting techniques, and add robustness checks.

Appendix

The following appendices provide additional details on S-curve derivations, parameter estimation methods, and bias analyses across different models and noise specifications. Appendices are organized as: **Part A** establishes theoretical foundations; **Part B** presents baseline Fisher-Pry results; **Part C** analyzes alternative S-curve and noise models; **Part D** examines NLLS misspecification across noise structures; **Part E** details the debiasing method; **Part F** examines sensitivity to observation count and noise level. Code is available in `s_curves_bias.ipynb` for replication and further analysis.

3.6 Theoretical Foundation

3.6.1 Deriving the logistic curve

Starting from the generalized logistic differential equation, described in eq. 3.1 as

$$\frac{dQ(t)}{dt} = kQ(t)^{\alpha}\left[1 - \left(\frac{Q(t)}{L}\right)^{\beta}\right]^{\gamma}. \tag{3.11}$$

When $\alpha = \beta = \gamma = 1$, we derive the logistic differential equation

$$\frac{dQ(t)}{dt} = kQ(t)\left[1 - \frac{Q(t)}{L}\right]. \tag{3.12}$$

By separation of variables, we derive the analytical solution

$$Q(t) = \frac{L}{1 + Ae^{-kt}}, \tag{3.13}$$

where A is some constant depending on an initial condition. Assume that A has the following form

$$A = e^{kt_0}. \tag{3.14}$$

Then A defines a shift in the curve, relative to the point of maximum growth (the inflection point), which occurs at $t = t_0$ (where $d^2Q(t)/dt^2 = 0$). Thus, t_0 is the inflection point of the S-curve. Substituting eq. 3.14 into eq. 3.13, we derive the standard expression for the 3-parameter symmetric logistic equation,

$$Q(t) = \frac{L}{1 + e^{-k(t-t_0)}}. \tag{3.15}$$

3.6.2 OLS estimator of the S-curve

Let $f_t(\hat{L})$ be the FP transform of a time series, $Y(t)$, conditional on some $\hat{L}$. Hence, for a given $L = \hat{L}$ the remaining parameter estimates can be written as functions of L. That is

$$\begin{aligned}
\hat{k} &= \sum_{t=0}^{m} \frac{\left(t - \bar{t}\right)\left(f_t(\hat{L}) - \bar{f}_t(\hat{L})\right)}{\sum_{t=0}^{m}\left[t - \bar{t}\right]^2} \\
\hat{t}_0 &= \frac{-1}{\hat{k}}\left[\bar{f}_t(\hat{L}) - \hat{k}\bar{t}\right] \\
\hat{\sigma} &= \sqrt{\frac{1}{m-1}\sum_{t=0}^{m}\left[f_t(\hat{L}) - \hat{k}(t - \hat{t}_0)\right]^2},
\end{aligned} \tag{3.16}$$

where $\hat{k}$, $\hat{t}_0$, and $\hat{\sigma}$ are the OLS estimates for k, t_0, and σ, respectively; $\bar{\cdot}$ denotes the mean. We estimate $\hat{L}$ by minimizing sum of squared residuals:

$$\hat{\theta} = \min_{\hat{L},\hat{k},\hat{t}_0}\left(\sum_{t}\left[f_t(\hat{L}) - \hat{k}(t - \hat{t}_0)\right]^2\right)$$

for $\hat{\theta} = \{\hat{L}, \hat{k}, \hat{t}_0\}$, where f is the FP transformation.

Bounds for the optimization. $\max(Y_t) \times 1.001 \leq \hat{L} \leq 10^6$. Lower bound ensures valid S-curve and numerical stability in FP transform $f_t(L) = -\log(L/Y_t - 1)$. Upper bound allows 100× overestimation (true $L = 10^4$), preventing unrealistic extrapolations.

3.6.3 The relationship between S-curve parameters

It follows from eq. 3.6 that if the asymptote is approximately known, such that the estimate $\hat{L} \approx L$, using the OLS estimators, $\hat{k} \approx k$ and $\hat{t}_0 \approx t_0$. Define the F-P transformation using $\hat{L}$, on data Q_t as

$$F_t(\hat{L}) := -\log\left[\frac{\hat{L}}{Q_t} - 1\right] = -\log\left[\frac{\hat{L}}{L} - 1 + \frac{\hat{L}}{L}e^{-k(t-t_0)}\right]. \tag{3.17}$$

where F_t is the Fisher-Pry transformation as defined in eq. 3.6 for $\sigma = 0$. [9]

How does the asymptote estimate affect the inflection point estimate?

Let $\alpha := \frac{\hat{L}}{L}$ and $c := \alpha - 1$. Substituting these into eq. 3.17, the F-P transform becomes

$$F_t(\hat{L}) = -\log\left[c + \alpha e^{-k(t-t_0)}\right].$$

To see the effect of $\hat{L} \neq L$ on $\hat{t_0}$, assume $t << t_0$, as is the case when estimating parameters based on early diffusion levels. Then $c \ll \alpha e^{-k(t-t_0)}$, and hence

$$\log\left[c + \alpha e^{-k(t-t_0)}\right] \approx \log\left[\alpha e^{-k(t-t_0)}\right] = -k(t - t_0) + \log\alpha.$$

By rearranging the right hand side, the OLS estimate for the inflection point parameter will be

$$\hat{t_0} = t_0 + \frac{\log\alpha}{\hat{k}}$$

[9] One can confirm that the results presented here hold for all noise models presented in the paper.

where $\hat{k} > 0$. Thus, when $\hat{L} > L$, $\alpha > 1$ and thus $\hat{t}_0 > t_0$, and vice versa when $\hat{L} < L$. This agrees with the intuition that when one overestimates the asymptote one should also overestimate the inflection point.

How does the asymptote estimate affect the growth parameter estimate? The estimate for the intrinsic growth rate, $\hat{k}$, is the slope of the F-P transformation; i.e. the time-derivative of the F-P transformation. That is, the parameter estimate for k, conditional on some estimate $\hat{L}$, is

$$\hat{k} := \frac{dF_t(\hat{L})}{dt} = \frac{ke^{-k(t-t_0)}}{1 + e^{-k(t-t_0)} - \frac{L}{\hat{L}}}. \tag{3.18}$$

Let

$$\beta := \frac{e^{-k(t-t_0)}}{e^{-k(t-t_0)} + 1 - \frac{L}{\hat{L}}}.$$

In the case where $\hat{L} > L$, $1 - \frac{L}{\hat{L}} > 0$ and thus $\beta < 1$, meaning that $\hat{k} < k$. Similarly, if $\hat{L} < L$, $\beta > 1$, meaning that $\hat{k} > k$.

The F-P transform is not always linear. Lastly, we analyze the behavior of the F-P transform under some estimate $\hat{L}$. That is, we analyze the second derivative of the F-P transform using an estimate for the asymptote. That is,

$$\frac{d^2F_t(\hat{L})}{dt^2} = -\left(1 - \frac{L}{\hat{L}}\right)\frac{k^2 \exp\left(-2k(t-t_0)\right)}{(\exp\left(-k(t-t_0)\right) + 1 - L/\hat{L})^2}$$

Note that

$$\left(e^{-k(t-t_0)} + 1 - L/\hat{L}\right)^2 > 0$$

and

$$k^2 e^{-2k(t-t_0)} > 0.$$

for all $\hat{L}$, k and t_0. Thus, to determine the behavior of the second derivative of the F-P transform we need only analyze

$$-c = \left(1 - \frac{L}{\hat{L}}\right).$$

Then, we can quickly confirm that when $\hat{L} \approx L$,

$$\frac{d^2 F_t(\hat{L})}{dt^2} \approx 0$$

meaning the F-P transformation is approximately linear and thus a OLS regression will generate approximately unbiased estimates for the remaining parameters. Similarly, when $\hat{L} > L$ or $\hat{L} < L$ the F-P transform will be concave down and convex respectively. This result provides intuition as to why we expect $\hat{k} > k$ for $\hat{L} < L$, and vice versa for $\hat{L} < L$. Take a small proportion of the s-curve and assume $\hat{L} < L$, such that the F-P transform is concave. Thus, when fitting an OLS regression to estimate the slope $\hat{k}$, you will over-estimate the growth parameter in the mean/median. The opposite will be the case when $\hat{L} > L$, in which case you will underestimate the slope as the F-P transform is convex.

3.6.4 Mapping two S-curves to one another

Consider two time series, $Y(t)^{(1)}$ and $Y(t)^{(2)}$ where all parameters are arbitrary.[10] Assume we are interested in the diffusion level for $\mathbb{E}\big[Y(t)^{(1)}\big]$ at some time t_1. Then the adoption level at time t_1 is

$$\frac{\mathbb{E}\big[Y(t_1)^{(1)}\big]}{L^{(1)}} = \frac{1}{1 + \exp\left(-k^{(1)}(t_1 - t_0^{(1)})\right)}.$$

We can find a time t_2 such that

$$\frac{\mathbb{E}\big[Y(t_1)^{(1)}\big]}{L^{(1)}} = \frac{\mathbb{E}\big[Y(t_2)^{(2)}\big]}{L^{(2)}}.$$

[10] Note that $\mathbb{E}[\sigma^{(\cdot)}\epsilon_t^{(\cdot)}] = 0 \ \forall \ t$ for $\cdot \in \{1, 2\}$.

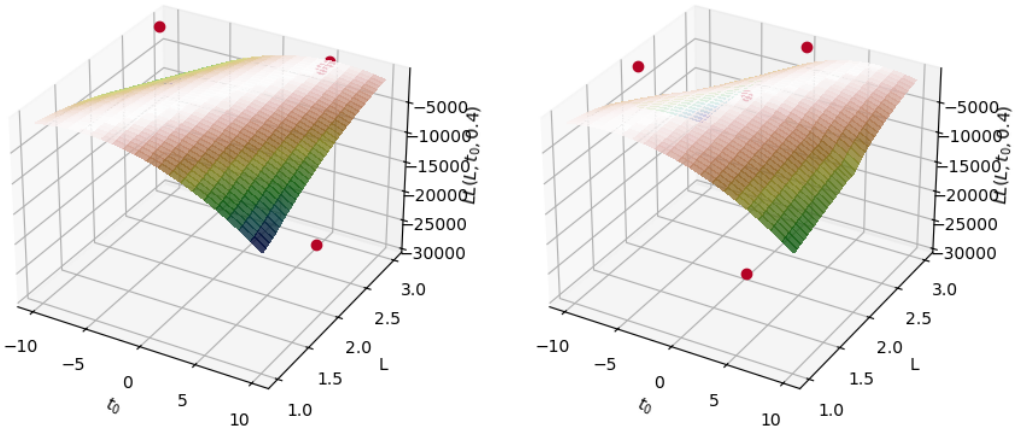
The equality holds when

$$t_2 := \frac{k^{(1)}}{k^{(2)}}\left(t_1 - t_0^{(1)}\right) + t_0^{(2)}.$$

Then the time t_2 is the time for time series $Y(t_2)^{(2)}$ that corresponds to the mean adoption level for $Y(t_1)^{(1)}$.

3.6.5 The optimization landscape at early diffusion

To understand the difficulty associated with estimating parameters of the logistic curve at early diffusion stages, we visualize the objective function from eq. 3.7 as a 3D surface over the parameter space. Figure 3.6 shows the sum of squared residuals as a function of (L, t_0) conditional on $k = 0.30$ for two observation windows. The flat ridge in the early-diffusion



(a) Sum of squared residuals at $t_m - t_0 = -5$ (approximately 10% diffusion).

(b) Sum of squared residuals at $t_m - t_0 = 5$ (approximately 75% diffusion).

Figure 3.6: Optimization function as a function of L and t_0 conditional on $k = 0.30$ for a time series with parameters $\{10^4, 0.3, 0, 0.1\}$. Red dots indicate the position of the true parameter values. **(a)** At early diffusion ($t_m - t_0 = -5$), the objective function exhibits a flat ridge where multiple (L, t_0) combinations produce nearly identical fit quality, making parameter identification difficult. **(b)** At later diffusion ($t_m - t_0 = 5$), the objective function becomes sharper around the true values, enabling more precise parameter estimation.

landscape (Figure 3.6a) demonstrates why parameter estimates exhibit high variance and systematic bias. Multiple parameter combinations along this ridge produce nearly identical sum of squared residuals, preventing the optimization from reliably identifying the true

values. As more of the S-curve is observed (Figure 3.6b), the objective function develops a sharper minimum near the true parameters, substantially improving estimation quality.

3.7 Baseline Fisher-Pry Results

3.7.1 Simulation configuration

Define the parameters, $\theta = \{L, k, t_0, \sigma\}$ as

$$\theta = \{10^4, 0.3, 0, 0.1\}.$$

for the number of observations

$$m = 50$$

and number of simulations

$$N = 10{,}000.$$

for each

$$T \in [t_0 - 15, t_0 + 15]$$

where T is the time of the last observation such that the first data point is observed at time $T - m$.

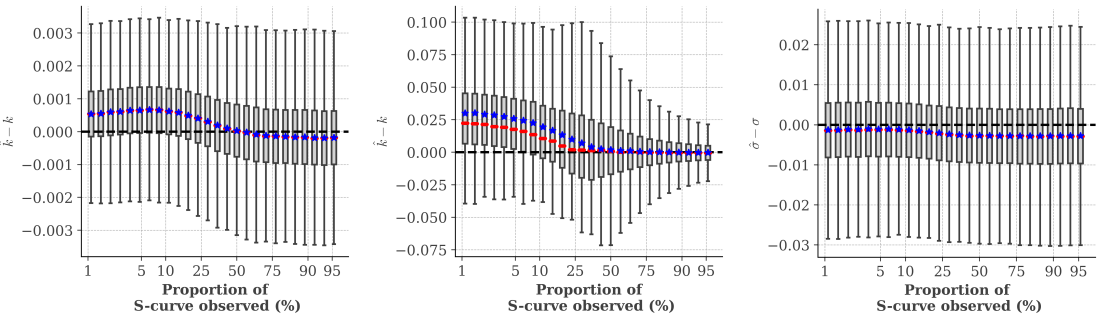
3.7.2 Potential dependence of bias on growth rate

All simulations use baseline $k = 0.3$ year^{-1}. Varying k with fixed observation times implicitly changes diffusion levels reached, so our results across diffusion proportions may partly capture effects of varying the growth rate. However, our approach (varying diffusion with fixed k) may not fully capture direct k-dependence through two competing effects: (1) *curvature and signal strength*: higher k produces steeper curves with stronger curvature.

For example, the logistic second derivative at inflection point $\frac{d^2Q(t=t_0)}{dt^2} = k^2L/4$ scales with k^2, providing more distinguishable asymptotic information; (2) *temporal discretization* - higher k reduces the samples per given diffusion level (i.e. temporal scale) by $(1/k)$. That is, a higher k provides less data per given diffusion level. For $k < 1$, the $1/k$ will have a stronger effect than the curvature benefit ($\propto k^2$), suggesting higher k for $0 < k < 1$ may increase estimation difficulty. Systematic investigation of k-dependence remains future work.

3.7.3 Baseline bias in growth and noise parameters

Growth parameter k and noise parameter σ remain approximately unbiased under well-specified FP estimation (main text focuses on L, t_0 bias), although deviations exist (Figure 3.7): baseline conditions $\theta = \{10^4, 0.3, 0, 0.1\}$, $m = 50$, 10,000 replications.



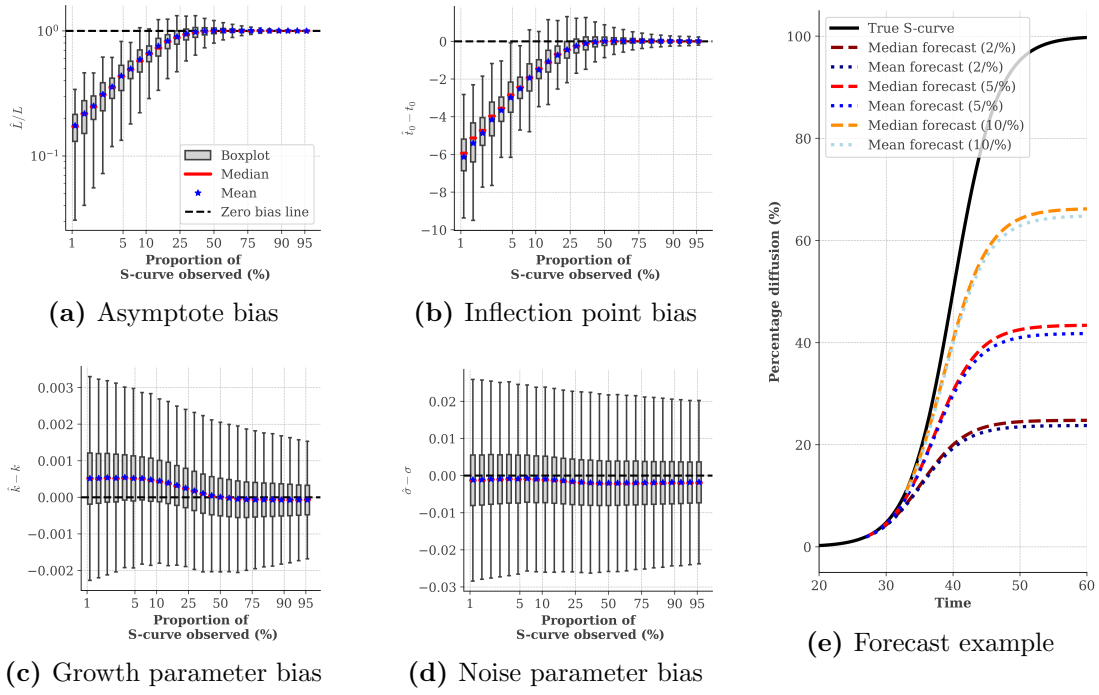
(a) FP: Growth parameter **(b)** NLLS: Growth parameter **(c)** FP: Noise parameter

Figure 3.7: Well-specified FP estimation produces approximately unbiased k **(a)** and σ **(c)** estimates, with median within 1% of true values across all diffusion levels. Misspecified NLLS applied to FP noise data **(b)** also produces approximately unbiased k estimates, validating robustness for debiasing. NLLS σ not shown—misspecified model does not estimate σ as defined in the surrogate DGP (eq. 3.5).

FP produces low-bias k and σ while L and t_0 are biased (main text). k transitions from slight upward to slight downward bias due to parameter interdependence (appendix 3.6.3). Critically, NLLS on FP data also produces approximately unbiased k. NLLS estimate for σ is not shown (misspecified model doesn't estimate σ from true DGP).

3.7.4 FP noise model keeping starting point constant

Figure 3.8 shows bias in parameters estimates with $m = 50$ datapoints at starting point, adding one observation at a time ($\theta = \{10^4, 0.3, 0, 0.1\}$, 10,000 simulations). Patterns are visually similar to constant $m = 50$ (main text), confirming diffusion proportion is primary bias determinant. Bias in L and t_0 are limited to ¡25% diffusion. Variance decreases faster than for constant $m = 50$, particularly for k (Figure 3.8c). This is expected, as more observations reduce variance.



(a) Asymptote bias **(b)** Inflection point bias

(c) Growth parameter bias **(d)** Noise parameter bias **(e)** Forecast example

Figure 3.8: Adding observations forward from $m = 50$. Bias patterns confirm diffusion proportion dominates. Variance decreases with more observations.

3.8 Alternative S-Curve and Noise Models

Noise parameter scaling across models. Table 3.3 shows noise magnitudes at 10% diffusion for different models.

Fisher-Pry noise operates in logit space: $f_t = k(t - t_0) + \sigma\epsilon_t$ where $f_t = -\log(L/Y_t - 1)$. The noise is constant in logit space but varies in levels, growing to maximum at the inflection

Noise Model	σ value	Noise at 10% diffusion ($Y = 1000$)
Fisher-Pry	0.1	$\pm$10% ($\pm$100)
Additive	100	$\pm$10% ($\pm$100)
Multiplicative	0.25	$\pm$28% ($\pm$280)

Table 3.3: Noise magnitude (for $\epsilon_t = \pm 1$) at 10% diffusion across noise models.

point. Well-specified estimation uses OLS on the Fisher-Pry transformation.

Additive noise operates in levels: $Y(t) = L/(1+e^{-k(t-t_0)}) + \sigma\epsilon_t$, with constant absolute noise ($\pm$100)—homoskedastic in levels. At 10% diffusion, this gives $\pm$10% relative variation, matching Fisher-Pry magnitude. Well-specified estimation uses NLLS directly on levels.

Multiplicative noise $Y(t) = Le^{\sigma\epsilon_t}/(1 + e^{-k(t-t_0)})$ produces proportional errors. Well-specified estimation: NLLS on $\log Y_t$. Unlike Fisher-Pry OLS, multiplicative NLLS computes $\log Y_t$ independently of parameters. With $\sigma = 0.1$, this produces approximately unbiased estimates at early diffusion. We use $\sigma = 0.25$ to demonstrate comparable underestimation, showing bias arises from S-curve geometry not estimation procedures.

3.8.1 The Gompertz curve

The Gompertz S-curve [11] is used in epidemic [80], cell biology [81], and population dynamics [59] models. It is defined as

$$Q(t) = L \exp\left(-\exp\left(-k(t - t_0)\right)\right)$$

where $Q(t)$ is the diffusion, L is the asymptote, k is the growth rate and t_0 the inflection point. The Gompertz curve has an inflection point at approximately 37% diffusion.

Gompertz noise model. Noise model:

$$Y(t) = L \exp\left(-\exp\left(-k(t - t_0) + \sigma\epsilon_t\right)\right), \quad \epsilon_t \sim \mathcal{N}(0, 1).$$

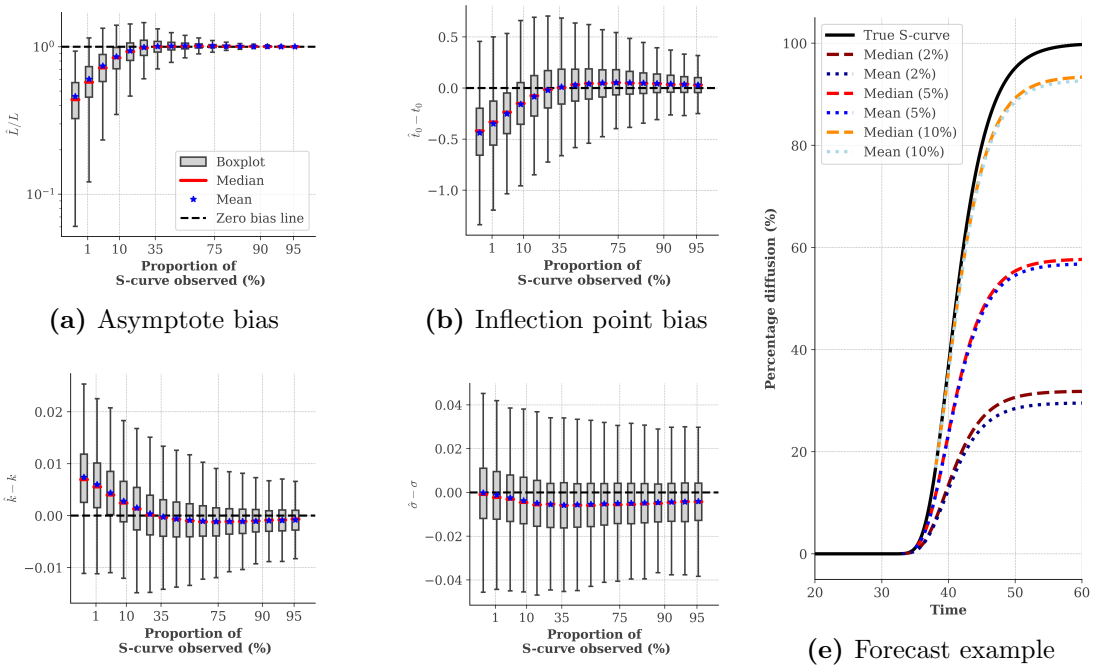
Transformation $-\log\log\left(L/Y(t)\right) = k(t-t_0) - \sigma\epsilon_t$ enables OLS estimation. Parameter relationships mirror logistic: $\hat{L} > L \Rightarrow \hat{k} < k, \hat{t}_0 > t_0$. This noise specification preserves scale invariance analogous to the logistic Fisher-Pry case, so the debiasing procedure (appendix 3.10.4) can be adapted to the Gompertz curve by generating surrogate data from the Gompertz DGP with this noise structure.

3.8.2 Bias in the parameter estimates

We test parameter estimates for

$$\theta = \{10^4, 0.3, 40, 0.1\}$$

with 1,000 simulations for constant $m = 50$ across the observation window (i.e. shifting the starting point when increasing last observed diffusion level). Figure 3.9 shows bias in all parameters and examples mean and median forecasts at approximately 2, 5, and 10% observed diffusion. Bias patterns are similar to the logistic case in the main paper, with systematic underestimation of L and t_0 at early diffusion levels. Winsor [269] documented that the Gompertz curve exhibits estimation challenges at early diffusion stages, consistent with our findings. Bias is confined to $< 20\%$ diffusion, which is less than the case shown in the main text for the logistic. This may be due to different model structures. Bias behaviour is similar. For both mean and median L and t_0 are underestimated at early diffusion, while k is marginally but negligibly overestimated, and σ near-unbiased throughout. This translates into underestimated point forecasts in mean and median.



(a) Asymptote bias

(b) Inflection point bias

(c) Growth parameter bias

(d) Noise parameter bias

(e) Forecast example

Figure 3.9: Gompertz bias under Fisher-Pry noise. Mean (blue) and median (red). **(a)** Asymptote and **(b)** inflection point are downward biased before ∼10% diffusion. **(c)** Growth and **(d)** noise parameters remain near-unbiased throughout. **(e)** Forecast examples at 2, 5, and 10% diffusion

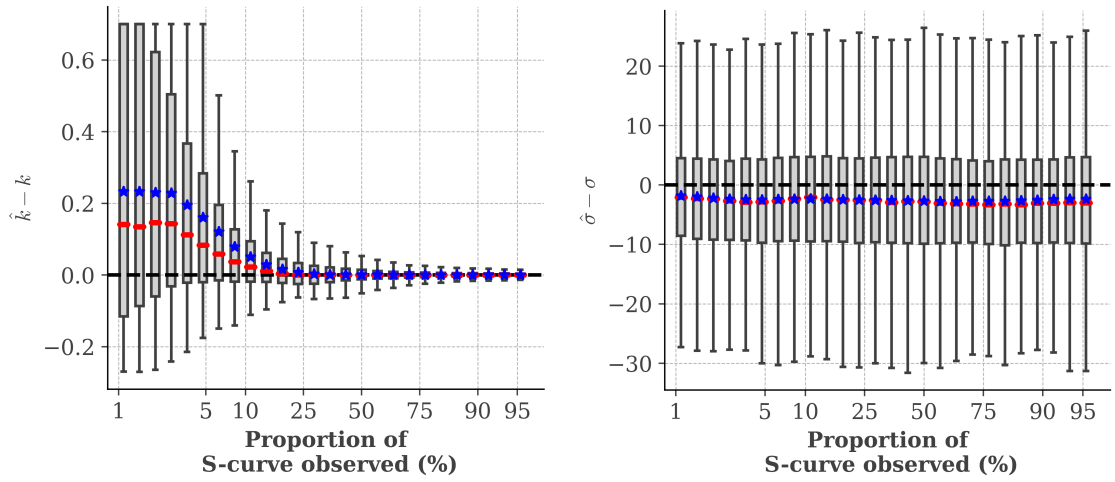
3.8.3 Additive noise model

An additive noise model for the logistic curve is

$$Y(t) = \frac{L}{1 + \exp\left(-k(t - t_0)\right)} + \sigma\epsilon_t$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and other parameters are as defined previously. A well-specified estimator is NLLS on Y_t. We run 10,000 simulations with $\theta = \{10^4, 0.3, 40, 100\}$, starting at $m = 50$ and adding one point at a time. Bias in L and t_0 shown in main text (Figure 3.3).

Growth and noise parameter bias. Figure 3.10 shows bias in k and σ under additive noise with $\sigma = 100$. Constant absolute noise becomes relatively larger at early/late diffusion, creating higher k variance than Fisher-Pry (where noise scales with signal). σ



(a) Bias in the growth parameter

(b) Bias in the noise parameter

Figure 3.10: Additive noise with well-specified NLLS and $\sigma = 100$ (calibrated to match Fisher-Pry noise magnitude at 10% diffusion). k shows upward bias with increased variance at early diffusion; σ remains approximately unbiased. The heteroskedastic error structure affects variance at early diffusion where relative noise is higher.

estimates remain approximately unbiased; variance ± 7.5 ($\pm 7.5\%$) constant across diffusion.

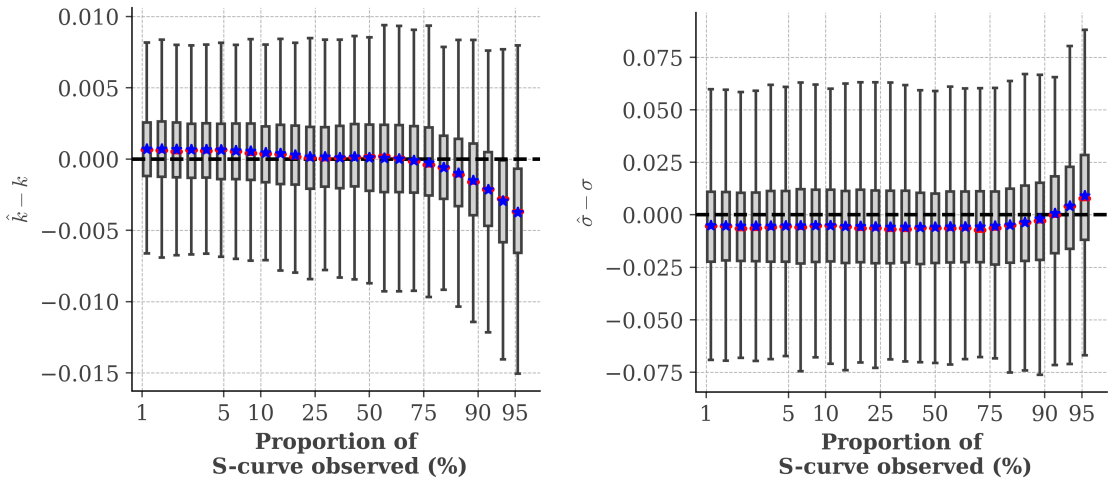
3.8.4 Multiplicative noise model

A multiplicative noise model is

$$Y(t) = \frac{L \exp(\sigma \epsilon_t)}{1 + \exp(-k(t - t_0))}$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and other parameters are as defined previously. Well-specified estimation uses NLLS on $\log Y(t)$. We run 1,000 simulations with $\theta = \{10^4, 0.3, 40, 0.25\}$, starting at $m = 50$ and adding one point at a time. We use $\sigma = 0.25$ rather than $\sigma = 0.1$ because multiplicative noise with lower σ and well-specified estimation produces approximately unbiased parameter estimates even at early diffusion, making it difficult to demonstrate the systematic underestimation phenomenon. Bias in L and t_0 are shown in main text (Figure 3.3).

Growth and noise parameter bias. Figure 3.11 shows bias in k and σ under multiplicative noise with $\sigma = 0.25$. $\sigma = 0.25$ produces comparable underestimation; L and



(a) Bias in the growth parameter

(b) Bias in the noise parameter

Figure 3.11: Multiplicative noise with well-specified NLLS on $\log Y(t)$ and $\sigma = 0.25$ (calibrated to show underestimation comparable to other noise models). Both k and σ remain approximately unbiased throughout diffusion. The proportional error structure provides stable estimation properties.

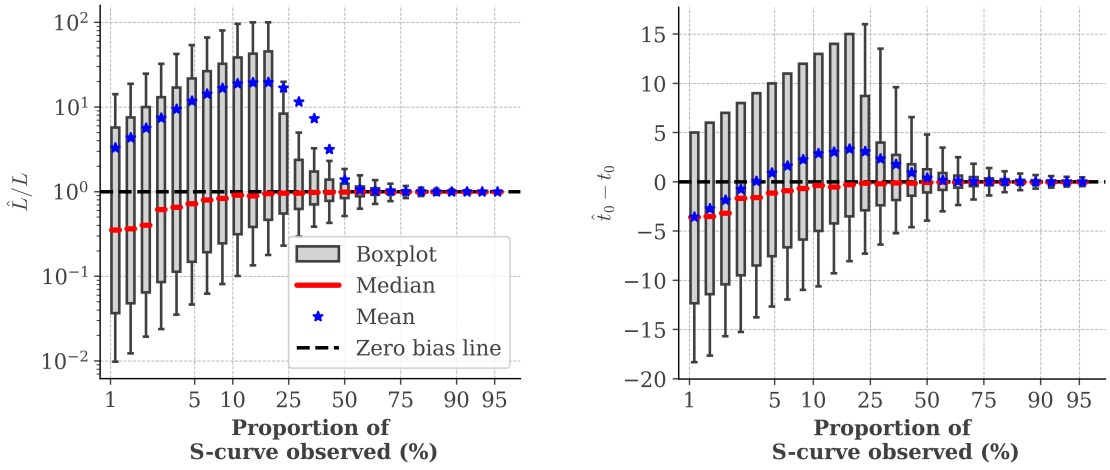
t_0 median bias persists to 10-15% diffusion. k and σ approximately unbiased throughout (well-behaved log-space error structure).

3.9 NLLS Misspecification

3.9.1 NLLS misspecification: poor performance across noise models

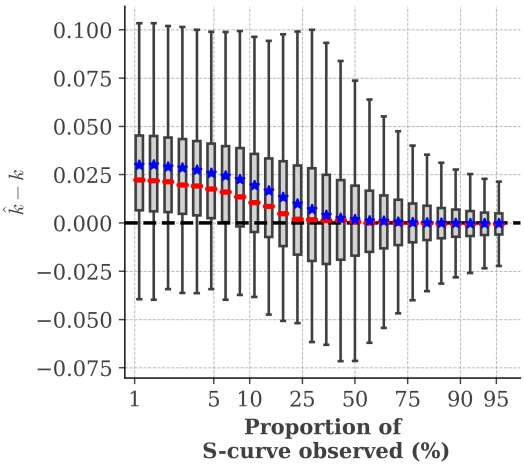
This section complements the main text NLLS analysis (Figure 3.2) by examining NLLS performance across all noise models. Figure 3.12 shows misspecified NLLS applied to multiplicative noise data (10,000 simulations, $\theta = \{10^4, 0.3, 40, 0.25\}$, $m = 50$), producing biased L and t0 estimates with increased variance.

Misspecification produces increased variance due to heteroskedasticity mismatch. Well-specified NLLS on additive (appendix 3.8.3) also exhibits bias at early diffusion.

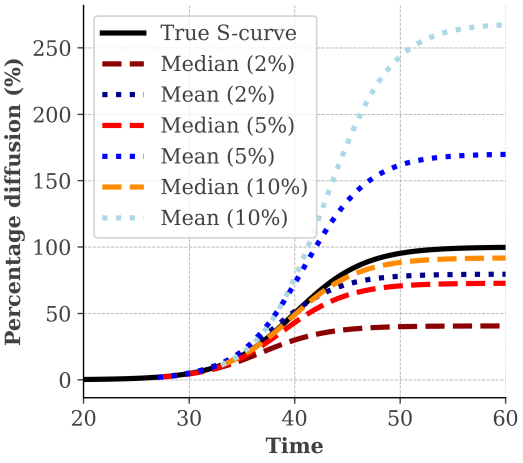


(a) Asymptote bias

(b) Inflection point bias



(c) Growth parameter bias



(d) Forecast example

Figure 3.12: Misspecified NLLS applied to multiplicative noise data. Median (red) and mean (blue). Median bias persists until 10-25% diffusion. **(a)** Asymptote downward biased in median, upward in mean at early diffusion. **(b)** Inflection point median downward biased; mean shifts from downward to upward. **(c)** Growth parameter upward biased. **(d)** Forecasts: median underestimates, mean overestimates.

3.10 Debiasing Methodology

3.10.1 Debiasing implementation

3.10.1.1 Algorithm pseudocode

For time-series Y_t with m observations, Algorithm 1 implements the complete debiasing procedure described in Section 3.2.

Algorithm 1: Debiasing procedure for S-curve parameter estimates.

Input: Time-series Y_t with m observations; surrogate parameter space $(k^{(s)}, \sigma^{(s)}, m, \mathcal{D})$ where $\mathcal{D}$ is the set of observation windows.

Output: Debiased parameter estimates $\hat{\theta} = \{\hat{L}, \hat{k}, \hat{t}_0, \hat{\sigma}\}$

$\hat{\theta}^{(b)} \leftarrow \arg\min_{\hat{L},\hat{k},\hat{t}_0} \sum_t [f_t(\hat{L}) - \hat{k}(t - \hat{t}_0)]^2$;

$t^{(b)} \leftarrow t_m - \hat{t}_0^{(b)}$;

for $d \in \mathcal{D}$ **do**

  Generate N surrogate samples $Y_t^{(s,i,d)}$ with $(L^{(s)}, k^{(s)} = \hat{k}^{(b)}, t_0^{(s)}, \sigma^{(s)} = \hat{\sigma}^{(b)})$;

  Estimate $\hat{\theta}^{(s,i,d)}$ for each sample $i = 1, \ldots, N$;

  Compute $\bar{t}^{(s,d)} \leftarrow \mathbb{E}_i[t_m - \hat{t}_0^{(s,i,d)}]$;

end

$d^* \leftarrow \arg\min_{d \in \mathcal{D}} |t^{(b)} - \bar{t}^{(s,d)}|$;

$\beta_L \leftarrow \mathbb{E}_i[\hat{L}^{(s,i,d^*)} / L^{(s,d^*)}]$;

$\hat{L} \leftarrow \hat{L}^{(b)} / \beta_L$;

$\{\hat{k}, \hat{t}_0, \hat{\sigma}\} \leftarrow \mathrm{OLS}(f_t(\hat{L}), Y_t)$;

return $\hat{\theta}$;

The notation $\hat{\cdot}^{(b)}$ denotes parameter estimates before debiasing, $\cdot^{(s)}$ denotes surrogate data parameters, and $f_t(\hat{L})$ denotes the Fisher-Pry transform using the asymptote estimate. The procedure uses fixed surrogate values $(L^{(s)}, t_0^{(s)}) = (10^4, 0)$ for all simulations, exploiting the scale-invariance property detailed in appendix 3.10.3.

3.10.1.2 Simulation configuration

We test the debiasing method on generated data using the data-generating process defined in eq. 3.6. We define the parameters, $\theta = \{L, k, t_0, \sigma\}$ as

$$\theta = \{10^4, 0.3, 0, 0.1\}$$

for the number of observations

$$m = 50$$

and number of simulations

$$N = 1{,}000.$$

for each

$$T \in [t_0 - 15, t_0 + 10].$$

[11] For each simulation $n \in \{1, \ldots, N\}$ and each T, we estimate the parameters as per the estimation procedure described in section 3.1.

Surrogate data for debiasing. For each parameter estimate, $\hat{\theta}^{(n,T)}$, we generate a surrogate dataset using parameters

$$\theta^{(s)} = \{10^4, \hat{k}^{(n,T)}, 0, \hat{\sigma}^{(n,T)}\}$$

for the number of observations

$$m^{(s)} = 50$$

and number of simulations

$$N^{(s)} = 200.$$

for each

$$T^{(s)} \in [t_0 - 20, t_0 + 15].$$

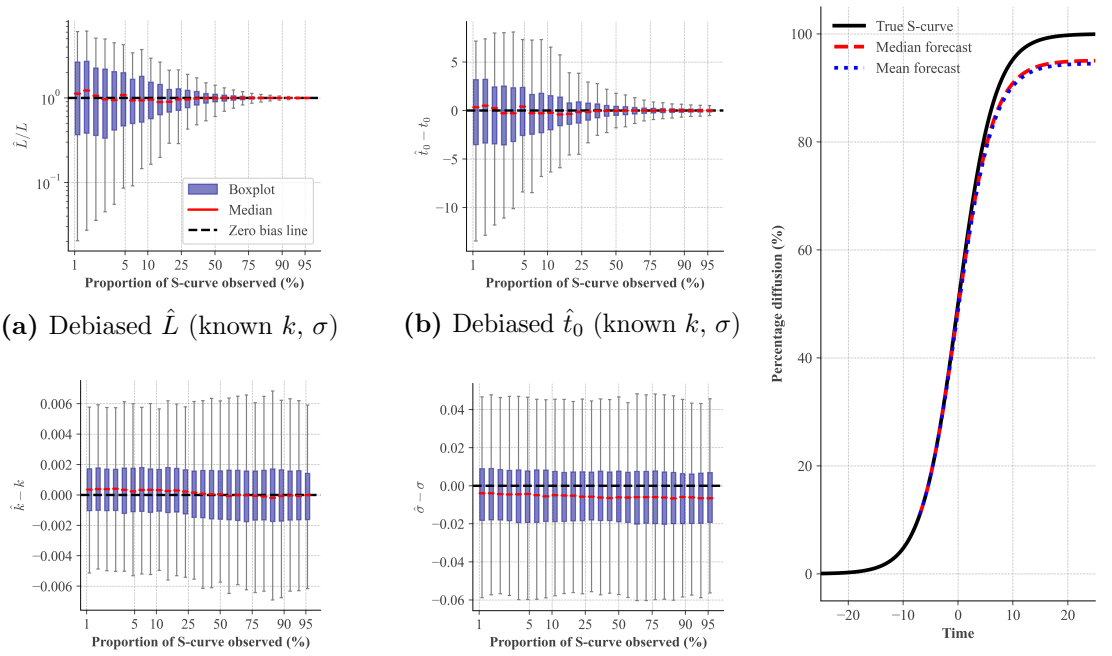
Then, using the surrogate data, we debias the asymptote estimate, $\hat{L}^{(n,T)}$.

Bounds for debiasing optimization. Surrogate data: $\max(Y_t) \times 1.01 \leq \hat{L} \leq 10 \times L^{(s)} = 10^5$. Since true $L^{(s)} = 10^4$ is known, upper bound allows 10× overestimation (vs. 100× for general plots). Well-specified estimates never exceed $2 \times L^{(s)}$ (main text $\sigma = 0.1$; appendix 3.11.2 $\sigma = 0.2$).

[11] Equivalent to the range of 1% diffusion to 95% diffusion observed.

3.10.2 Theoretical performance with known parameters

Figure 3.13: debiasing with *true* k, σ (vs. estimates). Shows method soundness—practical limitation stems from parameter estimation uncertainty, not method itself. Figure 3.13



(a) Debiased $\hat{L}$ (known k, σ) **(b)** Debiased $\hat{t}_0$ (known k, σ)

(c) Debiased $\hat{k}$ (known k, σ) **(d)** Debiased $\hat{\sigma}$ (known k, σ) **(e)** Forecast with known k, σ

Figure 3.13: Debiasing performance when k and σ are known rather than estimated. Results from $N = 1{,}000$ simulations with parameters $\{L = 10^4, k = 0.3, t_0 = 0, \sigma = 0.2\}$ and $m = 50$ observations. **(a-b)** Asymptote and inflection point show improved parameter estimates compared to before debiasing (Figure 3.16). **(c-d)** Growth and noise parameters remain unbiased. **(e)** Forecast (made at $\approx 10\%$ diffusion) with mean (blue) and median (red) closely following the true trajectory (black).

establishes the **theoretical performance limit**: when k and σ are known (not estimated), the debiasing procedure produces reasonably unbiased estimates with lower variance across all diffusion levels. At $\sigma = 0.2$ (higher noise than baseline $\sigma = 0.1$ in Figure 3.4), the procedure remains effective when true parameter values are used. This demonstrates that the **practical limitation** stems from uncertainty in estimating $\hat{k}$ and $\hat{\sigma}$, which propagates through surrogate data generation, rather than from fundamental failure of the debiasing method itself at higher noise levels (See appendix 3.11.2 for results with estimated parameters at $\sigma = 0.2$).

3.10.3 Mathematical Justification of the Debiasing Procedure

The debiasing procedure relies on three key theoretical properties that ensure the correction eliminates bias while maintaining statistical efficiency.

3.10.3.1 Property 1: Conditional unbiasedness of growth and noise parameters

From the Fisher-Pry transformation (eq. 3.6), conditional on $\hat{L} = L$, the transformed data becomes

$$f_t(L) = k(t - t_0) + \sigma\epsilon_t, \tag{3.19}$$

which is a linear model with additive Gaussian noise. Under standard OLS theory, the estimators satisfy

$$\mathbb{E}[\hat{k}|L] = k \quad \text{and} \quad \mathbb{E}[\hat{\sigma}|L] = \sigma. \tag{3.20}$$

Empirical validation through simulation (Fig. 3.7) confirms that $\hat{k} \approx k$ and $\hat{\sigma} \approx \sigma$ across all tested noise levels, with minimal bias compared to the bias in $\hat{L}$ and $\hat{t}_0$. This justifies treating $\hat{k}^{(b)}$ and $\hat{\sigma}^{(b)}$ as approximately unbiased proxies for the true values in the surrogate data generation step.

3.10.3.2 Property 2: Parameter Linkage through the FP Transform

The Fisher-Pry noise model couples parameter estimation through the transform $f_t(\hat{L}) = -\log(\hat{L}/Y_t - 1)$. Appendix 3.6.3 establishes that when $\hat{L}$ changes, all other parameter estimates change accordingly: correcting $\hat{L} \to L$ also corrects $\hat{t}_0$ (via $\hat{t}_0 = t_0 + \log(\hat{L}/L)/\hat{k}$ at early diffusion) and $\hat{k}$ (since $\hat{L} \neq L$ introduces curvature in the transform that biases slope estimates). Once $\hat{L}$ is corrected, re-estimating via OLS on the corrected transform yields improved estimates of all parameters.

This property is directly obvious for the Fisher-Pry noise structure. For additive noise, NLLS estimates all parameters simultaneously without transform-based coupling (appendix 3.10.4).

3.10.3.3 Property 3: Bias Characterization via Surrogate Data

The bias in $\hat{L}$ is a systematic function of:

1. The true parameter values (k, σ) — estimated from the data as $(\hat{k}^{(b)}, \hat{\sigma}^{(b)})$,
2. The observation window, quantified by $t^{(b)} = t_m - \hat{t}_0^{(b)}$,
3. The number of observations, m.

We construct a bias function $\mathcal{B}(k, \sigma, m, d)$ via parametric bootstrap, where we simulate N surrogate samples with parameters $(L^{(s)}, k^{(s)} = \hat{k}^{(b)}, t_0^{(s)}, \sigma^{(s)} = \hat{\sigma}^{(b)})$ and observation window d. For each window, we compute the empirical bias factor

$$\beta_L = \mathbb{E}_i \left[\frac{\hat{L}^{(s,i,d^*)}}{L^{(s,d^*)}} \right]. \tag{3.21}$$

where the expectation is taken over all N surrogate samples indexed by i.

Window Identification. The key insight is that $t^{(b)}$ serves as a sufficient statistic for identifying the observation window. Given that

$$\mathbb{E}[t_m - \hat{t}_0^{(b)}] \approx \mathbb{E}[t_m - \hat{t}_0^{(s,d^*)}] \tag{3.22}$$

for some window d^*, we identify the relevant bias regime by matching:

$$d^* = \arg\min_{d \in \mathcal{D}} |t^{(b)} - \bar{t}^{(s,d)}|. \tag{3.23}$$

3.10.3.4 Bias Correction Mechanism

The debiased asymptote is obtained by inverting the empirical bias factor:

$$\hat{L} = \frac{\hat{L}^{(b)}}{\beta_L}, \tag{3.24}$$

where $\beta_L = \mathbb{E}_i[\hat{L}^{(s,i,d^*)}/L^{(s,d^*)}]$ is the empirical bias factor at the matched window d^*.

Why the bias is multiplicative. The scale-invariance property (appendix 3.10.3.6) establishes that scaling the true parameters and data by a factor α produces proportionally scaled estimates:

$$\hat{L}(\alpha L, \alpha Y) = \alpha \hat{L}(L, Y). \tag{3.25}$$

Taking expectations on both sides yields $\mathbb{E}[\hat{L}(\alpha L)] = \alpha \mathbb{E}[\hat{L}(L)]$. Dividing both sides by the true asymptote gives:

$$\frac{\mathbb{E}[\hat{L}(\alpha L)]}{\alpha L} = \frac{\mathbb{E}[\hat{L}(L)]}{L}. \tag{3.26}$$

Thus the relative bias $\mathbb{E}[\hat{L}]/L$ is independent of L, but rather is a proportion. Define the bias factor $\beta_L(k, \sigma, m, d) := \mathbb{E}[\hat{L}]/L$, which depends only on (k, σ, m, d) per eq. 3.29. This gives

$$\mathbb{E}[\hat{L}] = \beta_L(k, \sigma, m, d) \cdot L, \tag{3.27}$$

where division by β_L removes this proportional error. For noise models lacking scale-invariance (e.g., additive), the bias is not multiplicative and the procedure requires modification (appendix 3.10.4).

3.10.3.5 Theoretical Guarantees

The procedure provides asymptotic unbiasedness under the following conditions:

Assumption 1: Known Data-Generating Process. The surrogate data is generated from the true noise model (eq. 3.5). If the DGP is misspecified, the bias correction may fail.

Assumption 2: (k, σ) Approximately Unbiased. We assume $\hat{k}^{(b)} \approx k$ and $\hat{\sigma}^{(b)} \approx \sigma$. This is empirically validated at baseline conditions ($m = 50$, $\sigma = 0.1$) in appendix 3.7.3, where median bias remains within 5% across all diffusion levels. Deviations at higher noise levels or sparse data may introduce secondary bias (appendix 3.11.1).

Assumption 3: Large N for Surrogate Samples. The empirical bias factor β_L converges to the true bias factor as $N \to \infty$:

$$\beta_L \to \mathbb{E}\left[\frac{\hat{L}^{(s,i,d^*)}}{L^{(s,d^*)}}\right] \quad \text{as } N \to \infty. \tag{3.28}$$

3.10.3.6 Scale-Invariance Requirements

A key requirement is that the bias factor β_L is invariant to (L, t_0) for fixed (k, σ, m, d):

$$\beta_L(L^{(s)}, k, t_0^{(s)}, \sigma, m, d) \approx \beta_L(L', k, t_0', \sigma, m, d). \tag{3.29}$$

When this property holds, we can use fixed arbitrary surrogate parameters $(L^{(s)}, t_0^{(s)})$ to investigate the bias.

Proof sketch. The key insight is that scaling both L and Y_t by the same factor α leaves the Fisher-Pry transform unchanged:

$$f_t(\alpha L) = -\log\left(\alpha L/(\alpha Y_t) - 1\right) = -\log\left(L/Y_t - 1\right) = f_t(L). \tag{3.30}$$

Since the FP transform determines parameter estimates, scaling $(L, Y_t) \to (\alpha L, \alpha Y_t)$ produces scaled estimates $\hat{L} \to \alpha\hat{L}$ with identical relative bias $\hat{L}/L$. Similarly, the S-curve depends only on $(t - t_0)$, making t_0 a pure time shift that does not affect bias patterns. Therefore $\beta_L = \mathbb{E}[\hat{L}/L]$ depends on (k, σ, m, d) but not on the absolute values of (L, t_0).

For applicability to different noise models and S-curve forms, see appendix 3.10.4.

3.10.4 Extension to other noise models and S-curve forms

The debiasing procedure can be adapted to other noise specifications than the FP-noise case presented in the main text, provided the appropriate data-generating process and estimation method are used. Scale invariance determines whether fixed surrogate parameters $(L^{(s)}, t_0^{(s)})$ can be used, which is crucial to the debiasing method.

Scale-invariant cases (debiasing procedure applicable).

1. **Fisher-Pry noise with Gompertz curve:** $Y(t) = L\exp(-\exp(-k(t - t_0) + \sigma\epsilon_t))$ where $\epsilon_t \sim \mathcal{N}(0, 1)$. Use OLS on Gompertz transform $-\log\log(L/Y(t)) = k(t-t_0) - \sigma\epsilon_t$. Scale invariance holds by identical reasoning to the logistic case: scaling preserves the transform structure. The debiasing procedure should apply, with surrogate data generated from the Gompertz FP DGP and Gompertz OLS estimation. See appendix 3.8.1 for bias patterns.

2. **Multiplicative noise in levels:** $Y(t) = Le^{\sigma\epsilon_t}/(1+e^{-k(t-t_0)})$. Use NLLS on $\log Y(t)$. Scaling $(L, Y) \to (\alpha L, \alpha Y)$ shifts $\log Y \to \log Y + \log\alpha$, producing $\hat{L} \to \alpha\hat{L}$ with identical relative bias $\hat{L}/L$. The debiasing procedure should apply, with surrogate data generated from the multiplicative DGP and NLLS estimation on $\log Y$. See appendix 3.8.4 for bias patterns.

Noise models without scale-invariance property (debiasing method not directly applicable):

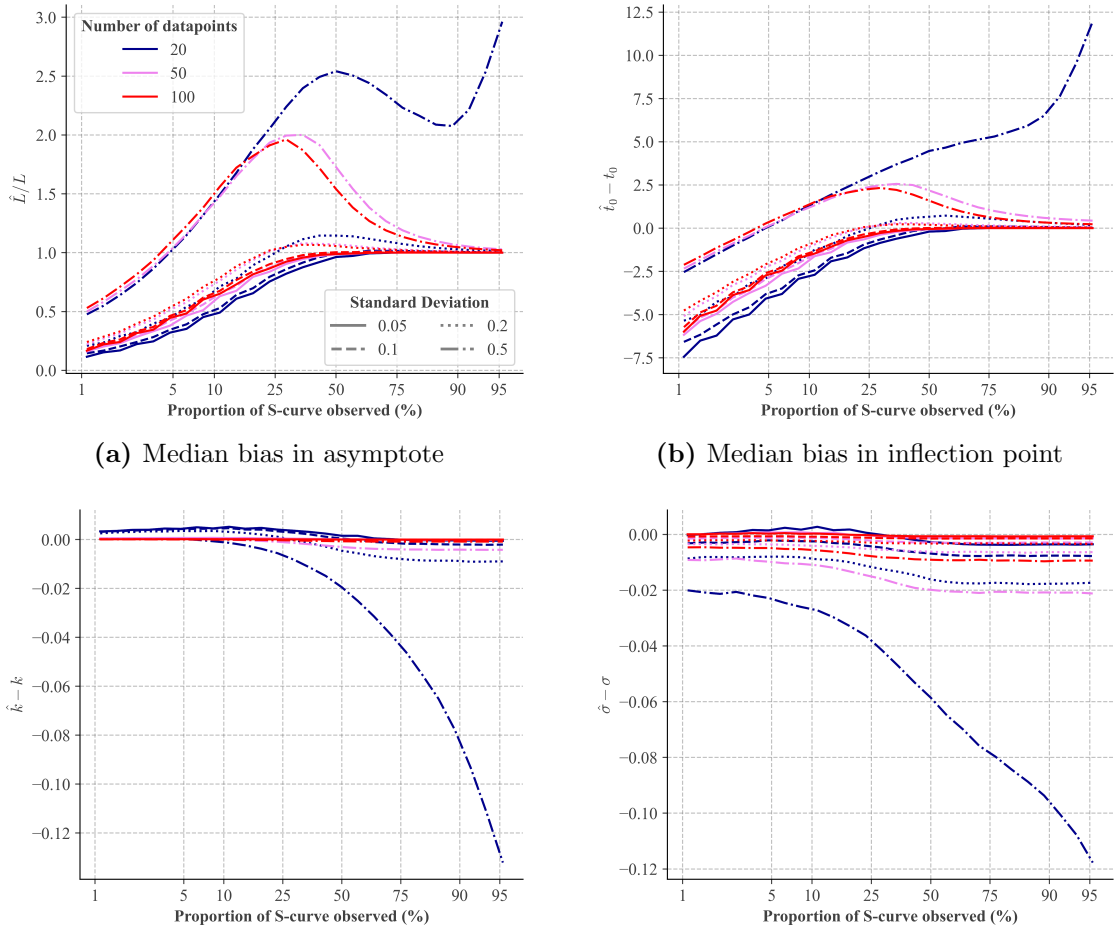
1. **Additive noise in levels:** $Y(t) = L/(1 + e^{-k(t-t_0)}) + \sigma\epsilon_t$. Well-specified estimator uses NLLS on Y_t. Scaling $L \to \alpha L$ changes $Y \to \alpha Q(t) + \sigma\epsilon_t$, not αY, because the absolute noise term $\sigma\epsilon_t$ remains unchanged. Larger L with fixed σ produces relatively lower noise and thus different bias patterns. The bias factor $\beta_L = \mathbb{E}[\hat{L}/L]$ therefore depends on both L and σ separately, not just on (k, σ, m, d) as in the scale-invariant cases. Thus the proposed debiasing method would not apply here. Extension to additive noise could potentially investigate generating surrogate data with $L^{(s)} = \hat{L}^{(b)}$ to match the absolute scale. See appendix 3.8.3 for bias patterns.

3.11 Sensitivity Analysis

3.11.1 Detailed analysis: variance and N effects on parameter bias

Comprehensive analysis of bias and variance dependence on m (observations) and σ (noise). Main text shows $m = 50$, $\sigma = 0.1$; here we examine robustness across parameter space.

Bias dependence on m and σ. Figure 3.14 shows median bias for all four parameters across $m \in \{20, 50, 100\}$ and $\sigma \in \{0.05, 0.1, 0.2, 0.5\}$ combinations, with 10,000 simulations per condition. At $\sigma = 0.5$, $m = 20$: median L bias remains positive across all diffusion (even 95%). Extreme noise distorts growth section; without observing leveling-off phase, estimation overestimates L. Similar anomalies occur for other parameters at $\sigma = 0.5$, especially $m = 20$.



(a) Median bias in asymptote

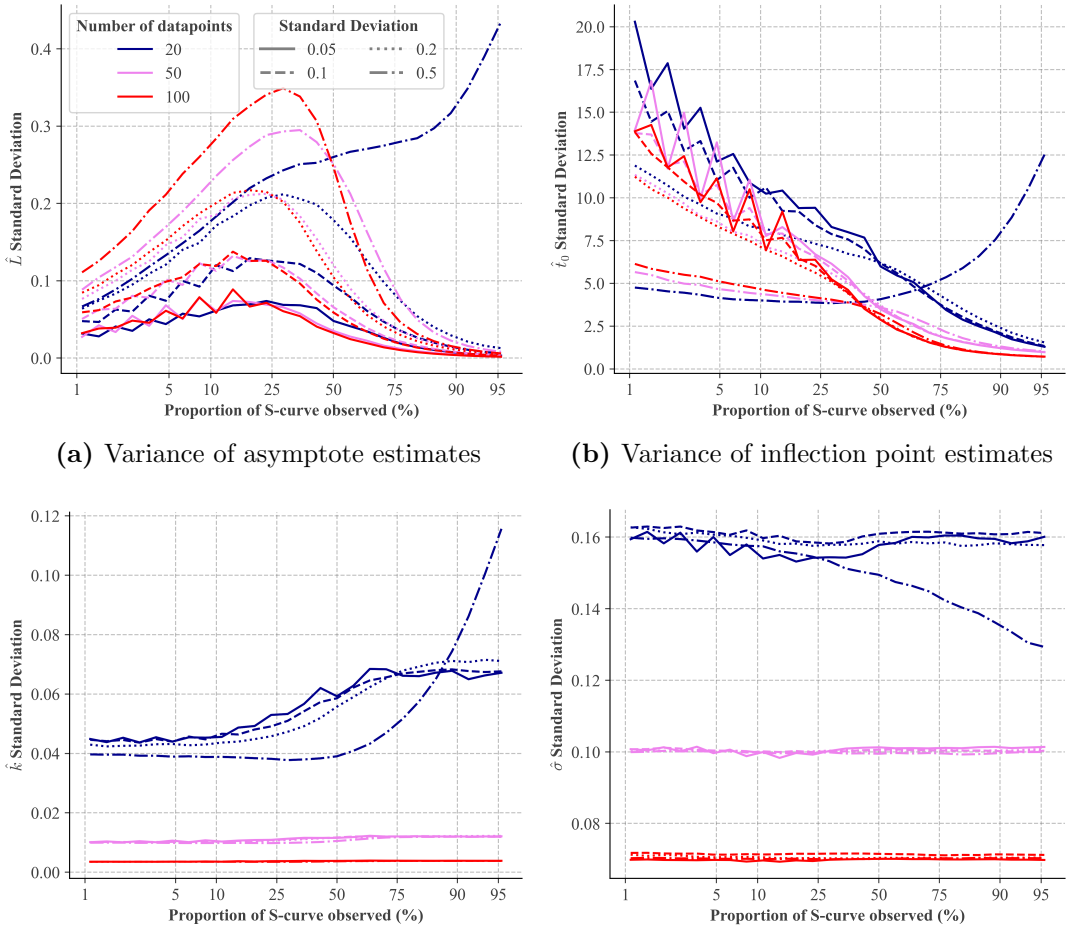
(b) Median bias in inflection point

(c) Median bias in growth parameter

(d) Median bias in noise parameter

Figure 3.14: Median bias vs. diffusion for $m \in \{20, 50, 100\}$, $\sigma \in \{0.05, 0.1, 0.2, 0.5\}$. **(a-b)** Asymptote and inflection point: diffusion proportion dominates; m and σ have smaller effects. At $\sigma = 0.5$, bias shifts positive (distortion from extreme noise). **(c)** Growth parameter transitions overestimate to underestimate, especially $m = 20$. **(d)** Noise parameter depends on true σ, weakly on m or diffusion.

Variance dependence on m and σ. Figure 3.15 shows normalized standard deviation (coefficient of variation) for all four parameters across $m \in \{20, 50, 100\}$ and $\sigma \in \{0.05, 0.1, 0.2, 0.5\}$ combinations, with 10,000 simulations per condition. L and t_0 variance decrease with diffusion; σ affects magnitude. k and σ variance depend on m, weakly on diffusion. Anomalous behavior at $m = 20$, $\sigma = 0.5$ (extreme noise overwhelms limited data).



(a) Variance of asymptote estimates

(b) Variance of inflection point estimates

(c) Variance of growth parameter estimates

(d) Variance of noise parameter estimates

Figure 3.15: Normalized standard deviation vs. $m \in \{20, 50, 100\}$, $\sigma \in \{0.05, 0.1, 0.2, 0.5\}$, diffusion. **(a)** Asymptote: diffusion dominates, decreases after 50%. **(b)** Inflection point: decreases with diffusion, lower noise, more observations. **(c-d)** Growth and noise: depend on m, weakly on diffusion or σ. Anomalies at $m = 20$, $\sigma = 0.5$.

3.11.1.1 Implications for debiasing procedure

At $m = 50$, $\sigma = 0.1$: (1) k bias is low with low variance; (2) $\hat{\sigma}$ has minimal bias and variance; (3) L, t_0 require diffusion-dependent correction. At $\sigma = 0.2$ there is much higher variance in $\hat{k}$ and $\hat{\sigma}$, which degrades debiasing (appendix 3.11.2.2). At $\sigma = 0.5$ there are more erratic patterns, showcasing that the debiasing method is unsuited for very high noise.

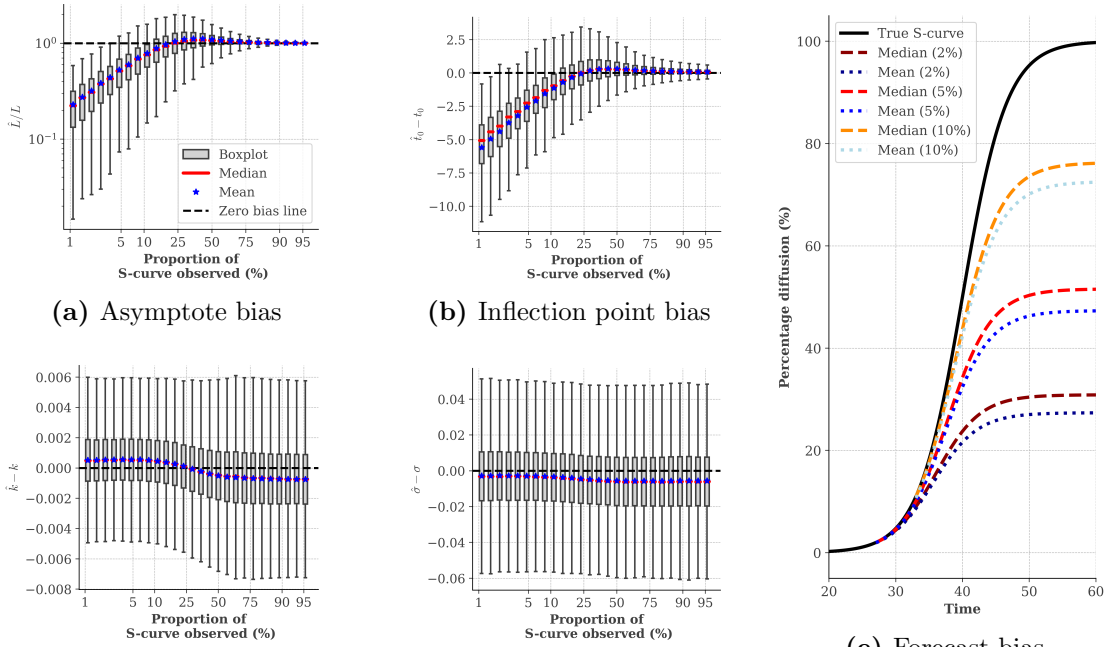
More data at early diffusion ($m = 100$ at 5%) reduces variance but can't overcome under-identification; higher diffusion ($m = 50$ at 20%) provides qualitatively new asymptotic information.

3.11.2 Bias in parameter estimates with higher noise levels

This section examines bias at $\sigma = 0.2$ compared to baseline $\sigma = 0.1$ shown in the main text, holding other parameters constant ($k = 0.3$, $L = 10^4$, $t_0 = 0$, $m = 50$). NLLS misspecification shows patterns consistent with $\sigma = 0.1$ with further increased variance (median bias persists to ∼25% diffusion and interquartile range widens), confirming robustness of the main text findings. We focus here on well-specified estimation performance degradation.

3.11.2.1 Well-specified FP estimation with higher noise levels

Figure 3.16 shows well-specified estimation at $\sigma = 0.2$. Median bias persists to ∼25% diffusion (similar to the main text case of $\sigma = 0.1$); variance increases but remains lower than NLLS.



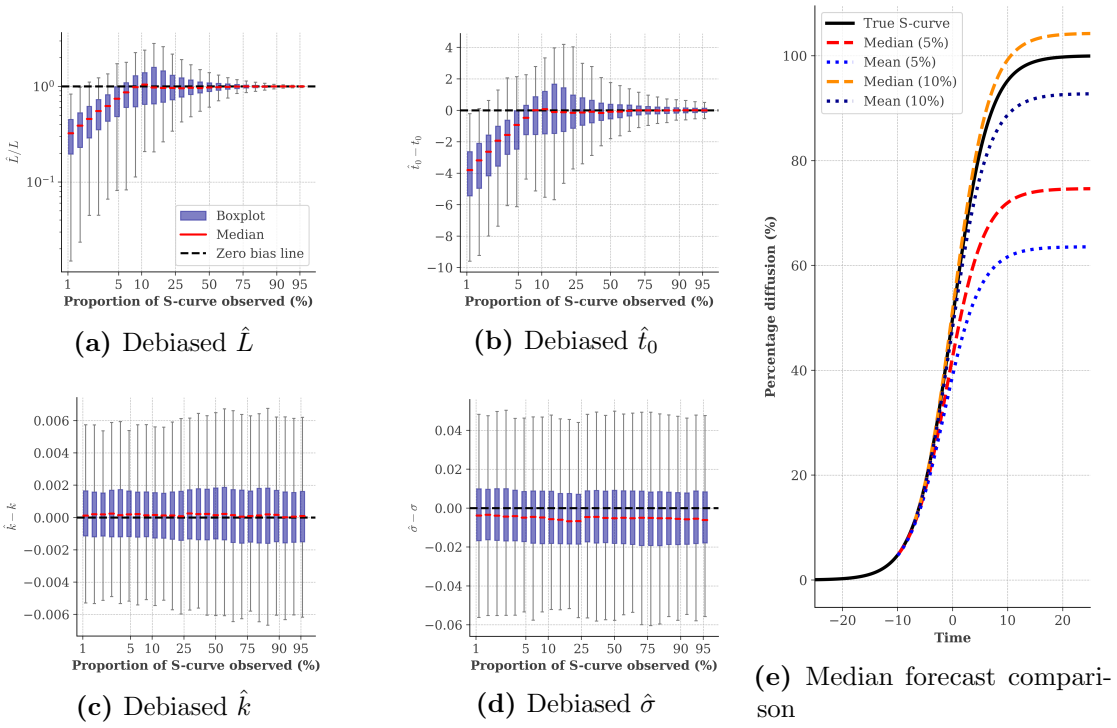
(a) Asymptote bias **(b)** Inflection point bias

(c) Growth parameter bias **(d)** Noise parameter bias **(e)** Forecast bias

Figure 3.16: Well-specified FP estimation with $\sigma = 0.2$. 10,000 simulations, $m = 50$. Boxes (25-75%) remain below zero until ∼5% diffusion, compared to ∼10% at $\sigma = 0.1$ (main text).

3.11.2.2 Debiasing results at higher noise levels

In figure 3.17 we plot the results of the debiasing procedure for simulations across different proportions of the S-curve observed with $\sigma = 0.2$, compared to the results shown in the main text for $\sigma = 0.1$ (figure 3.4).



(a) Debiased $\hat{L}$ **(b)** Debiased $\hat{t}_0$

(c) Debiased $\hat{k}$ **(d)** Debiased $\hat{\sigma}$ **(e)** Median forecast comparison

Figure 3.17: Debiasing at $\sigma = 0.2$ using estimated $\hat{k}$, $\hat{\sigma}$. $N = 1{,}000$ simulations, $m = 50$. **(a-b)** Median bias crosses zero at ∼10% diffusion. IQR at 10% diffusion: approx. 70-125%L . **(c-d)** Growth and noise parameters remain near-unbiased. **(e)** Forecast at ∼10% diffusion approximately tracks true trajectory.

Performance degradation from parameter uncertainty. Compared to main text ($\sigma = 0.1$), median bias persists longer at early diffusion and IQR widens. At 1-5% diffusion, median $\hat{L}$ ranges 30-80%L, compared to initial bias of 20-50%L. Estimates $\hat{k}$ and $\hat{\sigma}$ exhibit higher variance at higher noise levels (figures 3.15c, 3.15d), propagating uncertainty into surrogate data generation. Appendix 3.10.2 uses *true* k and σ at $\sigma = 0.2$ and shows that that would be sufficient information to reduce bias.

Chapter 4

Universality and Predictability of Technology Diffusion

This chapter is taken from: Wagenvoort, B., Dyer, J. C., Lafond, F., & Farmer, J. D. (2026). Universality and predictability of technology diffusion [284].

Many technologies grow along S-curves: diffusion is slow, then rapid, then levels off. Forecasting this growth is vital for renewable energy, AI and other technology transitions, but it has been unclear whether technologies follow a single universal process, and past forecasts have proved unreliable. We assemble a database of 120 mature technologies, from canals to mobile phones, and show their S-curve shapes are remarkably universal. Using Bayesian methods and extensive out-of-sample backtesting, we show that a Bertalanffy-Richards process does a good job of fitting the data and its forecasting outperforms popular alternatives. Its point forecasts are typically accurate to within a factor of two, even from a 5% diffusion origin and decades ahead. This gives a validated method to forecast any technology that follows an S-curve, with known accuracy. Our forecasts for solar PV and wind indicate that by 2050 they will supply 13–470 and 4–17 PWh globally each year (90% prediction intervals). Our median estimate for solar in 2050 is about 80 PWh, compared to 31 PWh of global electricity generation in 2024, and roughly the same as all the useful energy the world consumes today. Even the most aggressive IPCC AR6 scenarios are too pessimistic about solar, implying that ambitious climate targets will likely be met faster than widely believed.

Keywords: Technology growth; Bayesian forecasting; S-curves; Energy transition

4.1 Introduction

Technologies fuel economic growth and play a central role in our lives. Early work [7, 8, 285] showed qualitatively that many technologies exhibit "S shaped" growth behavior, in which

growth is initially slow, then rapid, and then slows again, and coined the term S-curve. However, many different functional forms have been postulated for technology growth [4, 14, 50] and previous studies have been unable to strongly narrow the space of possibilities [20], so there is no common agreement [41]. Anticipating technology growth (also called diffusion, adoption or deployment) is potentially very useful for planning and investment, but previous attempts to do so have yielded widely varying and unreliable results; naive S-curve forecasts based on fitting S-curves to historical data tend to be too pessimistic and many past forecasts are clearly wrong in hindsight [31]. While several recent studies apply probabilistic or Bayesian S-curve methods to renewable energy diffusion [69, 114, 242], prior results are based only on a few immature technologies, so there is no ground truth to test what happens in later stages. More recent studies find a handful of mature and similar technologies to solve this problem [113, 286], but have a small sample size and do not backtest out-of-sample against late-stage diffusion data, and other forecasts that have been made are essentially qualitative [260]. As a result, prior to our work here, the precise nature of the process of technology diffusion, including its universality and forecastability, remained highly uncertain.

Why do technologies follow S-curves? A common explanation emphasizes the technology-specific context: Each technology has its own story that depends on idiosyncratic factors such as the effort and competence of firms and individuals, global competition, supply chain stability, industrial policy, serendipity, and geopolitics [287]. For example, the diffusion of solar photovoltaic energy was inhibited by lack of materials during 2000-2010 [288], boosted afterward by policies in many countries such as Germany [289] and China [290], and is now being damped by U.S. industrial policy [291]. Because such factors are hard to anticipate, this has limited usefulness for forecasting.

Another approach is to search for a law governing technology growth. For instance, if $Y(t)$ is the diffusion level, e.g., the number of instances of the technology existing at time

t, consider a growth process of the form

$$\frac{dY(t)}{dt} = r(t)Y(t), \tag{4.1}$$

where $r(t)$ is a growth rate. If $r(t)$ is relatively constant, this is an exponential growth process. There are many reasons to expect exponential behavior during the early stages of growth. On the demand side, exponential growth holds if consumers are more likely to buy a given technology such as a mobile phone if others around them own the technology. On the supply side, it follows from the assumption that the willingness to invest is proportional to the size of the existing market. For example, the confidence that a new railroad project will succeed might increase proportionally to the number of successful projects. There are many possible explanations that predict exponential diffusion in early stages [16, 40, 292].

The characteristic S-curve shape of technology growth occurs because the growth rate $r(t)$ slows down as the market saturates. If many people already own mobile phones, for example, there are fewer people to buy new ones, and as more railroads are built, the number of new commercially viable routes diminishes. The functional form for $r(t)$ depends on the details of the saturation process. A useful analogy is bacterial growth, where $r(t)$ can sometimes be derived from first principles [293]. This depends on factors such as whether or not the food-bearing solution is well-mixed, as in a chemostat, or whether growth can only occur on a boundary, as in a Petri dish filled with agar. (Tumor growth is another interesting example.) For technology diffusion, different mechanisms underlying diffusion such as contagion, social influence or social learning, imply different shapes of S-curves [122, 294]. However, the empirical evidence and studies so far have not resolved whether technology growth obeys a common function $r(t)$ or whether each technology does something different.

Here we substantially resolve these long-standing questions about technology growth by

showing that the growth of technologies is remarkably universal and that this can be used to make useful forecasts of known accuracy. We assemble a much larger set of technologies than previous studies, consisting of mature technologies where the ground truth is known. We show that the shape of technology S-curves is approximately universal, test several candidate stochastic processes, perform extensive statistical testing, and develop a reliable forecasting method that can take advantage of historical diffusion data for any technology to make predictions of known accuracy. Finally, we apply this to solar and wind energy and show that, while many IPCC AR6 scenarios fall within our 90% prediction interval for wind, our median wind forecast lies below the typical 2050 scenario, while solar, in contrast, is likely to grow faster and higher than even the most optimistic scenarios.

4.2 Results

To provide a foundation for our work here we conducted an extensive search, finding a total of 1,362 time series, which we winnowed down to a data set with historical deployment of 120 mature technologies, ranging from canals to mobile phones, as detailed in the Data section. While there is inevitably selection bias, we have done extensive analysis to ensure that our results are robust, as described in the Data and Methods sections.

After testing several different functional forms for S-curves, we find that the Bertalanffy-Richards (B-R) process [13, 61, 64], in which growth declines according to $r(t) = k(1 - (Y(t)/L)^{\beta})$, is a good overall choice (see SI Sections 3.3 and 4.1 for model comparison). Here L is asymptotic deployment, k is the initial growth rate and β is a shape parameter. In this case the solution to equation 4.1 is

$$Y(t) = \frac{L}{\Big[1 + \exp\big(-\beta k(t - t_0)\big)\Big]^{1/\beta}}, \tag{4.2}$$

where t_0 locates the growth process in time [61]. An important special case is the logistic

model, where $\beta = 1$ [12, 14]. Figure 4.1 illustrates the Bertalanffy-Richards process for $\beta = 2/3$, which we find to be a good overall choice, and compares it to the two most widely used models for S curves, the logistic process ($\beta = 1$), and the Gompertz process[11], derived by scaling the Bertalanffy-Richards growth rate $r(t)$ by $1/\beta$ and taking the limit $\beta \to 0$, so that $r(t)$ declines exponentially with time. We also compare to pure exponential growth.

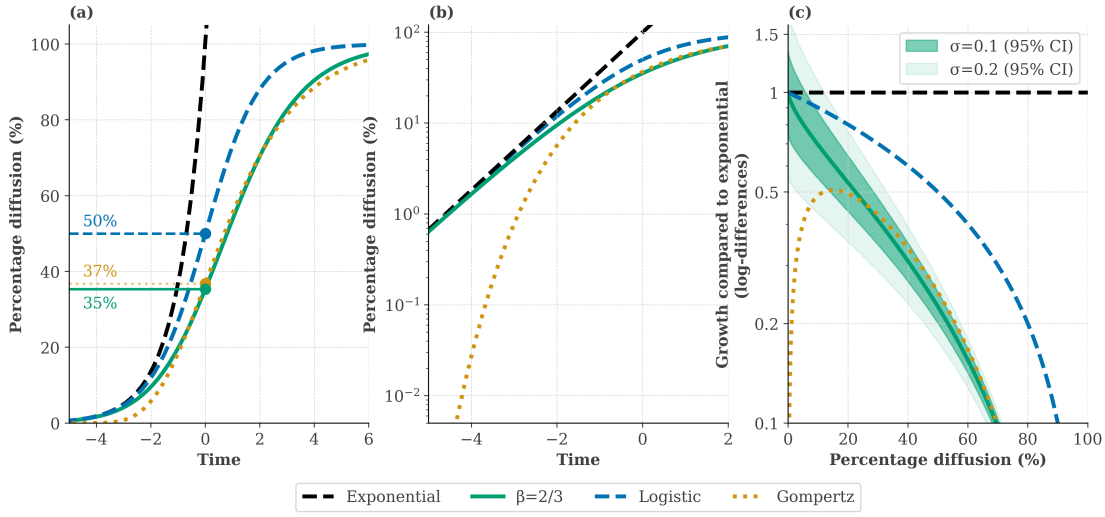


Figure 4.1: Comparison of S-curves to exponential growth. Panel **(a)** shows diffusion vs. time for exponential growth, the Bertalanffy-Richards model with $\beta = 2/3$, the logistic model ($\beta = 1$) and the Gompertz model. Panel **(b)** is the same as (a), but on logarithmic scale. Panel **(c)** shows the ratio of the growth rate for each process compared to exponential growth as a function of diffusion level. Shaded regions are the 95% confidence intervals for different noise levels for the Bertalanffy-Richards model with $\beta = \frac{2}{3}$. The exponential growth rate is chosen to match the initial growth of the logistic process, and the parameters of the S-curve models are chosen so they spend the same amount of time between 10 and 90% adoption.

Technology growth is inherently non-stationary and strongly non-linear. These problems are compounded by the fact that the time series are short and strongly autocorrelated, measurements are noisy and the fluctuations around deterministic behavior are heteroscedastic (meaning the fluctuations vary systematically with time over the growth path).

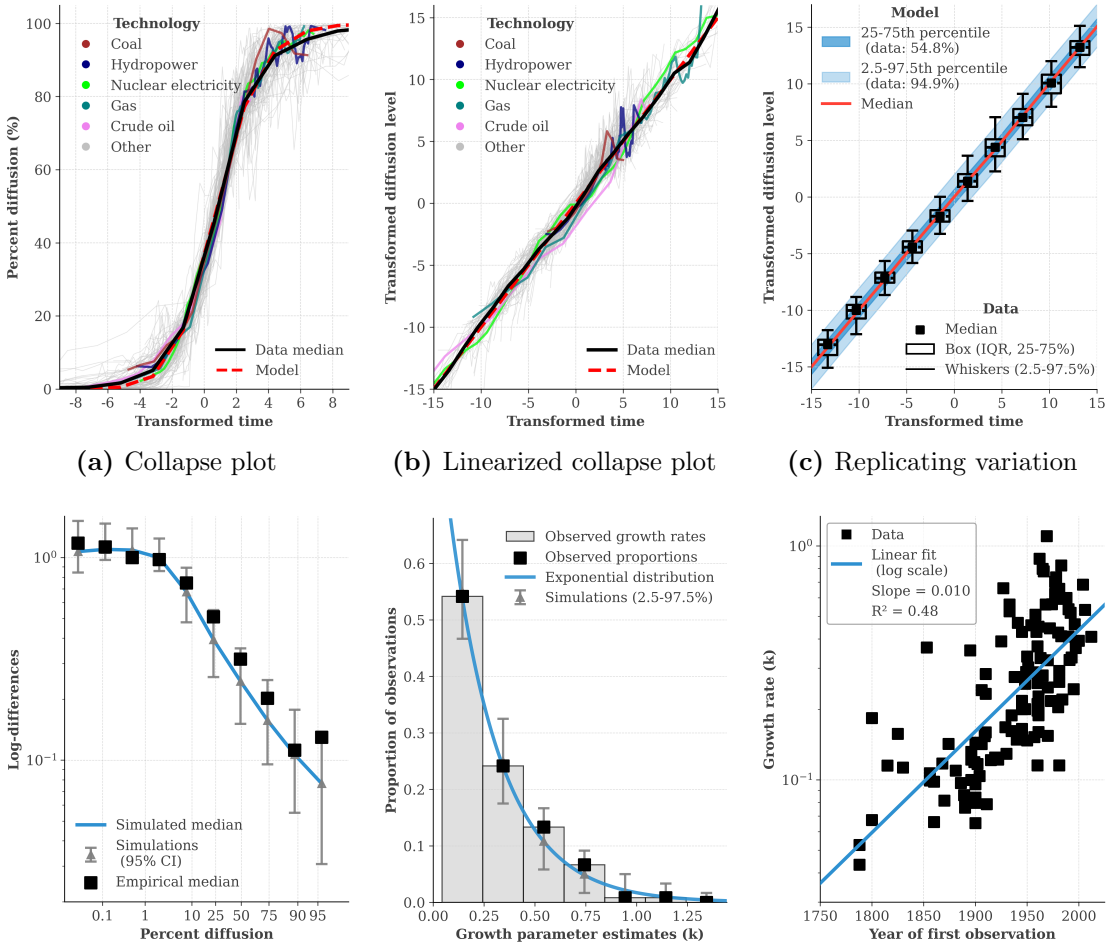
This leads us to recast the deterministic Bertalanffy-Richards process as a stochastic

model of the form

$$Y(t) = \frac{L}{\Big[1 + \exp\big(-\beta k(t - t_0) + \beta\sigma(\epsilon_t + \rho\epsilon_{t-1})\big)\Big]^{1/\beta}}, \tag{4.3}$$

where σ is the noise amplitude and ρ is an autocorrelation parameter. The noise process $\epsilon_t \sim \mathcal{N}(0, 1)$ is normal and IID; because the noise is inside the exponential, the noise process is effectively multiplicative and the resulting distribution for $Y(t)$ is roughly log-normal, which means it is strongly skewed and has relatively heavy tails. The noise process with amplitude σ represents idiosyncratic, technology-specific variations in growth rates, as well as data measurement errors. All of the parameters are potentially technology-dependent, but the evidence for technology-specific variations in β and ρ is weak, and because of the large variance in estimating parameters, for forecasting purposes it is better to assume universal values for β and ρ (see SI Section 3.1).

To compare the shape of the growth trajectories of different technologies, we fit equation 4.3 to each individual technology with $\beta = 2/3$ and $\rho = 0.8$ and rescale each time series in non-dimensional coordinates (denoted by tilde) so that $\tilde{t}_0 = 0$, $\tilde{L} = 1$, and $\tilde{k} = 1$, which makes the location, scale, and initial growth rate of each technology the same. We plot all 120 technologies on top of each other in non-dimensional coordinates and compare to the Bertalanffy-Richards model with $\beta = 2/3$ in Figure 4.2. The shape of the non-dimensional S-curves is remarkably similar, indicating that apart from differences in parameters, the growth processes are very close to each other. The Bertalanffy-Richards model explains 93% of the variance, showing that *to a good approximation the growth paths of different technologies are universal* in the sense that a simple deterministic law explains most of the growth path, at least in hindsight. Technology-specific, idiosyncratic factors may affect the parameters of the growth process and cause deviations from deterministic behavior, but the shape of the growth path is largely the same for all technologies.



(a) Collapse plot **(b)** Linearized collapse plot **(c)** Replicating variation

(d) Growth rate vs. diffusion level **(e)** Distribution of growth rates **(f)** Growth rates over time

Figure 4.2: The universal shape of technology growth paths. Based on fits using equation 4.3 with $\beta = \tilde{\beta} = 2/3$ and $\rho = \tilde{\rho} = 0.8$, panel **(a)** plots the data for each technology in non-dimensional coordinates (denoted by tilde), where $\tilde{t}_0 = 0$, $\tilde{L} = 1$, and $\tilde{k} = 1$ for all technologies. Energy technologies are highlighted. Panel **(b)** shows the same in nonlinear coordinates that make the growth trajectory linear (see SI Section 1.1). Panel **(c)** illustrates how the model captures the heteroskedasticity of the noise by comparing the variation in the data to that expected from the model in linearized coordinates. Boxes are 25% and 75% quantiles with black dots denoting the data median in each box and the red line denoting the model median. Shades of blue indicate 1, 2, and 3 standard deviations based on 1,000 realizations of the model using the fitted parameters for each technology. Panel **(d)** compares the median growth rates $r_i(t)$ for the data to the median growth for a similar simulation of 1,000 realizations of the model. Panel **(e)** shows the distribution of growth rates k for all technologies, which is approximately an exponential distribution shifted by the minimum empirical growth rate ($\min_i k_i \approx 0.04$). Panel **(f)** shows the growth rates k as a function of the year of first observation on a log-scale (excluding *monasteries*, which diffused during the 1100's). Growth rates of technologies increase roughly exponentially over time.

For a clearer view of the variation across technologies, Figure 4.2(b) shows the same

data under a nonlinear coordinate transformation that makes the growth linear in time (see Methods). To show that our model does a good job of capturing the heteroskedasticity of the data, in Figure 4.2(c) we study the cross-sectional standard deviation vs. time in nondimensional coordinates. To understand the expected variation due to the noise process, for each technology we use its fitted parameters to generate 1000 realizations of surrogate data whose length matches the empirical data. The noise amplitude of the empirical data under the linearizing transformation varies little, and is within the expected error bars, indicating that our model, which has a constant noise amplitude over time in the coordinates that linearize the dynamics, provides a good understanding of the heteroskedasticity of the underlying random process. (See SI Sections 3.1 and 3.2 for further quantitative tests and information).

As an alternate test of universality, in Figure 4.2(d) we plot the logarithmic growth rates $r(t) \approx \log Y(t) - \log Y(t-1)$ as a function of the normalized diffusion level $Y(t)/\hat{L}$. This provides a robust test as it removes the dependence on the estimated scale parameter $\hat{L}$ (though as one expects that growth rates are noisier than levels). The match between the model and the data is excellent. In SI Section 3.1 we use the surrogate data method to show that variations in estimated values of $\hat{\beta}_i$ are compatible with the hypothesis that they are statistical errors, indicating that it is a good approximation to use a constant value of $\beta \approx 2/3$.

Our analysis reveals an interesting regularity: As illustrated in Figure 4.2(e), the initial diffusion rates of the technologies have an approximately exponential distribution (SI Section 3.2). Figure 4.2(f) shows that diffusion rates for this sample of technologies are increasing with time, consistent with earlier observations[16, 211, 295]. The fit suggests that the rate of increase is exponential, with the caveat that there may be selection bias. (Our dataset only includes mature technologies that have reached at least 80% diffusion, so slower diffusing new technologies may be excluded). See Data section for more details.

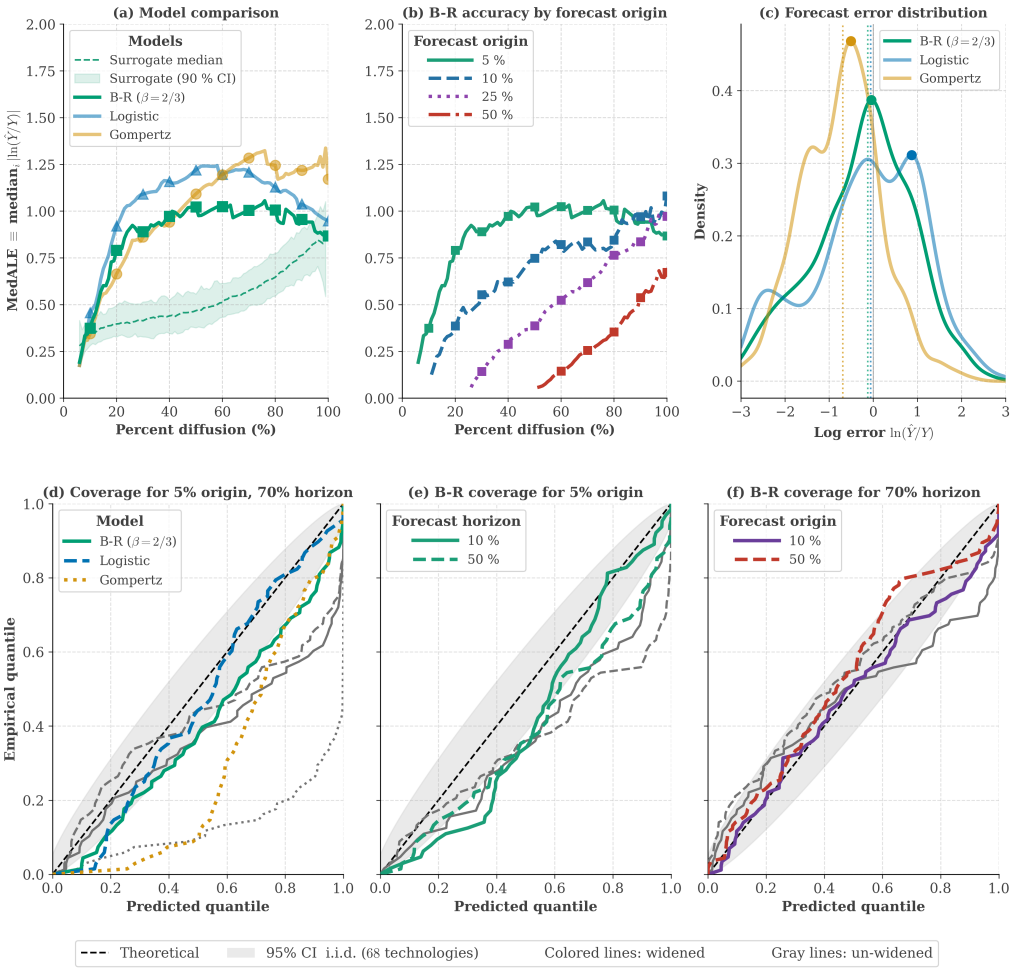


Figure 4.3: Evaluation of forecast accuracy. Panels **(a, b, c)** report point-forecast accuracy and bias for all technologies with at least five training points (yielding 68 technologies at 5% of asymptotic diffusion and 120 at 50%). Panels **(a)** and **(b)** measure point forecast accuracy based on the Median Average Log Error (MedALE) of the point forecasts, where the point forecasts are in turn based on the median of the predictive distribution $f_{t,t+\tau}(\hat{Y}_{t+\tau}|Y_{0:t})$. Panel **(a)** compares the accuracy of forecasts from a 5% diffusion origin for the B-R, logistic and Gompertz models; The teal band is the 5–95% confidence interval using the B-R model to generate surrogate data as described in Methods and the dashed teal line is its median (see also SI Section 1.4). Panel **(b)** compares the accuracy of the B-R model for forecasts made at 5%, 10%, 25% and 50% origins. Panel **(c)** shows the density of the log error $\ln(\hat{Y}/Y)$ for forecasts with a 5% origin and pooling all horizons. Dotted vertical lines mark medians for each model (dots are the modes), making the strong left-skew of the Gompertz model evident. Panels **(d-f)** evaluate distributional coverage by comparing the quantiles of the predictive distribution $f_{t,t+\tau}(\hat{Y}_{t+\tau})$ to the quantiles of the same distribution evaluated on the empirical data, $f_{t,t+\tau}(Y_{t+\tau})$; under perfect coverage this is the uniform distribution and the $q-q$ plot the identity. In each case we compare the coverage with and without the widening adjustment described in the Methods section. Gray lines show the un-widened coverage and colored lines show the widened coverage, with the gray shaded band the 95% pointwise i.i.d. confidence envelope. Panel **(d)** compares the B-R, logistic and Gompertz models at the 5% forecast origin and 70% target, panel **(e)** compares forecasts made with the B-R model at a 5% origin and the 10% and 50% horizons, and panel **(f)** compares B-R forecasts for a 70% horizon at 10% and 50% origins.

The approximate universality of S-curves suggests that historical data is useful for forecasting. We make forecasts with the logistic ($\beta = 1$), Gompertz, and Bertalanffy-Richards ($\beta = 2/3$, which we henceforth call "B-R") models using Bayesian methods to produce probabilistic forecasts $f_{t,t+\tau}(\hat{Y}_{t+\tau}|Y_{0:t})$ for individual technologies, where $\hat{Y}_{t+\tau}$ is a forecast made at *origin* time t for *forecast horizon* τ, based on historical data $Y_{0:t} = \{Y_0, \ldots, Y_t\}$ and $f_{t,t+\tau}$ is the distribution of the forecasts (see Methods and SI Section 1). Our priors for the growth rate k and the noise volatility σ reflect the range of values we have observed in the full-sample analysis (see discussion of in-sample vs. out-of-sample testing in the Methods section).

We make point forecasts based on the median of the posterior predictive distribution $f_{t,t+\tau}(\hat{Y}_{t+\tau}|Y_{0:t})$. To compare different methods, forecast horizons and origins, we score predictive accuracy based on the absolute log error (ALE) of the point forecast, and compare the median ALE ("MedALE") across technologies (see Figure 4.3(a-b); see SI Section 4.1 for the formal definition and sensitivity checks).

Because technologies follow intrinsically different timescales, we measure "time" for each technology based on percent diffusion, $100Y/\hat{L}$. (This is done retrospectively to compare accuracy across technologies, not to make forecasts.) In Figure 4.3(a) we compare MedALE for the logistic, Gompertz and B-R models at the 5% diffusion origin. The B-R model has lower MedALE than the logistic at every forecasting horizon and it is substantially better than Gompertz for horizons beyond 50%. Gompertz has somewhat lower forecasting errors for short horizons, but it is strongly biased, meaning its forecasts are systematically too low (see panel c, and SI Section 4.1, which also includes a comparison with the Bass-model). Because the B-R model has comparable forecasting skill and because of other factors discussed in the appendix, we believe it is overall a good choice. To assess statistical significance we compare these results to a B-R surrogate ensemble (see Methods). The empirical curves sit close to the surrogate 90% band at early diffusion and

exceed its upper bound beyond about 15% diffusion, indicating that none of these models fully capture the underlying dynamics at higher diffusion levels. A surprising feature is that forecasts at intermediate horizons of 60% or so are less accurate than at 100%; we do not fully understand this, but we suspect this is due to several reasons, including model mis-specification and the choice of priors. Notably, this peak becomes a plateau with no clear intermediate maximum when the forecast horizon is measured in years (SI Section 4.1).

In Figure 4.3(b) we compare forecasts of the B-R model made at the 5%, 10%, 25% and 50% diffusion levels. As expected, the errors are substantially smaller for forecasts made at larger diffusion levels. It is striking that errors for forecasts made at 5% diffusion are typically less than 100%, i.e. that the errors are typically within a factor of two of the empirical observations, even for horizons far into the future. While the timescales vary widely, in some cases this corresponds to forecasting horizons of up to 100 years.

Forecasts of unknown accuracy are useless. The fact that we are making probabilistic forecasts allows us to predict the accuracy of our point forecasts, and even better, to estimate the probability of different outcomes. If we forecast that the data will fall in a given range with probability x, do the observations fall within that range x% of the time? We assess this by comparing the forecast distribution $f_{t,t+\tau}$ evaluated on the predictions $\hat{Y}_{t+\tau}$ vs. the real data $Y_{t+\tau}$. More specifically, in Figures 4.3(d-f) we test the *coverage* of the probabilistic model by comparing the quantiles of $f_{t,t+\tau}(\hat{Y}_{t+\tau})$ and $f_{t,t+\tau}(Y_{t+\tau})$ using (Q-Q) plots (SI Section 1.3). For each forecast the probability integral transform (PIT) is the forecast CDF evaluated at the realized value. It is uniform when forecasts are calibrated, so its empirical CDF follows the diagonal. The gray bands quantify the uncertainty expected given the number of independent data points available. Figure 4.3(d) compares the three models for forecasts made at 5% diffusion for the 70% diffusion horizon. The coverage of the B-R model forecasts is far better than that of the Gompertz model

and marginally better than that of the logistic model. The Gompertz model's forecasts are extremely overconfident and strongly biased. Figure 4.3(e) shows how the coverage varies for forecasts made with the B-R model at the 5% origin, for several different horizons. The empirical coverage is good and mostly within the confidence intervals. Figure 4.3(f) compares different forecast origins. The B-R forecasts are somewhat overconfident in the upper tail, so henceforth we adjust the predictive distribution to compensate for this. This corrects the coverage problems, as described in Methods and SI Section 1.2. We use this for the remaining results.

We apply the method we have developed here to a problem of pressing interest, making forecasts for solar photovoltaic and wind energy, which are not mature technologies and are not in our original dataset (Figure 4.4). To check that the procedure is trustworthy for these technologies, we backtest it on the solar and wind time series themselves (SI Section 5.1) and stress-test it on deliberately adversarial energy cases, including supply shocks and abrupt regulatory slowdowns (SI Section 4.2). The diffusion of solar PV is currently still almost indistinguishable from noisy exponential growth, whereas wind growth rates appear to be in the early stages of slowing down. The first panel (a) shows our point predictions compared to the 95% prediction interval from an exponential growth model (a geometric random walk with drift with autocorrelated noise).

Our adjusted posterior for the asymptote L implies that solar PV is currently between 0.3 and 14% of its eventual saturation (5–95% interval, median 2.2%) and wind at 12–63% (median 43%). (See SI Sections 5.1 and 5.2). The uncertainty for solar is particularly wide; this is because solar is still close to exponential, as noted above, which makes it difficult to know when it will begin to fall off of exponential growth. The sharper lower bound corresponds to the case where this happens more or less immediately, whereas there is high uncertainty for the upper bound because it could continue approximate exponential growth for an uncertain time into the future. (This is the generic behavior for any technology still

within the noise band for exponential growth).

Even when L is hard to infer, as for solar and wind today, we can test whether a stated growth pathway is consistent with S-curve dynamics. To do this we fix L, estimate the other parameters from the historical data, and compute the probability of reaching each benchmark by its target year. Figure 4.4(b) shows this for the IEA 1.5°C net-zero scenario [296, 297]: 5.5 TW by 2030, 12.5 TW by 2035, 17.8 TW by 2040, 24.2 TW by 2050 (for comparison global solar was 2.2 TW in 2024 [298]). The four target years imply somewhat different effective asymptotes, ranging from 20-26 TW. The IEA pathway is unlikely if $L < 25$ TW and is met ahead of schedule if $L > 25$ TW. We show that 25 TW sits at the low end of asymptotes consistent with the historical data capacity forecast (See SI Section 5.2 and discussion below in terms of PWh/yr).

Panel 4(c) compares the predictive prior distribution that we use to initialize the Bayesian inference to the predictive posterior at three different times. This illustrates the heavy tails predicted for high solar diffusion and shows that even though the prior strongly weights low diffusion levels, the posterior strongly rejects this.

Panels (d-f) show our main results. We compare our distributional forecast, shown as the shaded fan, to the distribution of the 1,837 IPCC AR6 scenarios, which we summarize as box-and-whisker plots for each decade (for more detail see SI Section 5.2). The prognoses for wind and solar are very different. The forecasted median growth of wind between now and 2050 is a factor of over 2, from just over 2 PWh/yr in 2024 to about 5 PWh/yr, with a 90% prediction interval of 4–17 PWh/yr. (Note, though, that the upper tail is fat – the 97.5th percentile in 2050 exceeds 40 PWh/yr). In contrast, the median growth of solar between now and 2050 is a factor of over 40, from 1.9 PWh/yr in 2024 to about 80 PWh/yr, with a much wider 90% prediction interval of 13–470 PWh/yr. This median of about 80 PWh/yr is roughly the same as all the useful energy the world consumes today. For comparison, total global electricity generation in 2024 was 31 PWh [210].

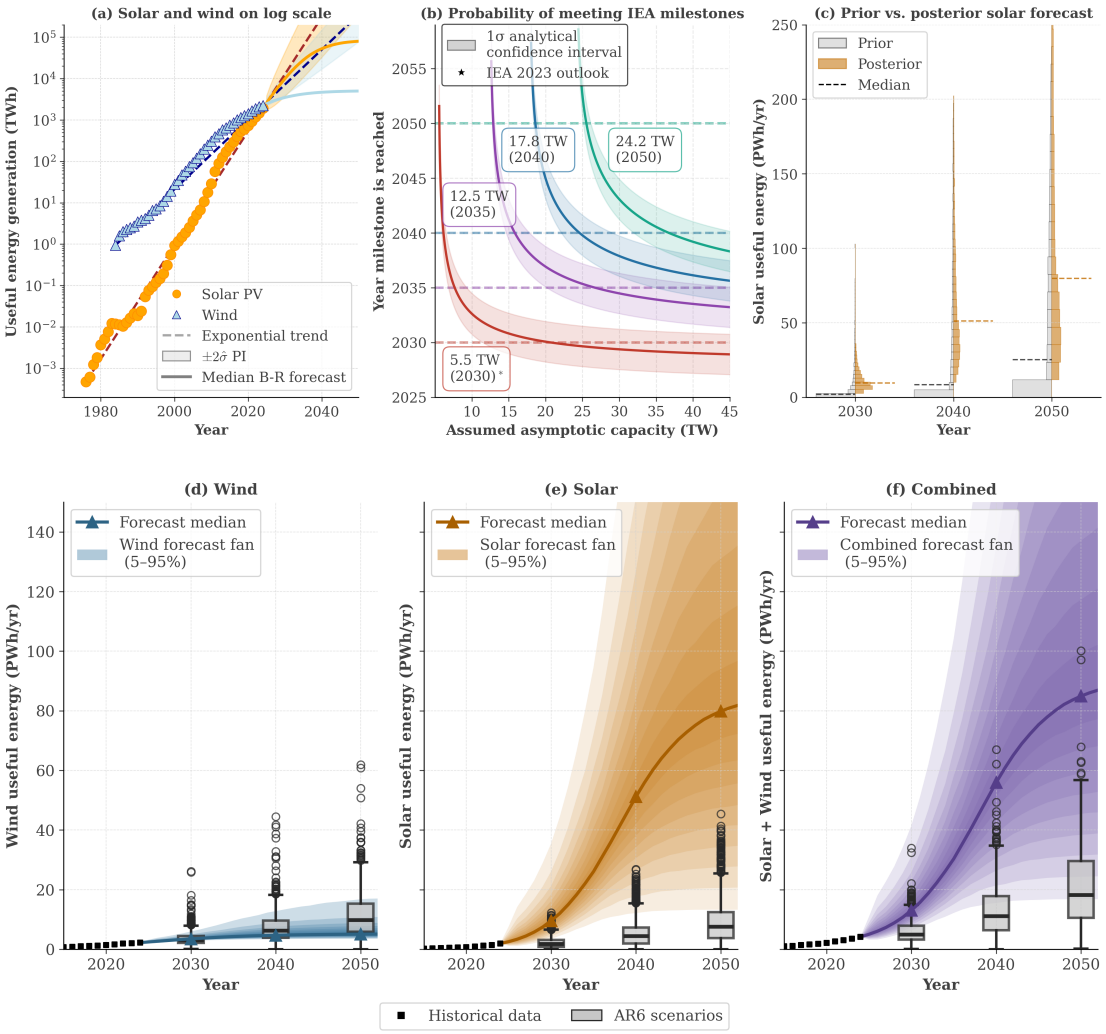


Figure 4.4: Forecasts for solar and wind energy and comparison to the AR6 IPCC scenarios. Panel **(a)** compares their historical diffusion and our median S-curve based forecast to a forecast based on exponential growth (shaded areas are the $\pm 2\hat{\sigma}$ prediction intervals, roughly 95%, using the method from [30]). Panel **(b)** plots the probability that solar reaches each IEA net-zero milestone (2030, 2035, 2040, 2050) by its target year, conditional on a prior for the asymptote L. Solid colored lines are medians. Panel **(c)** plots the posterior predictive distribution vs the prior predictive distribution for solar after applying the widening transform of SI Section 1.2. Panels **(d, e, f)** compare our distributional forecast, shown as the shaded fan (the 5–95% prediction interval), to the IPCC AR6 scenarios. The AR6 scenarios at each decade are summarized as standard box-and-whisker plots (box = interquartile range, whiskers = 1.5×IQR, with outliers beyond the whiskers shown as circles). Panel **(d)** is wind, panel **(e)** is solar, and panel **(f)** is solar and wind combined.

A comparison to IPCC AR6 scenarios shows that they are typically slightly too optimistic for wind and much too pessimistic for solar. The median forecast for wind is roughly equal to the typical IPCC scenario in 2040 and substantially below it in 2050. In contrast, the median solar forecast is a factor of about 10 higher in 2050 (80 vs 7.7

PWh/yr), and *even the most optimistic IPCC forecast* (out of the 1837 in the sample) *is at the pessimistic end of likely solar deployment.*

These results support fast net-zero energy transition pathways [29] that do not rely on negative emission technologies [68]. This is supported by statistically tested cost forecasting methods, which use Moore's law (which is unconditional on deployment) to predict a median LCOE cost for solar of \$8/MWh in 2050 vs. \$35/MWh for wind [31]. This provides independent supporting evidence that solar is likely to be deployed much more and much faster than wind

4.3 Discussion

Our approach has important limitations, which we discuss in full in SI Section 1.5. The model is based only on the time series of past diffusion and doesn't reflect important auxiliary variables such as cost and competition. Because the forecasts are unconditional, they cannot be used to make predictions conditional on a change of policy. Nonetheless, our forecasts allow for the possibility that policies accelerating technology diffusion, such as subsidies, could alter the parameters of the process and shift the realized growth path to the upper end of the forecast distribution, while policies that hinder growth, such as taxes, could shift it to the lower end. Data limitations introduce potential selection bias toward successful technologies. Treating technologies that compete in a single growing market independently may not sufficiently account for competition, or yearly market-growth after the technology market share has stabilized. Finally, while we have tested our methods out of sample as much as possible, we used some in-sample information, specifically to discover useful universal parameters $\beta = 2/3, \rho = 0.8$, for testing priors and for adjusting the width of the predictive distribution.

Despite these limitations, our study here goes well beyond earlier work by using a large

sample of mature technologies as ground truth, by understanding and accurately modeling the heteroskedasticity of the noise process, employing Bayesian methods, performing systematic out-of-sample statistical testing, and showing that technology S-curves are approximately universal, meaning they all have a very similar shape and therefore follow a similar growth process. This allows us to predict future technology growth, and to quantify the accuracy of our predictions. While we have not attempted here to explain S-curve dynamics from first principles, our results strongly constrain past and future attempts to do so – whatever the process is, it must yield dynamics that are close to the Bertalanffy-Richards process with $\beta = 2/3$ that provides a good fit to the data. As we have shown, the widely used Gompertz model produces strongly biased forecasts with poor coverage. However, as we show in SI Section 3.3, averaging across technologies with heterogeneous growth rates and timings makes the pooled trajectory look Gompertz-like even when the individual series are logistic or B-R, so aggregate fits can incorrectly favor the wrong model. We also tested the Bass model, which performed poorly (SI Section 4.1), but we do not rule out mechanism-based accounts such as innovation and imitation, cost feedback, or policy, which contribute to S-curve growth in many historical series. As the case of solar and wind demonstrates, technology forecasts can be very useful in guiding investment and planning for emerging technologies. Although we have emphasized energy technologies here, the methods developed can be applied to any technology, such as different methods of AI, and there are many applications that remain to be explored further.

4.4 Methods

We use empirical Bayesian inference [136, 144] to generate forecasts and check that they are well-calibrated by extensive backtesting (making forecasts as if we were in the past and comparing the forecast to future observations). We perform extensive sensitivity analysis

to test the robustness of our results, comparing to different S-curve models, varying priors, and varying the parameters β and ρ to test their technology-invariance. Further details of our method and out-of-sample robustness checks are given in SI Sections 1 and 4.

We use the surrogate method to test for statistical significance, comparing the empirical results to an idealized situation in which each technology obeys a true $\beta = 2/3$ Bertalanffy-Richards process with parameters matching each technology. To assess the median errors in Figures 4.3(a-b), we generate 100 realizations of the process for each technology using least-squares fits of the model to the full observed series (SI Section 1.4) and measure the MedALE on this surrogate data. For the Q-Q plots in Figures 4.3(d-f) we compare to an i.i.d. uniform null hypothesis.

We have done out-of-sample testing as much as possible, using only the data that would have been available for real forecasts. However, given the nonlinearity of the S-curve model and the difficulty of parameter estimation based on short data series, some components of the estimation are in-sample. In particular, we have used common values for all technologies for the autocorrelation parameter $\rho = 0.8$ and the shape parameter $\beta = 2/3$ – these values are based on an analysis of the full time series for all technologies. We have also done extensive testing for different methods of formulating priors. Because this analysis is based on the full ensemble of 120 technologies, and is not technology-specific, we believe that the level of overfitting is relatively small. It would of course be useful to test these methods on even more technologies.

Because we find that our out-of-sample forecasts are less accurate than they would be if the data obeyed the idealized model of equation (3), for our application to solar and wind we widened the probabilistic forecasts $f_{t,t+\tau}(\hat{Y}_{t+\tau}|Y_{0:t})$ with a median-preserving recalibration (SI Section 1.2). The amount of widening is fitted on the 120-technology backtest and is robust (SI Section 4.1). Refitting it with each technology left out one-at-a-time changes the widening negligibly, so it is not driven by any single technology,

and the larger, less-filtered 154-technology dataset needs a similar amount of widening. B-R needs less widening than the logistic and Gompertz models and its widening is more robust across the 120- and 154-technology sets. For example, the logistic results vary by quite a lot when calibrating the widening parameter on the 120 vs. 154 technology dataset while B-R does not. Surrogate data from a correctly-specified model (SI Section 1.4) needs virtually no widening, so this method corrects a genuine misfit between the model and the data rather than simply inflating uncertainty.

In many cases the potential market size for technologies grows with time, which complicates forecasting. For example, since about 1960, total energy usage has grown at roughly 2% per year. If this pattern continues into the future we should expect the asymptote of the S-curves of energy technologies to reflect this growth rate, in contrast to our assumption of a constant asymptote in equation 4.2. We have experimented with three methods for addressing this problem: (1) Using energy share rather than total energy consumption; (2) explicitly taking this into account by using a modified growth process, analogous to a bacteria culture whose food supply is increasing with time. (3) ignoring the problem and using a constant asymptote. In our forecasts for solar and wind we have opted for (3) on the grounds that it is consistent with what we did for all the other technologies and because it only becomes a serious problem for forecast horizons beyond 2050. This is because the current exponential growth rates of roughly 30% per year for solar and 20% per year for wind are large in comparison to a projected long-range 2% growth rate of total energy usage (which is uncertain in any case). We backtest these approaches on mature energy technologies in SI Section 4.2, and discuss the solar and wind forecasts in SI Section 5.2.

4.5 Data

We collected 1362 time-series from academic papers [15, 29, 35, 49], open-source databases [210, 276, 299–301], government reports [234, 296], and industry publications [302, 303], then filtered to 120 distinct technologies. To provide ground truth, we only consider mature technologies, i.e., those that are within at least 80% of their estimated asymptotic level of deployment. This allows us to test our methods by pretending to be in the past and comparing forecasts to real data where the answer is known. The data presented here is strongly limited by availability, and there are many important technologies missing. Many of our time series have problems such as missing data and measurement errors. While we included as many technologies as possible, there is inevitable bias toward historically successful technologies. To address these problems, we have done extensive sensitivity analysis that indicates our results are robust. In particular, we tested our filtering procedures by repeating the analysis on a less-filtered dataset with 154 technologies and found that the results are similar. Given the measurement errors in the data, our results here provide a lower bound on predictability in the sense that cleaner data should make our forecasts more accurate. However, selection bias could make forecasts for technologies not in our data set less accurate. Better and more data across all growth stages would strengthen future analysis. Further details are in SI Sections 2 and 3.

Supplementary Material

4.6 Mathematical modeling of technology diffusion

This section introduces and motivates the use of S-curves, providing an overview of their applications and the model we use to make well-calibrated forecasts.

Background The energy transition is a technological transition [205, 304], characterized by phasing out fossil fuels and adopting renewables. Technological adoption is often modelled as an *S-curve*. Throughout this SI, we use the term for any sigmoidal $Y(t)$ describing cumulative adoption — for example, the logistic, Gompertz, and Bertalanffy-Richards forms introduced below (Eqs. (4.5)–(4.6)). Technological shifts occur throughout history due to necessity or innovation [47]. Examples include shifts from horses to tractors in agriculture, diesel locomotives replacing steam engines, and transitions from wood to coal to oil to renewables for energy production. Understanding historical diffusion dynamics across industries helps us model and forecast renewable energy adoption.

Technology diffusion can be measured in stocks or flows. The flow measures production or sales of a technology per unit time—for example, the number of mobile phones sold in a given year. Technology stocks measure the *cumulative* production of a technology, minus retirements—for example, the number of mobile phones currently in operation. Stocks are typically the preferred measure for modeling technological diffusion [8, 40].

Griliches introduced S-curves to model technology stocks in 1957 [8]. Researchers have since used S-curves extensively [7, 36, 82, 305] and developed many S-curve models with increasing complexity, often introducing additional parameters to improve fits [19, 41, 306–309].

4.6.1 S-curve models: The role of symmetry

Consider the generalized logistic differential equation.

$$\frac{dY(t)}{dt} = kY(t)^{\alpha}\left[1 - \left(\frac{Y(t)}{L}\right)^{\beta}\right]^{\gamma}, \tag{4.4}$$

where $\frac{dY(t)}{dt}$ describes the technology flow, $Y(t)$ is the stock at time t, L is the asymptotic stock as $t \to \infty$, and k is the *intrinsic growth rate*. Throughout this SI, *growth rate* refers to the log time-derivative $d\log Y/dt$, with empirical analogue $\Delta \log Y_{i,t} = \log Y_{i,t} - \log Y_{i,t-1}$ (see Section 4.14.1); k is the asymptotic value of this rate as $t \to -\infty$. The parameters α, β, and γ determine the shape of the S-curve. In particular, β controls the steepness and affects the symmetry of the function. When $\alpha = \gamma = 1$, Eq. (4.4) reduces to the Bertalanffy-Richards (B-R) curve [61–63], described as

$$\frac{dY(t)}{dt} = kY(t)\left[1 - \left(\frac{Y(t)}{L}\right)^{\beta}\right]. \tag{4.5}$$

When $\beta > 1$, the inflection point occurs after 50% adoption, and when $\beta < 1$, it occurs before 50% adoption. The analytical solution to Eq. (4.5) is

$$Y(t) = \frac{L}{\left(1 + \exp\left(-\beta k(t - t_0)\right)\right)^{1/\beta}}. \tag{4.6}$$

The logistic ($\beta = 1$) and Gompertz (divide by β and take $\beta \to 0$ limit) curves are special cases [14]. Throughout, t_0 is a *location parameter* (typically in years) that shifts the curve along the time axis. For the symmetric logistic and the Gompertz limit, t_0 coincides with the inflection point: the time of maximum growth, or for technology applications, the year of peak annual addition. For the asymmetric B-R curve with $\beta \neq 1$, the inflection shifts to $t_0 - \log(\beta)/(\beta k)$ (Section 4.13), so t_0 remains a curve offset but no longer marks peak growth.

The logistic curve. When $\beta = 1$, Eq. (4.5) reduces to the symmetric logistic curve

$$\frac{dY(t)}{dt} = kY(t)\left[1 - \frac{Y(t)}{L}\right], \tag{4.7}$$

first introduced by Verhulst [12] in 1838 to model population growth. Integrating by separation of variables, we obtain the analytical solution

$$Y(t) = \frac{L}{1 + A\exp(-kt)}, \tag{4.8}$$

where A is a constant depending on the initial condition. The parameter A defines a shift in the curve relative to the point of maximum growth. Let A take the form

$$A = \exp(kt_0), \tag{4.9}$$

so that maximum growth occurs exactly when $t = t_0$, at $Y(t_0) = L/2$, recovering the location-parameter interpretation introduced above for the logistic case.

Substituting Eq. (4.9) into Eq. (4.8), we derive

$$Y(t) = \frac{L}{1 + \exp(-k(t - t_0))}. \tag{4.10}$$

We refer to this as the logistic curve. Applications include disease modeling [2, 74], predicting gender from weight and height [58], and ecology and cancer modeling [40, 44, 49, 59].

The Gompertz curve. Starting from Eq. (4.5), dividing by β and taking the limit as $\beta \to 0$, we obtain the Gompertz curve,

$$Y(t) = L\exp(-\exp(-k(t - t_0))). \tag{4.11}$$

Gompertz introduced this curve in 1825 [11] to model population dynamics. Researchers have since applied the model extensively to plant growth [84, 86], fish growth [65], tumor growth [3], bacterial growth [90], and other applications [83, 85, 87–89, 133].

Comparing the three S-curves. S-curves can overfit, yielding inconsistent out-of-sample performance [31]. We therefore fix $\beta = 2/3$ as a technology-invariant value, balancing the early-exponential property of the logistic with the asymmetry of the Gompertz without adding free parameters.

To characterize growth behavior, we compute log-growth rates $\frac{d}{dt}\log Y(t)$. For the B-R curve (Eq. (4.6)),

$$\frac{d}{dt}\log Y(t) = \frac{k}{1+\exp\left(\beta k(t-t_0)\right)}, \tag{4.12}$$

where setting $\beta = 1$ yields the logistic case and $\beta = 2/3$ yields our B-R case, with limiting behavior

$$\lim_{t\to-\infty}\frac{d}{dt}\log Y(t) = k, \qquad \lim_{t\to\infty}\frac{d}{dt}\log Y(t) \sim k\,\exp(-\beta k(t-t_0)). \tag{4.13}$$

For the Gompertz curve, the log-growth rate takes a different form:

$$\frac{d}{dt}\log Y^{(g)}(t) = k\exp(-k(t-t_0)), \tag{4.14}$$

with limiting behaviour

$$\lim_{t\to-\infty}\frac{d}{dt}\log Y^{(g)}(t) = \infty, \qquad \lim_{t\to\infty}\frac{d}{dt}\log Y^{(g)}(t) \sim k\,\exp(-k(t-t_0)). \tag{4.15}$$

The B-R and logistic curves exhibit exponential growth at early diffusion (constant log-growth rate k), while the Gompertz curve has no constant early-growth regime. At late diffusion, all three decay exponentially, at rate βk for the B-R curve (rate k when $\beta = 1$)

and rate k for the Gompertz curve. In particular, the logistic ($\beta = 1$) and Gompertz have identical late-stage decay.

Figure 4.5 accompanies main paper Figure 1. It shows the B-R curve ($\beta = \frac{2}{3}$), logistic ($\beta = 1$), and Gompertz curves in levels (Figure 4.5a), log-differences against time (Figure 4.5b), and log-differences against percentage diffusion (Figure 4.5c). All curves use $L = 100$ and $t_0 = 0$. We set $k = 1$ for the B-R curve and adjust the logistic ($k = 0.749$) and Gompertz ($k = 0.526$) growth rates so that all three curves spend the same time traversing from 10% to 90% diffusion.

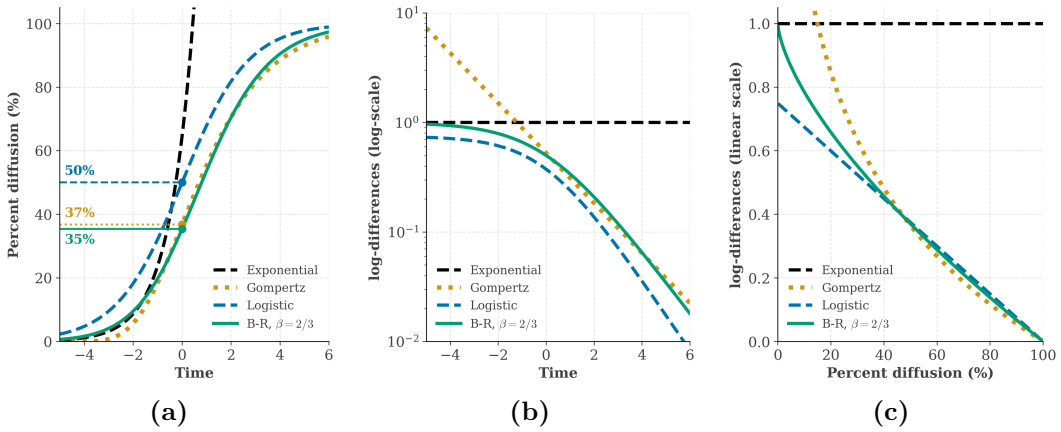


Figure 4.5: Comparison of the B-R curve with $\beta = \frac{2}{3}$ (green) to the logistic ($\beta = 1$, blue) and Gompertz (red) curves. Exponential growth (black) is plotted for reference. All curves use $L = 100$ and $t_0 = 0$; growth rates are $k = 1$ (B-R), $k = 0.749$ (logistic), and $k = 0.526$ (Gompertz), chosen so that time between 10–90% diffusion is the same across curves. **(a)** S-curves in levels; percent diffusion at $t = 0$ is highlighted. **(b)** Log-differences on log-scale against time. **(c)** Log-differences on linear scale against percent diffusion.

The three curves differ in two key respects. (1) *Inflection point*: for the logistic and Gompertz curves, $t = t_0$ defines the inflection point (point of maximum growth), occurring at exactly 50% and approximately 37% diffusion, respectively. For any asymmetric B-R curve ($\beta \neq 1$), the inflection point shifts to $t = t_0 - \frac{\log \beta}{\beta k}$. At $t = 0$, the B-R curve with $\beta = 2/3$ reaches approximately 35% adoption and is still accelerating, while the logistic and Gompertz curves begin to decelerate. For $k = 1$, $\beta = 2/3$, the B-R curve continues accelerating until $t \approx 0.6$, corresponding to approximately 46% adoption (Figure 4.5a).

(2) *Early growth*: under the log transform, the logistic curve grows at an approximately constant rate k for $t \ll t_0$ (approximate exponential growth) and then decays to zero, while the Gompertz curve's log-growth decays exponentially throughout. Figure 4.5b plots exponential growth at rate $k = 1$ (black) for reference. The logistic curve runs parallel to, but below, the exponential line because its adjusted growth rate is $k = 0.749 \neq 1$; this parallel structure confirms approximate exponential growth at early diffusion levels. The Gompertz curve, by contrast, decreases linearly with time on the log-scale. Figure 4.5c shows that the logistic curve's growth decreases linearly with percentage diffusion, while the B-R curve's growth decreases more gradually, lying between the logistic and Gompertz. All three curves show similar exponentially decaying growth rates in later stages, evident from the overlapping lines at high diffusion in Figure 4.5c.

Empirically, neither the logistic nor Gompertz model dominates. Comparative studies find results vary by technology, time period, and domain [69, 92–98]. Comin et al. [91] find the logistic curve systematically fails to capture early diffusion dynamics; we later show this also applies to the Gompertz curve under ensemble averaging. Each model captures one desirable property but not the other: many technologies exhibit approximate exponential growth at early diffusion levels [4, 29, 31, 41, 310] (logistic, not Gompertz), and many diffusion processes are asymmetric [20, 40, 50] (Gompertz, not logistic). Meade and Islam [20] compared 22 S-curve formulations and found curve symmetry matters more than functional form. The B-R curve combines both properties: it preserves exponential growth at early diffusion levels while allowing asymmetry [113, 311].

4.6.2 Technology adoption

The S-curve has three phases: an *initial* slow-growth phase as the technology enters the market, a *growth* phase of rapid expansion, and a *saturation* phase as adoption plateaus [40, 43, 49].

Phase-out of technologies. As technologies mature, better substitutes often emerge [40, 123]. The mature technology then declines, entering a fourth *death phase* as the substitute takes substantial market share. We do not model phase-out and therefore omit the death phase from our analysis. For the first three phases, the S-curve formulation suffices to describe and predict adoption.

4.6.3 Stochastic noise model

The Bertalanffy-Richards curve in Eq. (4.6) is deterministic. We extend it to a stochastic model below.

Transforming the S-curve. Fisher and Pry [37] introduced a transformation that linearises the logistic curve, later extended by Bhargava [46] with a time-dependent growth parameter. We generalise this transformation to the B-R curve and define the generalised Fisher–Pry (GFP) transformation and noise model:

$$f(Y_{i,t}; L_i, \beta) \equiv \frac{1}{\beta} \log\left(\frac{L_i^\beta}{Y_{i,t}^\beta} - 1\right) = -k_i(t - t_{0,i}) + \sigma_i \left(\epsilon_{i,t} + \rho\epsilon_{i,t-1}\right), \tag{4.16}$$

where ρ is a technology-invariant MA(1) correlation parameter, σ_i is the technology-specific noise standard deviation (i.e. the standard deviation of the noise in the GFP transformed space), and $\epsilon_{i,t} \sim \mathcal{N}(0,1)$ are independent noise terms. The function $f(Y_{i,t}; L_i, \beta)$ depends on the observed stock $Y_{i,t}$, the asymptote L_i, and the (a)-symmetry parameter β.

Then, the technology stock $Y_{i,t}$ follows the stochastic relationship

$$\begin{aligned} Y_{i,t} &= \frac{L_i}{\left(1 + \exp\left(-\beta k_i(t - t_{0,i}) + \beta\sigma_i(\rho\epsilon_{i,t-1} + \epsilon_{i,t})\right)\right)^{1/\beta}} \\ \epsilon_{i,t} &\sim \mathcal{N}(0,1) \quad \forall t, \end{aligned} \tag{4.17}$$

where $\{L_i, k_i, t_{0,i}, \sigma_i\}$ are technology-specific parameters and $\{\beta, \rho\}$ are technology-invariant

for parsimony (justified in Section 4.13).

Noise structure in logit space. The Fisher–Pry noise model is additive in transformed (logit) space. Since we observe stocks in levels, the variance of $Y_{i,t}$ depends on how GFP-space perturbations propagate to the level scale. Consider the logistic case ($\beta = 1$) with i.i.d. additive noise in logit space

$$Y(t) = \frac{L}{1 + \exp(-k(t - t_0) - \sigma\epsilon_t)},$$

where the GFP value is $f \equiv -\log(L/Y - 1) = k(t - t_0) + \sigma\epsilon_t$. The derivative is

$$\frac{\partial Y}{\partial f} = \frac{\partial}{\partial f}\left[\frac{L}{1 + e^{-f}}\right] = \frac{Le^{-f}}{(1 + e^{-f})^2} = \frac{Y(L - Y)}{L}. \tag{4.18}$$

Thus, the noise variance in levels is proportional to $Y^2(L - Y)^2$, which equals zero at $Y = 0$, reaches maximum at the inflection point $Y = L/2$, and decreases to zero as $Y \to L$. For the B-R curve with $\beta = 2/3$, the derivative $\frac{\partial Y}{\partial f}$ has the same qualitative structure: zero at boundaries, maximum at the inflection point.

We assume that the noise in technology flows follows a moving-average MA(1) process. Under MA(1), if the stock at time t exceeds its deterministic value, the stock at $t + 1$ is also more likely to exceed the deterministic prediction.

4.6.4 Normalizing S-curves

To create robust models, we pool data from various technologies. Technologies have unique adoption trajectories with different growth rates and saturation levels. Despite this, we can normalize technology-specific S-curves to analyze cross-technological regularities.

Collapsing S-curves. Let $Y_i(t)$ be the time-series stock data of technology i, and assume different technologies follow the same law with different parameters. We define the normalized technology stock as

$$Y_i^{(n)}(t) \equiv \frac{Y_{i,t}}{L_i}, \tag{4.19}$$

and normalized time as

$$\tau_i \equiv k_i(t - t_{0,i}). \tag{4.20}$$

Next, we normalize the noise by dividing it by its standard deviation. That is, the normalized residuals are defined as

$$\frac{\log\left[Y_i^{(n)}(t)^{-\beta} - 1\right] + \beta\tau_i}{\beta\sigma_i\sqrt{1+\rho^2}} = u_{i,t} \tag{4.21}$$

Then the normalized Bertalanffy-Richards curve becomes

$$Y_i^{(n)}(t) = \frac{1}{\left(1 + \exp\left(-\beta\tau_i + \beta\sigma_i\sqrt{1+\rho^2}u_{i,t}\right)\right)^{1/\beta}}, \tag{4.22}$$

where the curve still contains technology-specific noise terms. Crucially, $\frac{1}{\beta}$ remains in Eq. (4.22), allowing us to assess whether technologies collapse onto a similar asymmetric S-curve or exhibit unique asymmetries.[1]

4.6.5 Model limitations

Our model assumes independent S-curve diffusion for each technology and excludes declining phases. In reality, technologies emerge, compete, and phase out dynamically [37, 40, 123], and markets expand. The framework here is therefore a univariate benchmark; multivariate distributional forecasts for competing technologies in evolving markets remain for future

[1]For model parsimony and to reduce overfitting risk, we fix $\beta = 2/3$ as a universal constant rather than estimating technology-specific β_i values. Section 4.13 discusses this. We show the universal parameter $\beta = \frac{2}{3}$ replicates the data well in Sections 4.14 and 4.16.

work. S-curves also describe *how* technologies diffuse, not *why*; mechanistic explanations are likewise outside our scope.

4.7 Our Bayesian forecasting procedure

Here we describe parameter estimation and forecasting for the stochastic model of Eq. 4.17, using Bayesian inference to generate distributional forecasts.

Notation conventions. We adopt standard Bayesian notation throughout. $\pi(\cdot)$ denotes probability distributions over parameters: $\pi(\theta)$ is the prior and $\pi(\theta \mid Y)$ is the posterior. $p(\cdot)$ denotes probability densities for observed data and noise: $p(Y \mid \theta)$ is the likelihood and $p(Y)$ is the marginal likelihood. Throughout this section, $Y_{i,t}$ denotes the observed adoption level for technology i at time t, with vector notation $Y_{i,1:T} = (Y_{i,1}, \ldots, Y_{i,T})$ for the full time series. The transformation f maps raw data to transformed variables. We denote the first observation time by 1 to avoid confusion with the inflection-point parameter t_0.

4.7.1 Introduction to Bayesian inference

Bayesian inference is based on Bayes's theorem, which states

$$\pi(\theta \mid Y_{1:T}) = \frac{p(Y_{1:T} \mid \theta)\pi(\theta)}{p(Y_{1:T})}, \tag{4.23}$$

where $\pi(\theta \mid Y_{1:T})$ is the posterior distribution of parameters θ given observed data $Y_{1:T}$, $p(Y_{1:T} \mid \theta)$ is the likelihood, that is, the density of the observed data $Y_{1:T}$ viewed as a function of θ, $\pi(\theta)$ is the prior distribution of θ, and $p(Y_{1:T})$ is the marginal likelihood

(evidence). Since $p(Y_{1:T})$ serves as a normalizing constant, Eq. (4.23) can be rewritten as

$$\pi(\theta \mid Y_{1:T}) \propto p(Y_{1:T} \mid \theta)\pi(\theta). \tag{4.24}$$

Given the likelihood $p(Y_{1:T} \mid \theta)$ and prior $\pi(\theta)$, the posterior $\pi(\theta \mid Y_{1:T})$ is determined up to a normalising constant.

4.7.2 Estimating parameters of our stochastic model

We apply Bayesian inference to obtain posterior distributions for the parameter vector $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$ for each technology i.

Log-likelihood of the observed data. The B-R DGP of Eq. (4.16) specifies additive Gaussian noise in the GFP-transformed space. Define the centred latent variable

$$x_{i,t} \;\equiv\; f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i}),$$

which is Gaussian with MA(1) covariance under the DGP (Section 4.6); its closed form in terms of $Y_{i,t}$ appears in Eq. (4.26) below. By the change-of-variables formula [312], the log-likelihood of the observed technology stocks $Y_{i,1:T}$ is

$$\log \mathcal{L}(Y_{i,1:T} \mid \theta_i) = \log p_x(x_{i,1:T} \mid \theta_i) + \log \det \left(\frac{\partial f}{\partial Y_{i,1:T}} \right), \tag{4.25}$$

where p_x is the joint Gaussian density of $x_{i,1:T}$ under the MA(1) noise structure.[2] The Jacobian is taken with respect to f alone, since the deterministic trend $k_i(t - t_{0,i})$ has zero derivative in $Y_{i,t}$. Computing the log-likelihood therefore reduces to applying the GFP

[2]The textbook form of the change-of-variables theorem reads $p_Y(y) = p_X(g^{-1}(y))\,|dg^{-1}/dy|$, with $g : X \mapsto Y$ the level function (here the B-R formula). The formula above is identical: $g^{-1}(Y_{i,t}) = x_{i,t}$, and $|dg^{-1}/dY_{i,t}| = |df/dY_{i,t}|$ since the trend $k_i(t - t_{0,i})$ that separates g^{-1} from f has zero $\partial/\partial Y_{i,t}$.

transform f to the observed data, evaluating the latent Gaussian density at $x_{i,1:T}$, and adding the closed-form Jacobian.

4.7.2.1 The Bertalanffy-Richards curve log-likelihood

By adding $k_i(t-t_{0,i})$ to both sides of the GFP transformation $f(Y_{i,t}; L_i, \beta)$ from Eq. (4.16), we obtain a normal distribution

$$x_{i,t} = \frac{1}{\beta}\log\left(\frac{L_i^\beta}{Y_{i,t}^\beta} - 1\right) + k_i(t - t_{0,i}) = \sigma_i\left(\epsilon_{i,t} + \rho\epsilon_{i,t-1}\right) \tag{4.26}$$

where $\epsilon_{i,t} \sim \mathcal{N}(0,1)$ for all technologies i. Eq. (4.26) gives the closed form of the centred latent defined above. The (t,t') element of the Jacobian of f with respect to $Y_{i,1:T}$ is

$$\left(\frac{\partial f(Y_{i,1:T}; L_i, \beta)}{\partial Y_{i,1:T}}\right)_{t,t'} = \frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t'}} = \delta_{t,t'}\frac{-L_i^\beta}{Y_{i,t}(L_i^\beta - Y_{i,t}^\beta)} \tag{4.27}$$

where $\delta_{t,t'}$ is the Kronecker delta ($\delta_{t,t'} = 1$ if $t = t'$, 0 otherwise). Since the Jacobian is diagonal, its determinant equals the product of diagonal elements. The log-determinant is

$$\log\det\left(\frac{\partial f(Y_{i,1:T}; L_i, \beta)}{\partial Y_{i,1:T}}\right) = \sum_{t=1}^{T}\left[\beta\log L_i - \log\left(Y_{i,t}(L_i^\beta - Y_{i,t}^\beta)\right)\right]$$

which is independent of the noise model.

i.i.d. noise model First, consider the case where $\rho = 0$, so all noise is i.i.d. Eq. (4.16) reduces to

$$x_{i,t} = \sigma_i\epsilon_{i,t},$$

which implies

$$x_{i,t} \sim \mathcal{N}(0, \sigma_i^2).$$

The conditional probability of $Y_{i,t}$ given parameters θ_i is normal, yielding the log-likelihood function

$$\log p(Y_{i,t} \mid \theta_i) = \log \left(\mathcal{N}\left(x_{i,t}; 0, \sigma_i^2\right) \mathbb{I}[Y_{i,t} < L_i] \right) + \log \det \left(\frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t}} \right), \tag{4.28}$$

where $\mathbb{I}[Y_{i,t} < L_i]$ is the indicator function ensuring $Y_{i,t} < L_i$, which is required to ensure Eq. 4.26 remains well-defined. We next derive the log-likelihood for $\rho \neq 0$.

MA(1) noise model When $\rho \neq 0$, we calculate the joint distribution for the full sequence $x_{i,1}, \ldots, x_{i,T}$. The covariance matrix for an MA(1) process is

$$\mathrm{Cov}(x_{i,t}, x_{i,t-j}) = \begin{cases} (1+\rho^2)\sigma_i^2 & |j| = 0 \\ \rho\sigma_i^2 & |j| = 1 \\ 0 & |j| > 1 \end{cases} \tag{4.29}$$

The joint probability density function (PDF) is

$$p(x_{i,1}, \ldots, x_{i,T}) = \frac{1}{\sqrt{|V_i|}(2\pi)^{T/2}} \exp\left(-\frac{1}{2} X_i^\top V_i^{-1} X_i \right) \tag{4.30}$$

where

$$X_i = \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,T} \end{bmatrix}$$

is a vector of length T, and

$$V_i = \sigma_i^2 \begin{bmatrix} 1+\rho^2 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1+\rho^2 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1+\rho^2 & \rho & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho & 1+\rho^2 & \rho & 0 \\ 0 & \cdots & 0 & \rho & 1+\rho^2 & \rho \\ 0 & \cdots & 0 & 0 & \rho & 1+\rho^2 \end{bmatrix} \tag{4.31}$$

is a symmetric $T \times T$ covariance matrix. The first diagonal entry $(1+\rho^2)\sigma_i^2$ reflects the stationary MA(1) assumption, where $\epsilon_{i,0} \sim \mathcal{N}(0,1)$ contributes to $x_{i,1}$ even though it occurs before the observed data. For technology diffusion, we assume that $0 < \rho < 1$, indicating positive correlation between noise in consecutive periods. Since $\rho > 0$, V_i is positive definite and therefore invertible. The log-likelihood for the MA(1) model is

$$\log p(Y_{i,1:T} \mid \theta_i) = \log p(x_{i,1}, \ldots, x_{i,T}) + \sum_{t=1}^{T} \log \det \left(\frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t}} \right), \tag{4.32}$$

where the joint density $p(x_{i,1}, \ldots, x_{i,T})$ is given by Eq. (4.30) with covariance matrix V_i, and the Jacobian term is identical to the i.i.d. case.

Posterior distribution under an MA(1) process In an MA(1) model, current noise depends on the previous noise term. Specifically, the noise at time t is $\sigma_i \epsilon_{i,t} + \rho \sigma_i \epsilon_{i,t-1}$. Thus, the first forecast value $Y_{i,T+1}$ depends on both the future unobserved noise draw $\epsilon_{i,T+1}$ and the final in-sample noise term $\epsilon_{i,T}$. $\epsilon_{i,T}$ can be inferred from the data. To best initialize the forecast, we can estimate $\epsilon_{i,T}$ along with the parameters θ_i.

To do this, we define a posterior distribution that explicitly includes $\epsilon_{i,T}$:

$$\pi(\theta_i, \epsilon_{i,T} \mid Y_{i,1:T}) = \frac{p(Y_{i,1:T} \mid \theta_i, \epsilon_{i,T}) p(\epsilon_{i,T}) \pi(\theta_i)}{p(Y_{i,1:T})} \propto p(Y_{i,1:T} \mid \theta_i, \epsilon_{i,T}) p(\epsilon_{i,T}) \pi(\theta_i)$$

where $p(Y_{i,1:T} \mid \theta_i, \epsilon_{i,T})$ is the likelihood under the DGP, $\pi(\theta_i)$ is the parameter prior, and $p(\epsilon_{i,T}) = \mathcal{N}(0,1)$ matches the standard-normal distribution of $\epsilon_{i,T}$ under Eq. 4.17.

For all $s \in \{1, \ldots, T-1\}$,

$$x_{i,s} = \sigma_i \epsilon_{i,s} + \rho \sigma_i \epsilon_{i,s-1}$$

as before. For $s = T$,

$$x_{i,T} - \sigma_i \epsilon_{i,T} = \rho \sigma_i \epsilon_{i,T-1}.$$

For the time series $Y_{i,1:T}$, we can perform Bayesian inference on the transformed data $x_{i,1:T}$ to estimate parameters θ_i and $\epsilon_{i,T}$, where

$$\tilde{X}_i := \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,T-1} \\ x_{i,T} - \sigma_i \epsilon_{i,T} \end{bmatrix} = \sigma_i \begin{bmatrix} \rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & \rho & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & \rho & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \epsilon_{i,0} \\ \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,T-2} \\ \epsilon_{i,T-1} \end{bmatrix}.$$

The innovation vector includes $\epsilon_{i,0}$ (the implicit pre-sample innovation from the stationary MA(1) assumption), making the matrix $T \times T$ and the innovation vector of length T. We adjust the covariance matrix from Eq. 4.29 to account for estimating $\epsilon_{i,T}$. The modified

covariance matrix is

$$V_i^* = \sigma_i^2 \begin{bmatrix} 1+\rho^2 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1+\rho^2 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1+\rho^2 & \rho & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho & 1+\rho^2 & \rho & 0 \\ 0 & \cdots & 0 & \rho & 1+\rho^2 & \rho \\ 0 & \cdots & 0 & 0 & \rho & \rho^2 \end{bmatrix} \tag{4.33}$$

where the last diagonal element is now $\rho^2\sigma_i^2$ instead of $(1+\rho^2)\sigma_i^2$.

Using the modified covariance matrix V_i^* and vector $\tilde{X}_i$, we could update the joint PDF $p(x_{i,1}, \ldots, x_{i,T})$ from Eq. (4.30) and generate the posterior distribution through Metropolis–Hastings sampling.

Setting $\epsilon_{i,T} = 0$ in practice. We set $\epsilon_{i,T} = 0$ universally. Since $\mathbb{E}[\epsilon_{i,T}] = 0$ by construction, this simplification reduces model complexity without substantially affecting forecast quality (full justification in Section 4.13). The standard covariance matrix V_i from Eq. 4.29 then applies (rather than V_i^*), and the posterior reduces to $\pi(\theta_i \mid Y_{i,1:T})$.

4.7.3 Distributional forecasts

We generate distributional forecasts by sampling from the posterior $\pi(\theta_i \mid Y_{i,1:T})$. We draw N_{samples} samples (after thinning), and for each sample $m \in \{1, \ldots, N_{\text{samples}}\}$ use $\theta_{m,i}$ to forecast an adoption path $Y_{m,i,t}$ that the technology could follow.

We forecast for $t \in \{T+1, \ldots, T+\tau\}$, where τ is the forecast horizon. Since $\epsilon_{i,T} = 0$, the first forecast step at $T+1$ has noise term $\sigma_{m,i}\epsilon_{i,T+1}$ (as $\rho\epsilon_{i,T} = 0$). For a τ-step forecast

over $[T+1, T+\tau]$, the MA(1) noise vector is

$$\boldsymbol{\epsilon}_{T+1:T+\tau,m,i} = \sigma_{m,i} \begin{bmatrix} \epsilon_{i,T+1} + \rho\epsilon_{i,T} \\ \epsilon_{i,T+2} + \rho\epsilon_{i,T+1} \\ \vdots \\ \epsilon_{i,T+\tau} + \rho\epsilon_{i,T+\tau-1} \end{bmatrix} \tag{4.34}$$

with $\epsilon_{i,t} \sim \mathcal{N}(0,1)$ i.i.d. for $t \in \{T+1, \ldots, T+\tau\}$. The first component $\epsilon_{i,T+1} + \rho\epsilon_{i,T}$ shows why $\epsilon_{i,T}$ matters: it propagates into the one-step-ahead forecast through the MA(1) structure.

4.7.4 Implementation: forecasting procedure for a time-series

Algorithm 2 outlines the full procedure for a single technology time-series: data normalisation, Bayesian inference, and forecast generation. Key steps are detailed below.

Data normalization. To assess many technologies simultaneously and fairly, we normalize data to be able to set priors across technologies at each diffusion level. k_i and σ_i are approximately scale invariant across diffusion levels and technologies, so we do not require normalization. However, L_i and $t_{0,i}$ can vary significantly in scale across technologies, making it challenging to define a single prior for all technologies. For example, L_i can range from 10^2 (e.g., fraction of population adopting a behavior) to 10^8 (e.g., number of mobile subscriptions). Similarly, $t_{0,i}$ can vary from being in the 1100s (e.g., adoption of *monasteries*) to the 2000s (e.g., mobile phones). We perform two normalization procedures that help inform prior distributions. First, we normalize adoption data by the maximum observed value,

$$\tilde{Y}_i(t) = \frac{Y_{i,t}}{Y_{i,\max}},$$

where $Y_{i,\max} = \max_t(Y_{i,t})$. Hence, at any diffusion level, $\max_t \tilde{Y}_i(t) = 1$. The value of $\tilde{L}_i \equiv L_i / Y_{i,\max}$ is constant across technologies for a given diffusion level. For example, at 10% diffusion, $\tilde{L}_i \approx 10$ and at 50% diffusion, $\tilde{L}_i \approx 2$.

Second, we shift the time axis so that the first observation occurs at $\tilde{t} = 0$,

$$\tilde{t} = t - t_1.$$

This makes the time of the first observation consistent across technologies, reducing cross-technological variation in the $t_{0,i}$ prior. Some variation in $t_{0,i}$ remains because the inflection point can occur at different times relative to the first observation.

Defining prior distributions. We set weakly informative priors for the parameters $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$ of each technology. Each parameter has a three-parameter shifted Gamma prior. We write $X \sim \Gamma(\alpha, \ell, s)$ to mean $X = \ell + Z$ with $Z \sim \Gamma(\alpha, s)$ [313]. Here ℓ is the lower bound of the support, s is the scale, and α is the shape. The mode is $\ell + (\alpha - 1)\, s$ for $\alpha \geq 1$. When $\alpha = 1$ the prior reduces to a shifted exponential with origin at ℓ. For numerical stability we sample the logarithm of every parameter except L_i.

We apply the same prior to every technology and at every forecast origin. We first normalise each series so the last observed value $y_{T_{\max}} = 1$, where $T_{\max}$ is the time index of the last observation. The asymptote $\tilde{L}_i$ is then a multiple of current adoption. The prior is

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2,\, 1,\, 35), \\ t_{0,i} \sim \Gamma\big(3,\, \max(5,\, T_{\max} - 3/\hat{k}_{\mathrm{ssr},i}),\, 30\big), \\ k_i \sim \Gamma(1,\, 0,\, 1), \\ \sigma_i \sim \Gamma(1,\, 0,\, 1), \end{cases} \tag{4.35}$$

where $\hat{k}_{\mathrm{ssr},i}$ is a preliminary growth-rate estimate defined below. We call this the *universal*

prior, a one-size-fits-all rule, not a prior hand-tuned per technology or diffusion level.

The lower bound $\ell = 1$ on $\tilde{L}_i$ enforces $\tilde{L}_i > 1$. This encodes that the technology has not yet saturated, so its asymptote lies above current adoption.

We anchor the $t_{0,i}$ location to each series through the growth-rate estimate $\hat{k}_{\text{ssr},i}$. We obtain it from a least-squares fit of the GFP transform (Eq. 4.16) profiled over L_i. If this fit fails, we fall back to $\hat{k}_{\text{ssr}} = 0.15$. Anchoring the location at $T_{\max}$ alone would assume the inflection has not yet occurred. The offset $-3/\hat{k}_{\text{ssr},i}$ relaxes this. A slower growth estimate gives a larger offset, so the prior then allows an earlier inflection. The fallback value of 0.15 is slow, giving an offset of about 20 time steps. The lower bound then does not rule out the true inflection while the technology is below about 83% diffusion (for $\beta = 2/3$). Beyond that level the post-inflection stage of the S-curve is clearly visible, so the prior can be adjusted by hand. Section 4.16 gives the derivation.

The k_i and σ_i priors are unit exponentials. We use this same prior for the backtests of main paper Figure 3 and the solar and wind forecasts of main paper Figure 4.

We keep the prior simple and cross-technology, with no hand-tuning. This shows that a single automatic rule already gives well-calibrated forecasts across the 120-technology set. Section 4.16 gives the full design rationale and sensitivity analyses. It also backtests the prior on a larger 154-technology dataset and shows how more informative technology- or diffusion-stage-specific priors improve forecasts. Section 4.10 discusses this further.

Prior configuration files. We implement the prior distributions of Eq. (4.35) in `mcmc_setup.py` for the standard model. For forecasting market shares (the fraction of a market rather than absolute adoption levels), we suggest adjusting the L_i prior to reflect that shares are bounded in $[0, 1]$, as implemented in `mcmc_setup_shares.py`. For the asymptotic growth model of Section 4.17, which accounts for market growth via an exogenous exponential parameter, `mcmc_setup_ass_1.py` adjusts the L_i and $t_{0,i}$ location

parameters.

Metropolis–Hastings sampling. We perform Metropolis–Hastings MCMC sampling to generate samples from the posterior distribution [166, 314]. Algorithm 2 describes the full forecasting procedure. We base the numbers described in the algorithm on our implementation for the results shown in Figure 3 of the main text, but these can be adjusted as needed.

Algorithm 2: Distributional forecasting procedure for technology diffusion

Input: Technology time series $\{Y_{i,t}\}$ for technology i, forecast horizon τ

Output: Distributional forecast for $t \in \{T+1, \ldots, T+\tau\}$

Step 1: Load and normalize data:

Load observed time series: $\{t_1, Y_{i,1}\}, \ldots, \{t_T, Y_{i,T}\}$

Calculate normalization constant: $Y_{i,\max} = \max(Y_{i,1}, \ldots, Y_{i,T})$

Normalize data: $\tilde{Y}_{i,t} = Y_{i,t}/Y_{i,\max}$ for all $t \in \{1, \ldots, T\}$

Shift time: $\tilde{t} = t - t_1$ so first observation occurs at $\tilde{t} = 0$

Step 2: Bayesian inference:

Define prior distributions $\pi(\theta_i)$ for $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$

Optional: Set universal parameters: $\beta = 2/3$, $\epsilon_{i,T} = 0$, and $\rho = 0.8$

Calculate log-likelihood $\log p(\tilde{Y}_{i,1:T} \mid \theta_i)$ using Eq. (4.32)

Perform Metropolis–Hastings MCMC sampling:

Generate $N_{\text{MCMC}} = 1{,}000{,}000$ samples from posterior

Thin by a factor of 100 to adjust for autocorrelation between MCMC samples

Retain $N_{\text{samples}} = 10{,}000$ samples

Step 3: Generate forecast trajectories:

foreach *posterior sample* $m = 1$ *to* $N_{samples} = 10{,}000$ **do**

Extract parameters: $\theta_{m,i} = \{L_{m,i}, k_{m,i}, t_{0,m,i}, \sigma_{m,i}\}$

foreach *forecast time* $t \in \{T+1, \ldots, T+\tau\}$ **do**

Draw $\epsilon_{i,t} \sim \mathcal{N}(0,1)$

B-R forecast:

$\tilde{Y}_{m,i,t} = L_{m,i}\big[1 + \exp\big(-k_{m,i}\beta(t - t_{0,m,i}) + \beta\sigma_{m,i}(\epsilon_{i,t} + \rho\,\epsilon_{i,t-1})\big)\big]^{-1/\beta}$

Update lag: $\epsilon_{i,t-1} \leftarrow \epsilon_{i,t}$

end

Rescale: $Y_{m,i,t} = \tilde{Y}_{m,i,t} \cdot Y_{i,\max}$ for $t \in \{T+1, \ldots, T+\tau\}$

end

Step 4: Construct distributional forecast:

foreach *forecast time* t **do**

The distribution of $Y_{m,i,t}$ across $m = 1, \ldots, N_{\text{samples}}$ samples represents the forecast distribution at time t

end

Step 5 (Optional): Apply shifted correction:

foreach *posterior sample* m **do**

Generate model value at last training time using $\theta_{m,i}$: $\hat{Y}_{m,i,T}$

Compute GFP shift between observed and model values:

$\delta_{m,i} = \frac{1}{\beta}\log\left(\frac{L_{m,i}^{\beta}}{Y_{i,T}^{\beta}} - 1\right) - \frac{1}{\beta}\log\left(\frac{L_{m,i}^{\beta}}{\hat{Y}_{m,i,T}^{\beta}} - 1\right)$

Apply shift to all forecast times in GFP space and invert to get $Y_{m,i}^{\text{shifted}}(t)$

end

Distributional forecasts for the B-R curve. The τ-step forecast under posterior sample m is the vector

$$\boldsymbol{Y}_{m,i}(\tau) = \frac{L_{m,i}}{\left(1 + \exp\left(-k_{m,i}\beta(\boldsymbol{\tau} - t_{0,m,i}) + \beta\boldsymbol{\epsilon}_{T+1:T+\tau,m,i}\right)\right)^{1/\beta}}, \qquad \boldsymbol{\tau} = \begin{bmatrix} T+1 \\ \vdots \\ T+\tau \end{bmatrix},$$

with $\boldsymbol{\epsilon}_{T+1:T+\tau,m,i}$ the forecast-period noise vector from Eq. 4.34. The component-wise forecasts $\{Y_{m,i,t}\}_{m=1}^{N_{\text{samples}}}$ at each $t \in \{T+1, \ldots, T+\tau\}$ form the predictive distribution at time t.

Shifted and nonshifted forecasts. We consider two forecast variants. The *nonshifted* forecast uses the trajectory $\boldsymbol{Y}_{m,i}(\tau)$ directly, as described above. The *shifted* variant applies a post-hoc correction in GFP space. For each posterior sample m, let $\hat{Y}_{m,i,T}$ denote the model's fitted value at the last training time T. We compute the GFP transform (Eq. 4.16) of both the observed data point and the model value,

$$\delta_{m,i} = \underbrace{\frac{1}{\beta}\log\left(\frac{L_{m,i}^{\beta}}{Y_{i,T}^{\beta}} - 1\right)}_{\text{observed GFP at } T} - \underbrace{\frac{1}{\beta}\log\left(\frac{L_{m,i}^{\beta}}{\hat{Y}_{m,i,T}^{\beta}} - 1\right)}_{\text{model GFP at } T},$$

and add $\delta_{m,i}$ to every forecast-period GFP value before inverting back to adoption levels. Then the model fitted value at the last training observation is equal to the true value (i.e., the residual is zero by construction). The shifted variant can reduce point-forecast error and improve calibration, with the consequence of slightly overdispersed forecast distributions at short forecast horizons (See Section 4.16.)

4.7.5 Adaptive widening of the predictive distribution

This subsection develops the adaptive widening shown in the before/after coverage panels of main paper Figure 3(e,f). Bayesian prediction intervals are often miscalibrated when data are limited or the model is misspecified. A nominal $c\%$ interval then need not contain the outcome $c\%$ of the time, and post-hoc recalibration of the predictive distribution is a standard remedy [159, 315]. Our forecasts show this pattern. On the 120-technology backtesting set the shifted forecast distribution is mildly under-dispersed (Section 4.16). The empirical probability integral transform (PIT) CDF sits below the diagonal at high percentiles, so the prediction intervals are too narrow. Part of this overconfidence is attributable to the noise misspecification discussed in Section 4.9.

We correct it post hoc, without re-specifying the noise model. We apply a median-preserving transform to the per-technology posterior predictive samples, so every point forecast is unchanged. We deliberately use a single-parameter form rather than a flexible nonparametric recalibration. This limits the risk of overfitting the backtesting set. The transform is asymmetric and stretches the upper tail more strongly than the lower tail. The exponent is selected per technology from a small lookup table calibrated on the backtesting set. When applied to forecasts, the procedure uses only the training data and the posterior.

Out-of-sample robustness. The recalibration generalises beyond the technologies used to fit it. Leave-one-technology-out cross-validation reproduces the in-sample calibration (Table 4.15). On correctly-specified surrogate data the fitted exponent is close to one, so a well-specified model receives almost no widening (Table 4.16). Section 4.16.4 gives the details.

4.7.5.1 The widening transform

Given posterior predictive samples $\{y_i\}$ with median m, we apply a different exponent above and below the median:

$$y_i^{\text{cal}} = \begin{cases} m\,(y_i/m)^{w} & y_i > m, \\ m\,(y_i/m)^{\sqrt{w}} & y_i \le m, \end{cases} \qquad w \ge 1. \tag{4.36}$$

In log-space, log-deviations from the median are scaled by w in the upper tail and $\sqrt{w}$ in the lower tail. The median is preserved exactly because w does not affect the median forecast. We apply a smaller exponent to the lower tail because forecasts at lower percentiles are better calibrated. The adoption floor is easier to forecast correctly than the maximum saturation level. The empirical justification for the asymmetric form (versus a symmetric variant in which the same exponent is applied above and below the median) is reported in Section 4.16.4.

4.7.5.2 The diffusion-dependent exponent

The required widening depends on the technology's current diffusion stage. We therefore parametrise w as a function of the estimated current diffusion fraction $\hat{d}_i$.

We estimate technology i's current diffusion stage from the existing MCMC samples as

$$\hat{d}_i = \frac{1}{\tilde{L}_i^{\text{med}}}, \qquad \tilde{L}_i^{\text{med}} \equiv \mathbb{P}_{50}\big[\pi(\tilde{L}_i \mid Y_{i,1:T})\big], \tag{4.37}$$

the posterior median of the normalised asymptote. Since the data are normalised so $\max_t Y_{i,t} = 1$, the ratio $1/\tilde{L}_i^{\text{med}}$ is technology i's diffusion fraction at the last training observation. The estimate $\hat{d}_i$ is read off the existing MCMC samples and requires no additional fitting.

We then look up the widening exponent by log-log linear interpolation (linear in $\log \hat{d}$

and $\log w$) between four anchors calibrated by KS minimisation on the 120-technology backtesting set (Section 4.16.4, Table 4.13):

$$\log w^*(\hat{d}) \;=\; \log w_j^* + \frac{\log \hat{d} - \log d_j}{\log d_{j+1} - \log d_j}\left(\log w_{j+1}^* - \log w_j^*\right), \qquad d_j \le \hat{d} \le d_{j+1}, \quad (4.38)$$

with anchors

$$w^*(0.05) = 1.93, \qquad w^*(0.10) = 2.72, \qquad w^*(0.25) = 3.58, \qquad w^*(0.50) = 1.73.$$

Outside the anchor range we clip: $w^*(\hat{d}) = 1.93$ for $\hat{d} \le 0.05$ and $w^*(\hat{d}) = 1.73$ for $\hat{d} \ge 0.50$.

4.7.5.3 Per-technology calibration procedure

For technology i with posterior predictive samples $\{Y_{m,i}(t)\}_{m=1}^{N_{\text{samples}}}$ at any forecast time t:

1. **Estimate the current diffusion level.** Compute $\hat{d}_i = 1/\tilde{L}_i^{\text{med}}$ from Eq. (4.37).

2. **Look up the widening exponent.** Read $w^*(\hat{d}_i)$ from Eq. (4.38), clipped to the anchor range.

3. **Apply the asymmetric transform.** For each forecast time t let $m_{i,t} = \text{median}_m Y_{m,i}(t)$ be the predictive median. Each sample is transformed via Eq. (4.36) with $w = w^*(\hat{d}_i)$:

$$Y_{m,i}^{\text{cal}}(t) = \begin{cases} m_{i,t}\left(Y_{m,i}(t)/m_{i,t}\right)^{w^*(\hat{d}_i)} & Y_{m,i}(t) > m_{i,t}, \\ m_{i,t}\left(Y_{m,i}(t)/m_{i,t}\right)^{\sqrt{w^*(\hat{d}_i)}} & Y_{m,i}(t) \le m_{i,t}. \end{cases}$$

   The median is preserved exactly. The upper tail is stretched in log-space by w^*; the lower tail by $\sqrt{w^*}$.

4.7.5.4 Recommended practice

We recommend applying the adaptive widening transform (Eqs. (4.36)–(4.38)) to forecasts produced with this method. The widened intervals are better calibrated on the 120-technology backtesting set, with reductions at every forecast origin (main paper Figure 3f, Section 4.16.4).

We apply the adaptive widening to the solar PV and wind forecasts reported in the main text (Figure 4(d-f)) and in Section 4.19. We do not apply it in two SI contexts where the goal is to characterise the raw model. That is, in the empirical energy backtest of Section 4.17, which displays baseline behaviour against ground truth, and the un-widened calibration diagnostics of Section 4.16. Every figure states explicitly whether the widening is applied.

4.7.6 Software implementation: how to run the forecasting procedure

We implement the procedure in Python using NumPy [316], SciPy [317], Pandas [318], and statsmodels [319]. Below we list the MCMC hyperparameters and outline the run procedure.

4.7.6.1 MCMC hyperparameters.

We set the following hyperparameters for MCMC sampling:

1. Number of MCMC samples: $N_{\text{MCMC}} = 1{,}000{,}000$
2. Thinning factor: 100 (to reduce autocorrelation)
3. Target acceptance rate: 0.234 (optimal for Metropolis–Hastings sampling)
4. Warm-up \ burn-in period: 100,000 (to allow the scale to converge to the target

acceptance rate)

The values above are the defaults we use, but can be adjusted based on computational resources and convergence diagnostics.

Number of MCMC samples. The number of samples can be increased for more accurate posterior estimates, at the cost of longer computation time.

Number of MCMC chains. We run a single MCMC chain for each technology. Running multiple chains can help assess convergence, but increases computation time.

Target acceptance rate. The optimal acceptance rate of Metropolis–Hastings sampling is 23.4% [168, 169], although this is generally seen as an upper bound, and it may sometimes be lower [320, 321]. In general, an acceptance rate between $10 - 25\%$ is considered acceptable. If the acceptance rate is lower, it means that the proposed distribution is too wide (i.e. there are many samples in low probability regions), and if it is higher, the proposed distribution is too narrow (i.e. the chain is not exploring the parameter space sufficiently).

The thinning factor. The thinning factor can be modified based on the observed autocorrelation in the chains (See Ref. [322] for a discussion on optimal thinning parameter). Thinning is not necessary when conditions are optimal (e.g. sufficiently long sampling run and optimal acceptance rate [323]), but Talts et al. [322] show that generally thinning can significantly improve the posterior predictive performance.

The warm-up \ burn-in period. This period allows adjustment of the scale parameter of the proposal distribution (smaller scale for lower acceptance rates, larger scale for higher acceptance rates) to reach the target acceptance rate.

Sensitivity of hyperparameters. We recommend $\geq$ 1,000,000 samples, a thinning factor of $\geq$ 100, an acceptance rate of 10–25%, and a burn-in of $\geq$ 100,000. Sensitivity tests confirm robustness across reasonable ranges: increasing samples to 5,000,000, raising the thinning factor to 1,000, varying the acceptance rate within 10–25%, and extending burn-in to 500,000 all leave forecast quality essentially unchanged. Burn-in below 100,000 produces poorer mixing and less accurate posteriors; with burn-in above this floor, the thinning factor has only a marginal effect (mainly on the upper tail). Parallel chains (1 vs 4) likewise do not change forecast quality, so we run a single chain per technology.

4.7.6.2 How to run the forecasting procedure.

To run the forecasting procedure, one needs to run two scripts: (1) Either `backtest.py`[3] or `forecast.py`[4] and (2) `plot_forecast.py`. The `backtest.py` or `forecast.py` script performs Bayesian inference to obtain posterior samples, while `plot_forecast.py` generates and plots distributional forecasts using the posterior samples. The steps are as follows.

1. Install the required libraries:

```
pip install numpy scipy pandas statsmodels matplotlib
```

2. Set up the data location pathway for saving and loading files.

3. Load the observed data for the technology of interest and save it in two arrays: one for time points and another for adoption values.

4. Define training/testing split based on the desired diffusion level (e.g., 10%, 25%, etc.) or number of observations.[5]

[3] If backtesting for a specified observed diffusion level

[4] If forecasting only, or backtesting for a specific observed number of data points

[5] To forecast only, set training data to all available data points. That is, define sufficiently large diffusion level or number of observations to include all data points in training.

5. Set up the prior distributions for the model parameters using the appropriate configuration file if forecasting (`mcmc_setup.py` for standard models, `mcmc_setup_shares.py` for share models, or `mcmc_setup_ass.py` for asymptotic growth models), following the structure outlined in Eq. (4.35). For backtesting results presented in the main paper and SI, we always used `mcmc_setup.py` and the universal prior unless otherwise stated.

6. Set specific hyperparameters for MCMC sampling, such as the number of samples, thinning factor and acceptance rate.

7. Run the MCMC sampling to obtain posterior samples (saved in a CSV file).

8. The separate script (`plot_forecast.py`) implements the forecasting procedure:

   (a) Load the posterior samples from the CSV file.

   (b) Set the forecast horizon, τ (number of time steps to forecast).

   (c) Generate forecast trajectories using the posterior samples, as described in Step 3 of Algorithm 2.

   (d) Construct the distributional forecast from the generated trajectories.

   (e) Generate a figure to visualize the forecast trajectories (saved as `forecast_`τ`.png`).

   (f) If backtesting, compare the forecast trajectories to the actual observed data beyond the training period (saved as `backtest_`τ`.png`), where τ defines the forecast horizon.

9. Save the forecast trajectories to a CSV file for further analysis.

4.8 Evaluating distributional forecast performance

We assess whether our forecast distributions are *probabilistically calibrated* [159]. A calibrated forecast neither over- nor under-states uncertainty. We measure calibration with the probability integral transform (PIT) [160, 324], the methodology underlying the coverage panels of main paper Figure 3(d-f).

The PIT. Let $F_{Y_{T+\tau}|Y_{1:T}}$ be the CDF of the predictive posterior $p(Y_{T+\tau} \mid Y_{1:T})$ for a future value $Y_{T+\tau}$ given past data $Y_{1:T}$. The PIT of an observed value $y_{T+\tau}$ is its quantile under this CDF, $F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau})$. If the forecast is calibrated, the PIT is uniform, $F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) \sim \mathcal{U}(0,1)$.

Estimating the PIT. The predictive posterior is rarely available in closed form. Marginalising over the posterior and exchanging the order of integration writes the predictive CDF as the posterior expectation of the per-parameter CDF. We approximate this expectation by Monte Carlo over posterior draws $\theta^{(n)} \sim \pi(\theta \mid Y_{1:T})$:

$$\begin{aligned} F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) &= \int_{-\infty}^{y_{T+\tau}} p(Y_{T+\tau} \mid Y_{1:T})\, dY_{T+\tau} = \int F_{Y_{T+\tau}|\theta}(y_{T+\tau})\, \pi(\theta \mid Y_{1:T})\, d\theta \\ &\approx \frac{1}{N} \sum_{n=1}^{N} F_{Y_{T+\tau}|\theta^{(n)}}(y_{T+\tau}), \end{aligned} \tag{4.39}$$

where $F_{Y_{T+\tau}|\theta^{(n)}}$ is the predictive CDF of $Y_{T+\tau}$ under parameters $\theta^{(n)}$. Equivalently, we take the empirical CDF of the corresponding predictive draws $Y_{T+\tau}^{(n)}$,

$$F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) \;\approx\; \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}\Big(Y_{T+\tau}^{(n)} \le y_{T+\tau}\Big), \tag{4.40}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Both are consistent estimators of the same predictive CDF. We use the empirical-CDF form in the calibration analysis of Section 4.16.

Confidence band. Under perfect calibration the PIT values are $\mathcal{U}(0,1)$ [325], so a Q-Q plot of the empirical PIT distribution against $\mathcal{U}(0,1)$ lies on the 45° line. A convex deviation below the diagonal indicates under-dispersion (overconfidence); a concave deviation indicates over-dispersion (underconfidence). We place a 95% confidence band on the plot under this uniform null. For a sample of n independent PIT values, the k-th order statistic of n i.i.d. $\mathcal{U}(0,1)$ draws follows $U_{(k)} \sim \text{Beta}(k,\, n+1-k)$ [326], and the pointwise 95% band of these Beta quantiles gives the grey envelope in the calibration figures. The binomial counterpart for a PIT histogram is the Clopper–Pearson interval [327], under the same null. When a panel pools several diffusion targets or forecast origins, each technology contributes one correlated PIT value to each, so we size the band with the smallest n entering the panel, the most conservative choice. The empirical Q-Q analysis is in Section 4.16.

Effect of the prior. Calibration depends on how well the prior matches the true parameter distribution. Under a correctly specified model, a prior that is too narrow tends to give overconfident (underdispersed) forecasts, and a prior that is too wide tends to give overdispersed forecasts. The effect grows over the forecast horizon. This is why we keep the priors weakly informative (Section 4.7). On correctly-specified surrogate data forecast with our universal prior, the fitted widening exponent is close to one. This confirms that little recalibration is needed when the model is correct (Section 4.16).

4.9 Empirical surrogate datasets

Purpose. Several analyses compare empirical features to a benchmark generated by the Bertalanffy-Richards model itself. These features include parameter distributions, ensemble growth, residual variance, and forecast errors. We call the benchmark dataset an *empirical surrogate*. The empirical surrogate inherits per-technology parameter estimates from the

data, and gives us a null distribution under the assumed DGP. Unlike a forecast for a real technology, where the true parameters are unknown, the surrogate is generated under the model itself, so the model is correctly specified by construction. We then compare the empirical features against this null.

Generating procedure. We work with N empirical technologies (typically $N = 120$) and fix the technology-invariant parameters $(\beta, \rho, \epsilon_{i,T}) = (2/3,\ 0.80,\ 0)$ from Section 4.13. We estimate $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ in two steps. First, we fit $\hat{L}_i$ by non-linear least squares of Eq. 4.6 to $\{(t, Y_{i,t})\}$. Second, conditional on $\hat{L}_i$, we run OLS of the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ on time t. The OLS slope and intercept give $\hat{k}_i$ and $\hat{t}_{0,i}$, and the residual standard deviation gives $\hat{\sigma}_i$ after the $\sqrt{1+\rho^2}$ correction for MA(1) noise (Section 4.13). We then draw M surrogate datasets, each with N time-series,

$$Y_i^{(s)}(t) \sim \mathbf{P}\Big(Y_{i,t} \,\Big|\, \hat{\theta}_i,\ \beta,\ \rho,\ \epsilon_{i,T}\Big), \qquad s = 1, \ldots, M, \tag{4.41}$$

where $\mathbf{P}$ is the DGP of Eq. 4.17. Some sections test whether a technology-invariant choice is empirically justified, and there we vary that parameter locally. For example, Section 4.13 sweeps β and ρ to assess invariance. Section 4.15 substitutes the Gompertz and logistic DGPs for direct model comparison.

Estimating $\hat{\sigma}_i$ on the 1–95% diffusion range. The shock scale $\hat{\sigma}_i$ sets the surrogate's intrinsic noise floor. It also bounds the forecast accuracy any well-specified model can attain. When we use the surrogate to benchmark forecast performance (Section 4.16), we estimate $\hat{\sigma}_i$ from the GFP-transformed residuals over the central portion of the diffusion range,

$$d_i(t) := Y_{i,t}/\hat{L}_i \in [0.01,\ 0.95], \tag{4.42}$$

Here $d_i(t)$ is the empirical diffusion share, the estimated counterpart of the normalised form $Y_i^{(n)}(t) = Y_{i,t}/L_i$ defined in Section 4.6. Two model-misspecification sources motivate this truncation. *(i) Early diffusion.* The B-R model assumes approximately exponential growth at low d (Section 4.6). Empirically, a minority of technologies grow faster than exponential at $d \lesssim 0.5\%$ and decay to exponential thereafter [91] (Section 4.14). *(ii) Near saturation.* The additive-Gaussian-in-GFP noise specification of Eq. 4.17 forces level-space noise to vanish as $d \to 1$, but empirical series retain noise there. Including either tail forces the least-squares fit to absorb structurally unmodelled variance into a larger $\hat{\sigma}_i$. We set the lower bound just above the Comin threshold so we retain observations at the forecast origins ($d_{\text{train}} \in \{0.05, 0.10, 0.25, 0.50\}$) used in Section 4.16. On the 120-technology dataset this truncation reduces the cross-technology median of $\hat{\sigma}_i$ from ≈ 0.47 on the full series to ≈ 0.28 on the 1–95% range. We also use two alternatives as sensitivity checks in Section 4.16. A tighter window $\hat{\sigma}_i^{5\text{–}95}$ ($d \in [0.05, 0.95]$, cross-technology median ≈ 0.24) drops more low-diffusion data. A wider window $\hat{\sigma}_i^{0\text{–}95}$ ($d \in [0, 0.95]$, cross-technology median ≈ 0.33) keeps all observations below the saturation tail. The wider window tests whether the choice of the lower cut matters, and Section 4.16 confirms it does not change the conclusions.

Analytical expected error (interpretive approximation). The decomposition below is a first-order approximation, not a rigorous bound. We use it to build intuition for the empirical error patterns of Section 4.16. We validate all numerical claims there against exact surrogate-based ALE. A first-order Taylor expansion of the absolute log-error $\ell_{i,t} \equiv \log(\hat{Y}_{i,t}/Y_{i,t})$ around the true parameters $(L_i, k_i, t_{0,i})$ yields

$$\ell_{i,t} \;\approx\; \frac{\delta L_i}{L_i} \;+\; q(d_{i,t})(\delta k_i\, h \;-\; k_i\, \delta t_{0,i} \;+\; \epsilon_{i,t})\,, \qquad q(d) \;\equiv\; 1 - d^{\beta}, \tag{4.43}$$

where $\delta L_i, \delta k_i, \delta t_{0,i}$ are parameter biases and h is forecast horizon. We denote the three first-order contributions as

$$\begin{aligned} L_0 &\equiv |\delta L_i / L_i| && \text{(level)}, \\ U(h) &\equiv q(d)\,|\delta k_i\, h - k_i\, \delta t_{0,i}| && \text{(curvature, grows with horizon)}, \\ S(\sigma) &\equiv q(d)\,\sigma_{\epsilon,i}\sqrt{2/\pi} && \text{(shock)}, \end{aligned}$$

with $\sigma_{\epsilon,i} = \sigma_i\sqrt{1+\rho^2}$ the MA(1)-corrected shock scale: the marginal standard deviation of the GFP-space residual at any single time point. Since the noise satisfies $\eta_{i,t} = \sigma_i(\epsilon_t + \rho\epsilon_{t-1})$ with $\epsilon_t \sim \mathcal{N}(0,1)$ i.i.d. (Section 4.6), $\mathrm{var}(\eta_{i,t}) = \sigma_i^2(1+\rho^2)$ and thus $\mathrm{sd}(\eta_{i,t}) = \sigma_{\epsilon,i}$. Operationally, $\sigma_{\epsilon,i}$ is the typical magnitude by which the observed GFP value deviates from the deterministic B-R trend; the underlying innovation scale is $\sigma_i = \sigma_{\epsilon,i}/\sqrt{1+\rho^2}$. We estimate σ_i from the empirical residual standard deviation after the OLS fit to the GFP-linearised model (see Section 4.13 and the procedure described in this section). Treating the three sources as approximately independent, the expected absolute log error scales as

$$\mathbb{E}[\mathrm{ALE} \mid d] \;\approx\; \sqrt{L_0^2 + U(h)^2 + S(\sigma)^2}. \tag{4.44}$$

Under correct parameters ($L_0 = U = 0$) this reduces to the pure-shock floor

$$\mathbb{E}[\mathrm{ALE} \mid d, \text{correct params}] \;=\; (1 - d^{\beta})\,\sigma_{\epsilon,i}\sqrt{2/\pi}, \tag{4.45}$$

which decreases monotonically from $\sigma_{\epsilon,i}\sqrt{2/\pi}$ at $d \to 0$ to 0 at $d \to 1$.

Worked surrogate example. Using the cross-technology median noise scale on the 1–95% range, $\hat{\sigma}^{1\text{–}95} \approx 0.28$ (median of $\hat{\sigma}_i$ across the 120 technologies), the MA(1) shock standard deviation is $\sigma_\epsilon = \hat{\sigma}^{1\text{–}95}\sqrt{1+\rho^2} \approx 0.28\sqrt{1+0.8^2} \approx 0.36$. Evaluating Eq. (4.45) at the

typical forecast origin $d = 0.05$ with $\beta = 2/3$ gives the irreducible pointwise ALE floor

$$\mathbb{E}[\text{ALE} \mid d{=}0.05, \text{ correct params}] = \underbrace{(1 - 0.05^{2/3})}_{q(d) \approx 0.86} \cdot \underbrace{0.36}_{\sigma_\epsilon} \cdot \underbrace{\sqrt{2/\pi}}_{\approx 0.80} \approx 0.25,$$

where $q(d)$ is the GFP variance-reducing factor of Eq. (4.43) (close to 1 at small d), σ_ϵ is the MA(1) shock standard deviation, and $\sqrt{2/\pi}$ converts the standard deviation of a centered Gaussian to its mean absolute value. The product corresponds to a $\approx 28\%$ relative error in level space ($e^{0.25} - 1$). This is the expected pointwise ALE under perfect parameters and noise scale $\hat{\sigma}^{1\text{–}95}$, and therefore the best accuracy any model can reach at $d = 0.05$ forecast origin. The tighter alternative $\hat{\sigma}^{5\text{–}95} \approx 0.24$ lowers the floor at $d = 0.05$ to ≈ 0.21, and a wider alternative $\hat{\sigma}^{0\text{–}95} \approx 0.33$ increases the floor. We use $\hat{\sigma}^{0\text{–}95}$ as a sensitivity check in Section 4.16.

Shape of the well-specified surrogate ALE. The forecasts in Section 4.16 use a per-sample FP shift that pins each posterior curve to the last training observation. After the shift, all parameter-uncertainty contributions vanish at the training boundary and grow with forecast horizon. The noise floor $S(\sigma)$ decreases with diffusion. The empirical surrogate datasets in Section 4.16 are approximately monotone-increasing in diffusion within the observable range up to 95%. The pronounced hump observed in the empirical pointwise ALE in Section 4.16 (main paper Figure 3a, b) is therefore not a feature of the B-R null. It reflects departures of the empirical data from the B-R DGP.

Caveats. Eqs. 4.43–4.45 are interpretive approximations, not proven bounds. They break down in three ways. *(1)* The decomposition is first-order in $(\delta L_i, \delta k_i, \delta t_{0,i})$ and treats the biases as independent. In practice the GFP-space regression couples them. Overestimating L_i implies underestimating k_i and overestimating $t_{0,i}$ (opposite signs for the symmetric case). Second-order terms become non-negligible for large δL_i. *(2)* The decomposition collapses the MA(1) covariance to its variance contribution $\sigma^2_{\epsilon,i} = \sigma^2_i(1 + \rho^2)$,

dropping the lag-1 covariance $\rho\sigma_i^2$. For averaged residuals over the relevant window lengths, this understates the standard deviation by a factor of ≈ 1.4. *(3)* Eq. 4.45 mixes the L_1 shock scale ($\sqrt{2/\pi}$) with quadratic aggregation. This is convenient for the additive decomposition but is not the exact moment of $|\ell_{i,t}|$. Numerical claims derived from these equations are therefore approximate. The surrogate datasets of Eq. 4.41 retain the full MA(1) structure and parameter coupling, and we report exact ALE on those surrogates in subsequent figures.

Diagnostic-comparison estimate. For diagnostic comparisons that test whether the DGP can replicate an empirical feature of the ensemble itself (Sections 4.13, 4.14, 4.15), we retain the full-series estimate of $\hat{\sigma}_i$. These comparisons are not benchmarks of forecast-error magnitude, so the truncation argument does not apply. Section 4.10 discusses the broader implications of the noise-model misspecification and its link to the post-hoc widening of prediction intervals in Section 4.7.5.

4.10 Limitations

With finite data, prior distributions influence Bayesian posteriors and forecasts [136, 328]. We normalise across technologies so that a single transparent prior applies to all, regardless of the units of L.

We use empirical Bayesian inference [144]: prior hyperparameters are estimated from the same data later used to compute the posterior. This double use can bias results [145, 329], but it may also yield more accurate forecasts and improved robustness on short series [137, 144]. Section 4.16 shows reasonable out-of-sample performance, indicating that the posterior accurately balances prior and data evidence [157].

Our Bayesian inference relies on MCMC methods, which are inherently stochastic [173, 330, 331]. We run a single chain per technology for computational efficiency; in

real-data backtesting we confirmed that running parallel chains does not materially change forecast quality (results not shown). Finite computational resources mean errors remain in posterior samples [331], and finite sampling from the posterior predictive distribution adds variability to the computed prediction intervals [136]. In particular, observations near the boundary of the 95% prediction interval may switch between inside and outside across runs: small MCMC-driven shifts in the posterior predictive move the interval edge across them.

Throughout the text, we avoid over-interpreting results that may depend on a particular seed or prior assumption. For example, forecasting from data covering only the first 1% of diffusion provides so little information that the prior dominates the posterior.

Noise model. The DGP uses a single additive Gaussian shock in GFP space (Eq. 4.17), which is structurally misspecified at the diffusion endpoints (Section 4.9). This contributes to the moderate overconfidence of the B-R prediction intervals reported in Section 4.16 (main paper Figure 3(e,f)), which we correct post hoc via the widening transform of Section 4.7.5. A fully model-based remedy would replace the single shock with a heteroscedastic specification, or augment it with a second source of variance acting in level (or log-level) space, to capture residual variability where the GFP shock vanishes near saturation. Boswijk and Franses [310] propose a heteroscedastic error structure for the Bass diffusion model on essentially these grounds; a similar extension could in principle be applied to the B-R model used here. Such specifications, however, introduce identification challenges and add parameters per technology that are difficult to pin down on the short data sets considered here. We leave this extension to future work.

Universal shape parameter. The B-R curve uses a single technology-invariant exponent $\beta = 2/3$ (Section 4.13). The cross-section supports this value on average (main paper Figure 2d), but true technology-specific exponents would shift the inflection diffusion

and the asymmetry of each fitted curve. Technology-specific estimation of β_i worsened PIT calibration in our backtests (Section 4.16.6), indicating that β_i is weakly identified per series under our priors. A coarser grouping by sector or by diffusion stage is a plausible refinement that we leave to future work.

Saturation prior at low diffusion. The universal $\tilde{L}$ prior is wide and has its mode well above the observed maximum. At low forecast origins the prior dominates the likelihood and the cross-technology posterior median of $\tilde{L}$ can sit above the true normalised asymptote on average (Section 4.16.2.3). This is by design. The universal prior demonstrates that a single prior, combined with technology-specific heuristics for the t_0 location, produces well-calibrated forecasts across forecast origins and horizons (Section 4.16, main paper Figure 3(e,f)). It is a transparent default, not a ceiling. We encourage practitioners to exploit any prior information available for their application. For example, those with diffusion-stage knowledge can replace the universal prior with the tighter, stage-conditional alternative of Section 4.16.7, which lowers MedALE on the calibration set. A complementary direction for future work is to adjust the k prior to compensate for the noise-misspecification bias identified in Section 4.16.2.3.

4.11 Data selection

We compile a technology diffusion dataset comprising time-series spanning the 12^{th} to 21^{st} centuries at global, regional (e.g. EU), and country levels. To provide ground truth, we keep only technologies with data on both early and late stages of diffusion. We perform a six-step cleaning procedure that reduces the dataset from 1362 to 120 time-series: (1) remove death phases (post S-curve asymptote), (2) ensure sufficient sample size, (3) require complete S-curve representation (i.e., enough coverage of early and late diffusion), (4) remove duplicated technologies and non-technology time-series, (5) remove outliers, and (6)

eliminate highly correlated time-series. We discuss limitations at the end of this section.

Defining technology diffusion time-series. We do not impose strict definitions of *technology* or *diffusion*, to keep as many series as possible and avoid bias. Technology diffusion time-series measure the adoption level of a specific technology over time, represented as stocks (cumulative adoption). In some cases, the data is in shares (percentage of all potential adopters); we keep these but note that in- and out-of-sample performance is broadly similar. Some time-series from our initial dataset arguably do not represent technology diffusion (e.g., *GDP*, a measure of *democracy*, *governmental legislation*), and we remove these. Where the case is borderline, however, we tend to keep the series (e.g., *gold reserves*, *monasteries*). Our results are not sensitive to these choices.

4.11.1 Datasets

Our initial dataset combines data from CHAT [299], HATCH [276], World Bank [301], Our World in Data [332], IEA [333], Grubler and Nakićenović [15, 35, 49], IHS [302], RethinkX [303], and Way et al. [29], with monastery records from Janauschek [334]. Grubler and Nakićenović provide European-level, country-level and global-level data; CHAT and HATCH offer country-level and global data; and the remaining datasets provide global-level data.

Aggregating country-level data to global proxy. For technologies in CHAT and HATCH with only country-level data, we construct a global diffusion proxy by aggregation. Let $Y_{i,t}^{(c)}$ denote diffusion of technology i in country c at time t. We compute a global technology stock as

$$Y_{i,t} = \sum_{c} Y_{i,t}^{(c)},$$

treating *NaN* values as zero. This is a simplifying assumption: a missing country entry may reflect absent reporting rather than true zero adoption, so the aggregated level may understate actual global diffusion, particularly in early years when reporting coverage is sparse. After aggregation, we replace remaining zeros with *NaN* to exclude them from model fitting.

Converting flow data to stock data. When the data is already in stocks, we keep it as is. When the data is in flows, we convert it into cumulative stocks. For flow data $\Delta Y_{i,t}$, we compute stocks as

$$Y_{i,t} = \sum_{T=1}^{t} \Delta Y_{i,T},$$

which assumes zero initial stock. If a technology was already present before the start of the sample, this understates the level and can alter the apparent diffusion fraction. The final dataset contains few aggregated flow series, and the cleaning procedure ensures that most series start near a technology's introduction, so the effect is small but cannot be ruled out for all series.

4.11.2 Shares and non-shares

The data comprises both shares and non-shares. For example, the datasets contain data on the total km of railroad tracks (a non-share) as well as the share of the population vaccinated with certain vaccines (shares). Shares are a normalized metric between 0 and 100 (or 0 and 1) accounting for differences in potential adopter base, while non-shares are absolute measures of adoption level. We use both types of data in our analysis, noting that about a third of the data is measured in shares. We do not observe any qualitative differences between the two.

Calculating useful energy shares. For energy, we have experimented with estimating S-curves for both levels and shares; this affects results for specific technologies but does not affect the qualitative overall conclusions. To compute shares, we rely on the concept of *useful energy* to put fossil and electricity-based energy sources on a common footing. We use useful energy data from Way et al. [29], who calculate useful energy from primary energy production, and extend their series (ending in 2020) through 2021–2024 using energy data from Our World in Data [210, 300] and efficiency factors from Way et al. For primary energy consumption $P_{i,t}$ of energy technology i with useful energy factor $f_{i,t}$, useful energy is $U_{i,t} = f_{i,t}P_{i,t}$. Total useful energy consumption is $U_t = \sum_i U_{i,t}$. The share of useful energy from technology i is

$$Y_{i,t} = \frac{U_{i,t}}{U_t}.$$

We use useful energy shares in our dataset for backtesting. We only use useful energy (non-share) data for forecasting future energy technology diffusion levels and when discussing energy technology diffusion levels (Sections 4.17, 4.18, and 4.19). We compare results using useful energy shares vs. levels in Section 4.17.

When the technology measure is a share and is considered a proxy aggregate (i.e., summing across countries; for example *HIB3 vaccine* and *RCV1 vaccine*), we compute an average. There are two such cases in the final cleaned dataset. In these two cases, the data contains two countries. For share data $Y_{i,t}^{(c)} \in [0, 100]$, we proxy the global share as

$$Y_{i,t} = \frac{1}{C}\sum_{c=1}^{C} Y_{i,t}^{(c)},$$

where $C = 2$.

Shares vs. non-shares: statistical comparison. Of the 120 technologies in the cleaned dataset, 42 (35%) are measured as shares and 78 (65%) as absolute levels.

Table 4.1 shows the distribution by sector. Using the same statistical methods applied to the full dataset in Section 4.14, we tested whether shares and non-shares exhibit different diffusion dynamics. Table 4.2 summarizes the results. We found no statistically significant differences for R^2, normalized residuals, or growth rates (k_i) (all p-values > 0.05). The inflection point $t_{0,i}$ (measured relative to first observation) occurs earlier ($p < 0.01$) for shares. This indicates that technologies measured in shares have less data at very low adoption levels. Given these findings, we treat shares and non-shares together in our analysis.

Sector	**Total**	**Shares**	**% Shares**
Industry	30	8	27%
Transport	19	7	37%
Agriculture	19	4	21%
Health	17	6	35%
Energy	15	7	47%
ICT	10	4	40%
Household technologies	4	4	100%
Financial services	2	0	0%
Education	2	2	100%
Military	1	0	0%
Other services	1	0	0%
Total	**120**	**42**	**35%**

Table 4.1: Distribution of shares and non-shares technologies by sector in the cleaned dataset.

Metric	Test	Statistic	p-value
R^2	t-test	-0.47	0.64
	KS-test	0.17	0.39
	Mann-Whitney U	1760	0.50
Normalized residuals	KS-test	0.02	0.62
	Mann-Whitney U	3232347	0.58
Growth rates, k_i	t-test	-0.01	0.99
	KS-test	0.19	0.25
	Mann-Whitney U	1655	0.93
Inflection point $t_{0,i}$ (rel. to first obs.)	t-test	-2.70	0.008
	KS-test	0.35	0.002
	Mann-Whitney U	1111	0.004

Table 4.2: Statistical tests comparing diffusion dynamics between shares and non-shares. The only statistically significant difference is in the inflection point $t_{0,i}$ (measured relative to first observation), where shares have earlier inflection points ($p < 0.01$).

Computing adjusted R^2 for fit quality assessment. We evaluate S-curve fit quality across all 1362 time-series using adjusted R^2 from Bertalanffy-Richards curve fits (Eq. 4.6).[6] For S-curve fit $\hat{y}_{i,t}$ to transformed data $y_{i,t}$ (defined via the GFP transformation in Eq. 4.16), we compute the adjusted R^2 as

$$\bar{R}_i^2 = 1 - \frac{\text{SSE}/(N_i - p)}{\text{TSS}/(N_i - 1)}, \quad \text{SSE} = \sum_t \left(\hat{y}_{i,t} - y_{i,t}\right)^2, \quad \text{TSS} = \sum_t \left(y_{i,t} - \bar{y}_i\right)^2,$$

where $\bar{y}_i = \frac{1}{N_i} \sum_t y_{i,t}$ is the sample mean, N_i is the number of data points, and $p = 2$ is the number of parameters estimated ($\hat{k}_i$ and $\hat{t}_{0,i}$).

Figure 4.6 shows the distribution of the adjusted R^2 for the raw dataset. While the fit explains $> 90\%$ of the variance for many technologies, a significant number of technologies exhibit adjusted $R^2 < 0.5$, potentially due to incomplete S-curves, outliers, or data quality issues.

[6] Running `save_raw_data.py` generates the dataset with 1362 time-series and corresponding R^2 values.

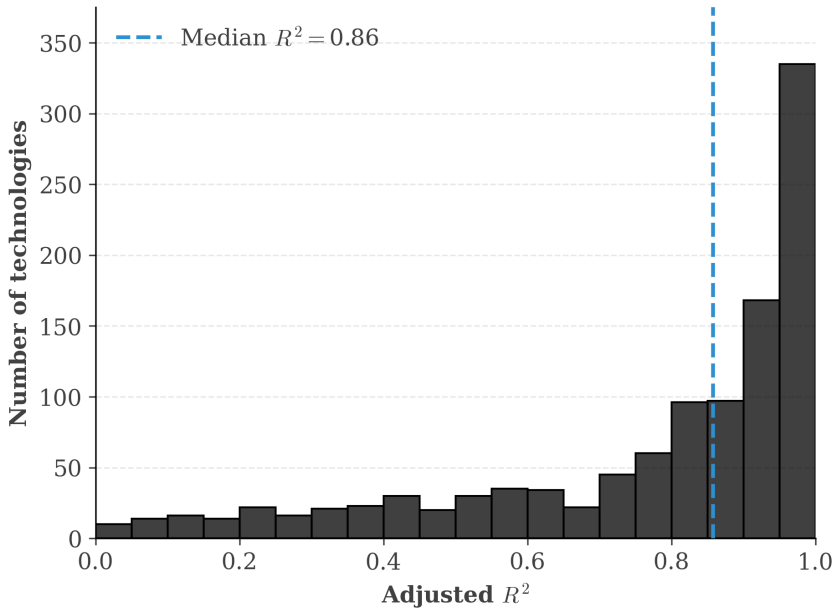


Figure 4.6: Distribution of adjusted R^2 for the fit of the Bertalanffy-Richards curve to all 1362 raw time-series.

4.11.3 Data cleaning

Our cleaning procedure comprises six systematic steps.[7] Table 4.3 summarizes the impact of each cleaning step on the dataset. Earlier steps naturally remove more time-series, as the pool of candidates diminishes progressively. The folder `datasets/` contains the data after each cleaning step.

Step	Description	Removed	Remaining	Data pts.	Unique technologies	% Ret.
–	Raw data	–	1,362	56,014	339	100%
1	Remove death phase	0	1,362	49,605	339	100%
2	Enough data points	404	958	45,665	290	86%
3(a)	Saturation	290	668	31,998	216	64%
3(b)	Early S-curve	211	457	26,125	155	46%
4(a)	Unique technologies	302	155	7,398	155	46%
4(b)	Incorrect saturation	34	121	6,084	121	36%
5	Remove outliers	0	121	6,032	121	36%
6	Correlated time-series	1	120	5,991	120	35%
Total	**Final dataset**	**1,242**	**120**	**5,991**	**120**	**35%**

Table 4.3: Summary of the data cleaning procedure. We identify *unique technologies* by manually grouping time-series based on title and description. Technologies are classified into 11 sectors (Table 4.5) using [335–340] as guidance.

[7] Running `data_cleaning.ipynb` generates the final cleaned dataset with 120 time-series.

4.11.3.1 Computing parameters for cleaning

Estimating asymptote and asymmetry parameters via NLLS. We fit Eq. 4.6 using NLLS regression[8], excluding the first and last data points (this makes our estimates more robust, as we have noticed the estimate of L is very sensitive to these points). This yields $\hat{L}_i$ and $\hat{\beta}_i$. We apply the linearization transformation (Eq. 4.16) with

$$L = \max[\hat{L}_i, 1.2 \max_t(Y_{i,t})], \quad \beta = \min[\max[\hat{\beta}_i, 0.15], 50].$$

The factor 1.2 provides a 20% margin above maximum observed value (values between 1.1 and 1.3 would yield the same technology set). For plotting the distribution of residuals, we use $L = \hat{L}_i$ where $\hat{L}_i$ is the NLLS estimate.

We set $\beta \geq 0.15$ to avoid extreme parameter estimates. The lower bound $\beta = 0.15$ corresponds to inflection at $\approx 1\%$ of asymptote; the upper bound $\beta = 50$ to $\approx 99\%$ diffusion. As discussed in Section 4.14.1, early technology growth can exhibit initial rapid growth slowing to constant exponential behavior after the first 1% of diffusion. Such behavior may lead to estimating very small $\hat{\beta}_i$, due to model mis-specification rather than actual diffusion dynamics. We have performed a sensitivity analysis, and varying the lower bound between 0.1 and 0.2, or setting $\beta = 2/3$ for all technologies yield similar results.

Computing normalized residuals via OLS. We transform data using the GFP transform (Eq. 4.16) and regress the transformed data $f(Y_{i,t}; L_i, \beta_i)$ on time t using OLS to estimate $\hat{k}_i$ and $\hat{t}_{0,i}$. We calculate normalized residuals

$$\tilde{R}_{i,t} = \frac{R_{i,t}}{s_i(R_i)}, \quad \text{where} \quad R_{i,t} = y_{i,t} - \hat{k}_i(t - \hat{t}_{0,i}), \tag{4.46}$$

[8]We use a `piecewise_regression` package that uses bootstrapping, which may lead to varying results upon replication. Under tests the resulting cleaned dataset remains, but number of data points at each step may vary slightly.

and $s_i(R_i) = \sqrt{\frac{1}{N_i-2}\sum_t R_{i,t}^2}$ is the sample standard deviation of residuals $R_{i,t}$ (with a $N_i - 2$ degrees-of-freedom correction, for estimating $\hat{k}_i$ and $\hat{t}_{0,i}$). [9] Algorithm 3 provides the complete procedure.

Algorithm 3: Computing normalized residuals for technology i

Input: Raw time-series data $\{(t, Y_{i,t})\}_{t=1}^{N_i}$ for technology i
Output: Normalized residuals $\tilde{R}_{i,t}$ for $t = 1, \ldots, N_i$
Step 1: Initial parameter estimation via NLLS
Fit Bertalanffy-Richards curve (Eq. 4.6) via NLLS to obtain $\hat{L}_i$ and $\hat{\beta}_i$
Step 2: Adjust parameters for robust transformation
Set $\beta = \min[\max[\hat{\beta}_i, 0.15], 50]$ (avoid extreme asymmetry parameter estimates)
Set $L = \max[\hat{L}_i, 1.2\max_t(Y_{i,t})]$ (ensure asymptote exceeds max observed value)
Step 3: Apply GFP transformation
Compute $y_{i,t} = -\frac{1}{\beta}\log\left[\left(\frac{L}{Y_{i,t}}\right)^{\beta} - 1\right]$ for each $t = 1, \ldots, N_i$
Step 4: Estimate linear model parameters via OLS
Fit $y_{i,t} = k_i(t - t_{0,i}) + \epsilon_{i,t}$ by OLS to obtain $\hat{k}_i$ and $\hat{t}_{0,i}$
Step 5: Calculate raw residuals
Compute $R_{i,t} = y_{i,t} - \hat{k}_i(t - \hat{t}_{0,i})$ for each $t = 1, \ldots, N_i$
Step 6: Compute residual standard deviation
Calculate $s_i(R_i) = \sqrt{\frac{1}{N_i-2}\sum_{t=1}^{N_i} R_{i,t}^2}$
Step 7: Normalize residuals
Compute $\tilde{R}_{i,t} = \frac{R_{i,t}}{s_i(R_i)}$ for each $t = 1, \ldots, N_i$
return $\{\tilde{R}_{i,t}\}_{t=1}^{N_i}$

4.11.3.2 Step 1: Removing death phase data

Some technologies (e.g., *telegrams*, *steamships*) have complete life cycles including phases where the stocks decline. In this paper, we model only the first three phases (up to maturity), omitting *death-phase* data. We identify the death phase by analyzing growth trends in transformed data $y_{i,t}$ (under GFP transformation, Eq. 4.16). To do this, we cannot use $\hat{L}_i$ and $\hat{\beta}_i$ from NLLS fitting directly, as the death phase may distort them. Instead, we use fixed values $L_i = 1.2\max_t(Y_{i,t})$ and $\beta_i = 2/3$. This L_i includes all data points in the transformation; $\beta_i = 2/3$ ensures plausible asymmetry. Other values ($\beta_i \in \{0.5, 1\}$ or $L_i = 1.1\max_t(Y_{i,t})$) yield similar results.

[9] Under the MA(1) noise model (Eq. 4.16), the OLS residuals $R_{i,t}$ are autocorrelated, so the $N_i - 2$ divisor does not yield an exactly unbiased variance estimator. This does not materially affect the cleaning procedure, as s_i is used only for residual normalization, not for model parameter estimation.

For $y_{i,t}$, the start of the death phase represents a shift from positive to negative growth. We detect this via piecewise linear regression with a single breakpoint following Muggeo [341]:

$$y_{i,t} = \begin{cases} k_{i,1}(t - t_{0,i}) + \xi_{i,t}, & t \leq b_i \\ k_{i,1}(b_i - t_{0,i}) + k_{i,2}(t - b_i) + \xi_{i,t}, & t > b_i \end{cases}$$

where $k_{i,1}$ is the slope before breakpoint b_i, $k_{i,2}$ the slope after, and $\xi_{i,t}$ the error term. We estimate b_i and its 95% confidence interval $[b_i^{(0.025)}, b_i^{(0.975)}]$. If $k_{i,1} > 0$ and $k_{i,2} < 0$, we omit data for $t > b_i^{(0.975)}$. If both slopes ≥ 0, the time-series exhibits no death phase. If both < 0, the time-series exhibits only the death phase, and we remove the technology.

Rationale for removing death phase before other steps. We remove the death phase first because S-curve models do not address the death phase: including it introduces strong model misspecification. We start with death phase removal so that all further steps can use statistically estimated values of $\hat{L}_i$ and $\hat{\beta}_i$.

Manual correction of misidentified death phases. Piecewise linear regression may misidentify death phases due to noise or few observations in the saturation phase. We use $b_i^{(0.975)}$ as the cutoff and visually inspect all 306 technologies with identified death phases. We override the automated breakpoint for 85 of them. In 80/85 cases we restore data points that the automated procedure had classified as post-saturation, since they represent noise around the asymptote rather than a sustained decline. This occurs because the noise model underlying the regression (GFP transform) assumes homoscedastic errors, which implies noise diminishes to zero as diffusion approaches the asymptote. In reality, some noise may remain in the saturation phase, leading to misidentification of breakpoints. Including additional data points in the saturation phase helps determine the quality of forecasts in later steps. In 5/85 cases we remove additional data points, as the death phase

starts earlier than initially identified. An example is shown in Figure 4.8g.

Figure 4.10b shows an example of misidentification for *Steamship capacity*, where regression identifies a death phase starting in 1910, but visual inspection indicates 1930 (blue dots are those data points added back into the dataset). Figure 4.11 shows an example for *cane sugar production (Japan)*, where regression identifies a death phase starting in 2006; visual inspection indicates 2003 (brown dots are additional data points removed in this step). We adjust both breakpoints based on these inspections.

4.11.3.3 Step 2: Enforcing minimum data requirement

We require $N_i \geq 10$ data points. We do this because using fewer points yields more unstable parameter estimates, but the results are not sensitive to the threshold, e.g. analysis with $N_i \geq 8$ or $N_i \geq 12$ yields similar results. Of the 1362 time-series, 958 satisfy the $N_i \geq 10$ requirement.

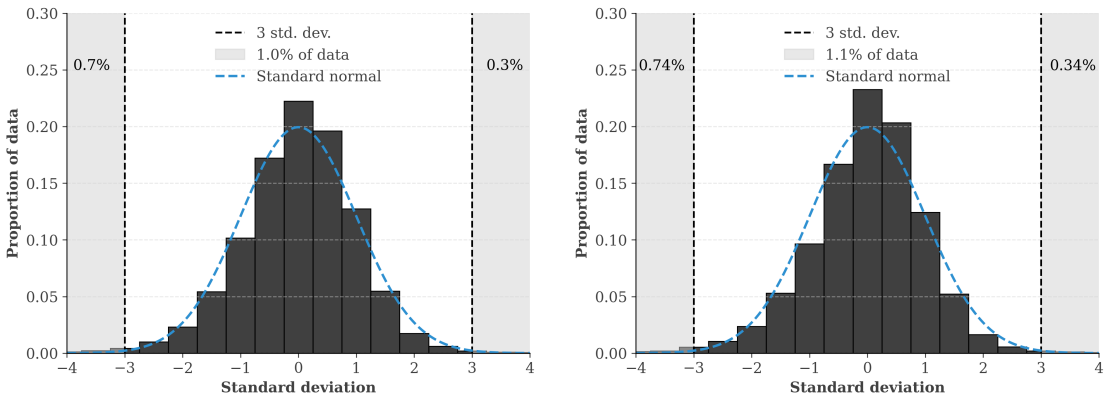
4.11.3.4 Step 3: Requiring complete S-curve coverage

Testing full S-curve dynamics requires both early and late phase data. We require early phases to analyze initial growth patterns (and train the model for backtesting), the full curve to analyze the general shape of adoption, and later phases to analyze saturation dynamics and validate forecasts. Specifically, we require (1) at least 5 data points before 35% diffusion, and (2) diffusion beyond 80% of estimated L.

To find which time series satisfy these requirements, we first estimate the asymptote $\hat{L}_i$ using NLLS. Next, for saturation, we require that $\max_t(Y_{i,t}) > 0.8\hat{L}_i$. This removes 290 time-series. For early diffusion, we keep time series that have at least 5 data points before 35% diffusion, $|\{t : Y_{i,t} < 0.35\hat{L}_i\}| \geq 5$. This removes 211 time-series. Of 958 time-series, 457 satisfy both requirements.[10]

[10]Saturation thresholds of $0.7\hat{L}_i$ or $0.9\hat{L}_i$, and early-phase requirements of $0.3\hat{L}_i - 0.5\hat{L}_i$ yield similar results.

At this point in the data cleaning procedure, all technologies contain both early and late adoption data, no death phase, and at least 10 data points, which should make them suitable for testing hypotheses about S-curves. We observe the distribution of residuals before removing outliers to highlight the effect of outlier removal. Figure 4.7 shows the normalized residuals after cleaning steps 1-3. Table 4.6 summarizes statistics of the residuals. As previously noted, these residuals are computed using the GFP transformation with $L = \hat{L}_i$ and $\beta = \hat{\beta}_i$ from NLLS fitting, unlike the residuals used in the cleaning steps which used adjusted L and β for robustness. The normalized residuals (standard deviations on the



(a) Normalized residuals for 958 time-series with $N_i \geq 10$ (Step 2)

(b) Normalized residuals for 457 time-series representing full S-curve (Step 3)

Figure 4.7: Histograms of normalized residuals under the Bertalanffy-Richards curve fit at different cleaning stages. **(a)** Residuals for 958 time-series with at least 10 data points (Step 2). **(b)** Residuals for 457 time-series that have at least 10 data points and cover a majority of the adoption curve (Step 3). In both cases, the residuals are well-behaved with $\approx 1\%$ of the residuals greater than 3 standard deviations from the mean (slightly fatter tails than a standard normal distribution).

x-axis) behave similarly between steps 2 and 3, with $\approx 1\%$ beyond 3 standard deviations in both cases (compared to 0.3% in a standard normal distribution). In both cases, there are more extreme negative residuals than positive ones. Table 4.6 indicates that the median is positive, but extreme left-tail observations give the distribution negative skewness. This is likely due to early-phase diffusion, where some technologies exhibit rapid initial growth before stabilizing to an exponential regime, a phenomenon that is not fully captured by the B-R curve. We discuss this in detail in Section 4.14.

4.11.3.5 Step 4: Selecting unique technologies and removing poor-quality data

Step 4(a): Applying scoring system to select unique technologies. Some technologies appear multiple times under different names or from different sources (e.g., *mobile phones*, *railroads*, *television*). We manually group technologies into unique technologies based on title and description. For the grouping, see the data dictionary in `data/cleaned/data_dictionary_final.csv` for grouping the remaining 457 time-series into 155 unique technologies. See the notebook `data_cleaning.ipynb` for the code used to generate the data dictionary.

After grouping technologies into unique technologies, we retain one representative time-series per unique technology using a scoring system that prioritizes global coverage, number of data points, early-phase representation, and temporal resolution:

- +1 point if the time-series is at the global level (over country-level data)
- +0.5 point if $N_i \geq 20$, +1 point if $N_i \geq 50$
- +1 point if $\{Y_{i,t} : Y_{i,t} < 0.05\hat{L}_i\} \geq 5$ (sufficient early data to train on)
- +1 if $\Delta_{\max} = 1$ year; -1 if $\Delta_{\max} \in [5, 10]$ years; -2 if $\Delta_{\max} > 10$ years, where $\Delta_{\max} = \max_{j=1,\ldots,N_i-1}(t_{j+1} - t_j)$ for ordered observation times $t_1 < \cdots < t_{N_i}$

When multiple time-series for the same technology have the same highest score, we select the time-series with the longest early data period (for training in the validation exercise) or longest overall time-series (for validation forecasts in backtesting). If still equal, we use adjusted $R^2 \geq 0.9$ as a tiebreaker. The selection procedure described here reduces the dataset from 457 to 155 technologies (302 time-series removed).

Exceptions to scoring system selection. In some cases, we manually select lower-scoring time-series. For *Coal*, *Hydropower*, and *nuclear electricity*, we select Way et al. [29]

data (useful energy shares) over series with higher-scoring primary energy or absolute-TWh data, because the useful-energy-share representation yields a consistent unit across the energy technologies used in the forecasting exercise (Section 4.19). For *Sulphuric acid production*, we select Canadian data because it has longer time-series and more early-phase data over world data.

Step 4(b): Removal of poor-quality time-series. We manually flag 34 of the 155 unique technologies for potential removal through visual inspection, then validate these using quantitative checks (with some overlap). We consider four rationales for removal:

(1) Non-technology time-series. Remove: *GDP (UK).* appears in CHAT/HATCH as potential diffusion indicators but do not represent, in our view, technology adoption. Other non-technology time-series exist, such as *Democracy Index.* These were removed in earlier filtering steps.

(2) Incorrectly specified saturation. Remove time-series where the asymptote $\hat{L}_i$ does not reflect true saturation. All checks below use only data statistics, not model outcomes, and are therefore not circular. The 154-technology robustness check (Section 4.16) confirms that including the flagged technologies does not materially affect calibration or accuracy. (a) The asymptote estimate, $\hat{L}_i$, changes by $> 50\%$ when excluding the last 20% of data (flags 7 technologies); this tests for instability in the NLLS fit that signals the saturation level is not well identified from the data. (b) Average year-over-year log growth does not decrease: $\overline{\Delta \log Y/\Delta t}_{\text{final}} \geq 0.8 \cdot \overline{\Delta \log Y/\Delta t}_{\text{mid}}$, where subscripts denote final 5 data points and middle 40–60% of the time-series respectively. This is distinct from Step 3's requirement that the technology has *reached* 80% diffusion: a technology can satisfy Step 3 yet still show insufficient deceleration in growth rate, which this test detects (flags 7 technologies). (c) Visual inspection: technologies that show growth without any visible curvature toward an asymptote are removed on judgement grounds: *sodium*

battery storage, *high-speed rail*, *satellite launches*, *CCS*, *offshore wind (Germany)*, *solar thermal*, *lithium-ion batteries*, *space launches*. This complements criterion (b) — (b) is statistical, (c) is visual — and together they flag cases where neither the data statistics nor the shape of the curve provide evidence of saturation. (d) NLLS cannot identify $\hat{L}_i$ from data confined to the early-exponential regime ($Y_{i,t} \ll L_i$): the B-R model then reduces to $Y_{i,t} \propto \exp(k_i t)$, in which L_i enters only through curvature that has not yet manifested [31]. This is an identifiability constraint, not a model-outcome judgement. We remove technologies whose data show no curvature toward saturation.

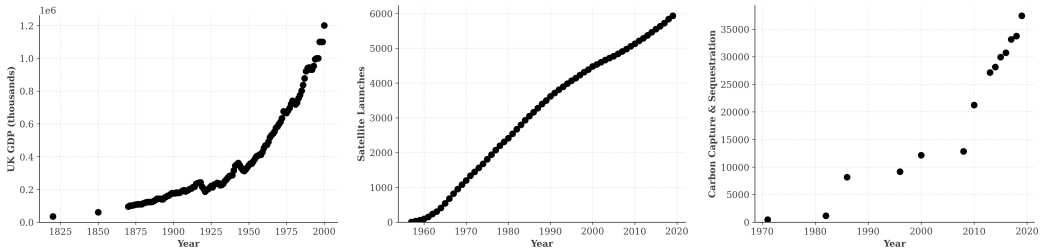
(3) Poor data quality. We remove time-series with erratic behavior inconsistent with smooth diffusion, applying two standard outlier-detection thresholds to year-over-year changes that are independent of model outcomes: (a) $|\Delta Y_{i,t}| > 10\sigma_{\Delta Y}$ where $\sigma_{\Delta Y}$ is the standard deviation of year-over-year changes (flags 10 technologies); (b) coefficient of variation $\mathrm{CV}(\Delta Y_{i,t}/Y_{i,t-1}) > 3$ (flags 14, with overlap). The S-curve model assumes smooth, monotone diffusion; time-series with sudden jumps or persistent fluctuations violate this assumption and would produce unreliable fits regardless of model choice. This step removes, for example, *hospital beds* and *cheques* (sudden jumps) and *iron ore*, *cobalt*, and *nickel* (persistent fluctuations). See Figure 4.8.

Nine technologies pass all automated checks above but are flagged and removed through manual visual inspection: *hernia surgery*, *coronary bypass*, *EFT (UK)*, *sensible heat storage*, plus five others. These are relatively short time series ($N_i < 15$) with insufficient data to distinguish a genuine asymptote from noise, so we opt to remove them. We also remove several battery-, CCS-, and solar-related technologies. We remove these for two reasons. (1) Our application to solar and wind energy forecasting is a key motivation for this study, so we ensure the dataset does not contain technologies in the same energy storage or solar sector, which we reserve for out-of-sample validation (Section 4.19). (2) These technologies have very few data points and no observable curvature toward an asymptote, making the

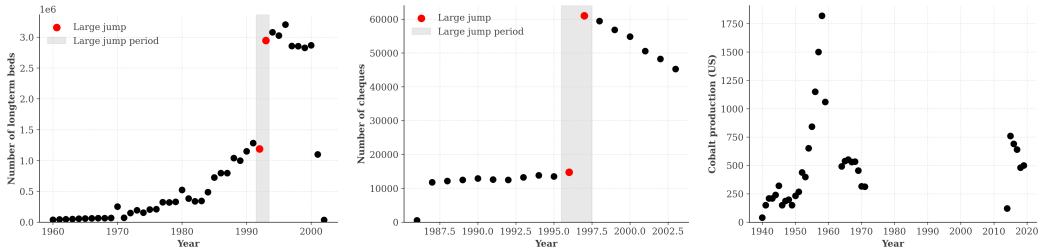
NLLS asymptote estimate unreliable (the same criterion as (d) above). For example, CCS has fewer than 15 global data points with no sign of saturation [342]. The same applies to sodium battery storage [343] and other early-stage storage technologies such as *lithium-ion battery storage*, *latent heat storage*, and *solar thermal energy* [344].

(4) Removal of individual outlier observations within retained technologies. We remove individual data points that reflect events unrelated to diffusion—wartime disruptions, financial crises, or data-entry errors—while retaining the technology itself. Specific cases are documented in Section 4.11.3.6.

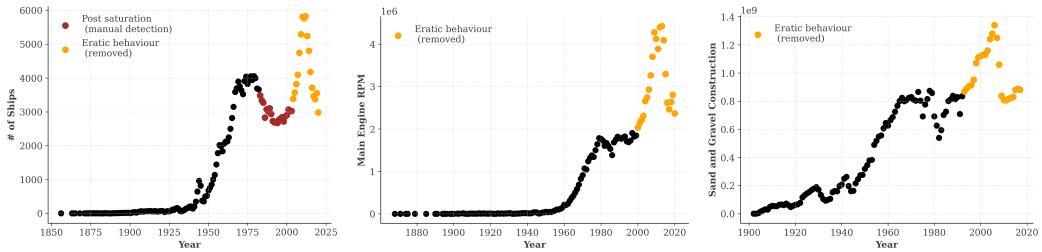
For three technologies (*ships*, *main engine RPM*, *sand/gravel construction*), we retain the technology but remove points that are obvious outliers. Figure 4.8 shows examples of removed time-series (figures 4.8g–4.8i show those where we removed specific data points and used the remaining time-series for analysis). After Step 4(b), 121 technologies remain. See `data/cleaned/step_4b` and `data_cleaning.ipynb` for complete removal details and plots.



(a) Non-technology time-series: UK GDP (USD) **(b)** No evidence of saturation: satellite launches **(c)** No evidence of saturation: CCS



(d) Poor data quality: number of long-term beds in hospitals **(e)** Poor data quality: cheques **(f)** Poor data quality: cobalt production (US)



(g) Poor data quality: Number of ships **(h)** Poor data quality: Main engine RPM **(i)** Poor data quality: Sand and gravel construction

Figure 4.8: Examples of time-series removed during Step 4(b). **(a)** Non-technology. **(b-c)** Growth without clear saturation. **(d-e)** Sudden jumps to saturation. **(f)** Persistent fluctuations. **(g-i)** Erratic fluctuations in final data points (orange: removed in Step 1; black: retained; red: removed manually).

Comparison of dataset properties before and after Step 4(b). Figure 4.9 compares the adjusted R^2 distribution and the normalized residuals before Step 4(b) to the final dataset. The left panel shows that removing the 34 technologies marginally increases the median adjusted R^2 (from 0.93 to 0.94), confirming that most removed series had lower fit quality. The right panel shows boxplots of normalized residuals (Eq. 4.46, under the GFP best fit with $\beta = 2/3$) binned by nondimensional time. The residual distributions are nearly identical between the two datasets, indicating that the removal does not materially alter the overall statistical properties of the data.

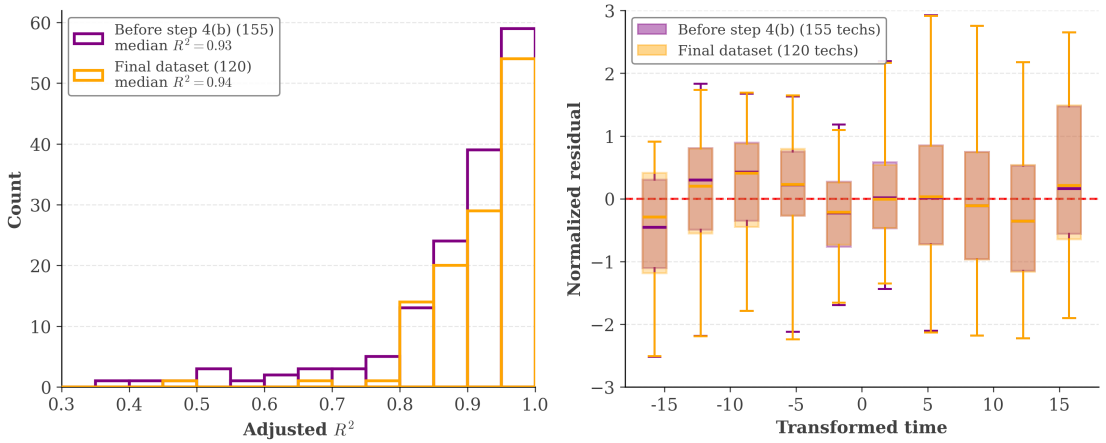


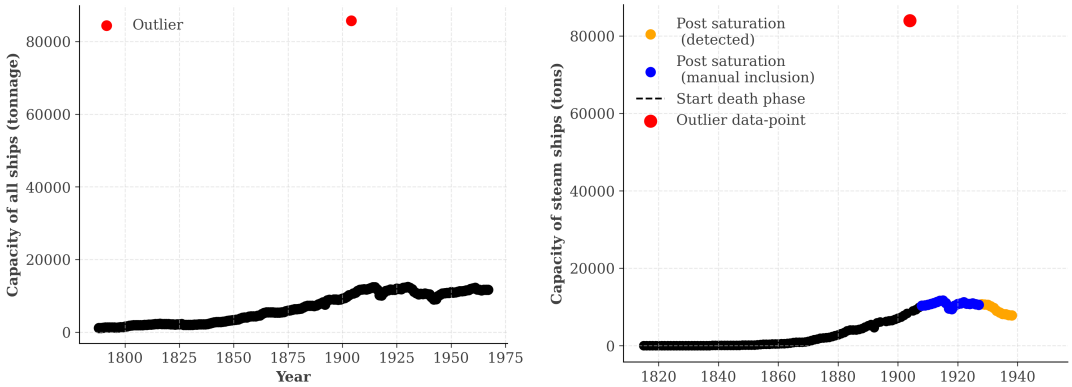
Figure 4.9: Comparison of dataset properties before and after Step 4(b). **Left:** Histogram of adjusted R^2 for the 155 technologies before Step 4(b) (purple) and the final 120 technologies (orange). The median R^2 increases slightly from 0.93 to 0.94. **Right:** Boxplots of normalized residuals (2.5th–97.5th percentile whiskers) binned by nondimensional time (same transformation as for main paper figure 2(b) and (c)). The large overlap across nondimensional time bins indicates that the distribution of residuals is nearly identical before and after Step 4(b), suggesting that the removal of technologies does not materially alter the overall residual distribution.

4.11.3.6 Step 5: Manual identification and removal of outliers

Manual inspection identifies 9 outlier data points for removal: one from *capacity of ships (UK)*, one from *capacity of steamships (UK)*, three from *primary bauxite production (US)*, and four from *gold reserves (South Korea)*. Figure 4.10 shows the removed outliers. The ship capacity outlier in 1910 (Figures 4.10a, 4.10b) is likely a data entry error. Bauxite production anomalies during 1942–1944 (Figure 4.10c) reflect WWII aluminum demand [345] and subsequent peacetime conversion[346]. Gold reserve outliers in South Korea (Figure 4.10d) correspond to financial difficulties [347, 348] preceding the 1997 IMF agreement[349], followed by a 1998 gold-collecting campaign[350]. These outliers reflect extraordinary economic events unrelated to the diffusion process itself; their removal corresponds to criterion (4) of Step 4(b).

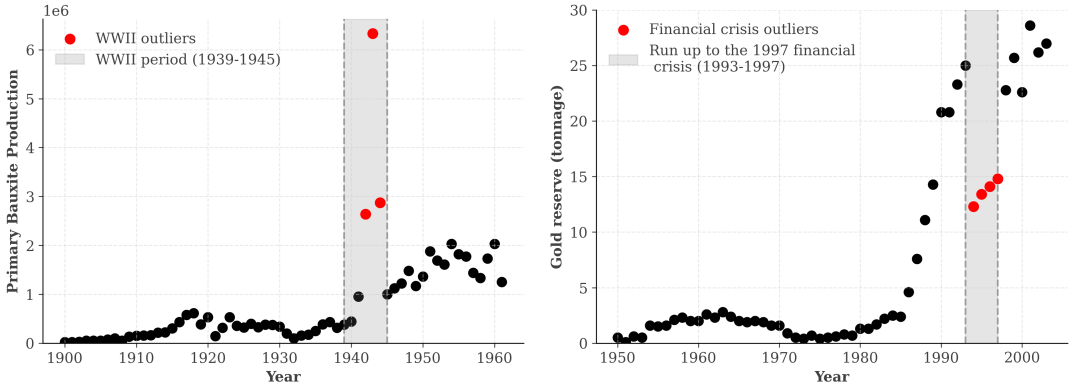
11

[11]The time-series excludes 2011–2013 gold purchases by the South Korean government[351].



(a) Capacity of ships (tons)

(b) Capacity of steamships (tons)



(c) Primary bauxite production

(d) Gold reserves (tons)

Figure 4.10: Diffusion curves for technologies where outliers were removed. **(a)** Capacity of ships (tons) in the UK with one outlier removed in 1910. **(b)** Capacity of steamships (tons) in the UK with outlier removed in 1910. Outlier is likely a data collection error. Death phase detection was too stringent (blue data are those data-points detected as post-saturation that were manually included again in step 1). **(c)** Primary bauxite production in the US with three outliers removed: World War II disruption pattern. **(d)** Gold reserves (tons) in South Korea with four outliers removed, due to 1993-1997 lead up to financial crisis.

Special case: Cane sugar consumption (Japan). For *cane sugar consumption (Japan)*, removal is more extensive. Figure 4.11 shows the time-series with removed data points highlighted. During WWII, cane sugar consumption in Japan dropped sharply due to supply disruptions and rationing [352, 353], falling to a historical low [354] from approximately 15 kg per capita [355]. Post-war recovery increased consumption, reaching pre-war levels by the mid-1950s [354, 356]. We remove all data before 1943 and treat diffusion as starting post-WWII. By the 1990s, Japan experienced long-term decline in sugar consumption [357], indicating death phase onset, reaching a per-capita low in consumption

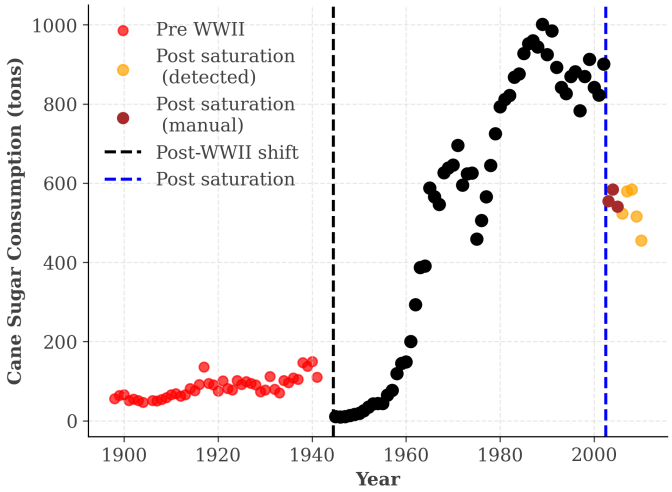


Figure 4.11: Cane sugar consumption (kg per capita) in Japan. Red dots are pre-WWII removed data points. Orange dots are post-2006 removed data points (post-saturation) that were detected in Step 1. Brown dots are additional post-saturation points that were manually removed in Step 1, for death-phase detection checks.

in 2010 [358].[12][13] We analyze data from 1943 to the early 2000s.[14]

[12]Death phase is shown to indicate the full extent of removed data.

[13]Cane sugar consumption stabilized around 17 kg per person in 2021[358] (post-2010 data not included).

[14]Per capita sugar consumption peaked in the early 1970s [358, 359], then declined, aligning with trends in Figure 4.11.

4.11.3.7 Step 6: Removing highly correlated time-series

Some technologies may be highly correlated (e.g., different vaccines in one country). We calculate Pearson correlation coefficients between residuals of all technology pairs (i, j):

$$r_{ij} = \frac{\sum_t (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)}{\sqrt{\sum_t (R_{i,t} - \bar{R}_i)^2 \sum_t (R_{j,t} - \bar{R}_j)^2}},$$

where $\bar{R}_i = \frac{1}{N_i} \sum_t R_{i,t}$ is the mean residual for technology i. We investigate technologies where $r_{ij} > 0.95$. We identify one such pair. We remove 1 time-series: *Polio vaccine (India)*.[15]

The final dataset includes 120 unique technologies; Table 4.4 lists each technology with its region, primary data source, observation count, and year span.

Table 4.4: The 120 cleaned technology time-series. Display names are abbreviated; full identifiers are kept in the data repository. N_i is the number of cleaned observations; the span gives the first and last observed year. Citations are to the database from which each series was drawn; where known, the original primary source is also indicated.

Technology	Region	Source	N_i	Span
Air conditioning (cars)	US	Grübler [15, 35]	30	1954–1984
All Biofuels	Germany	HATCH [299]	30	1992–2021
Ammonia synthesis (gas feedstock)	World	Rouwenhorst et al. [360] / RethinkX [303]	18	1935–2020
Ammonia synthesis (HFO/naphtha)	World	Rouwenhorst et al. [360] / RethinkX [303]	11	1935–1985
Ammonia synthesis (total)	World	Rouwenhorst et al. [360] / RethinkX [303]	22	1915–2020
Aquaculture Production	UK	HATCH [299]	61	1960–2020
ATMs (thousands)	World	Bessen [361] / RethinkX [303]	27	1973–2009
Automobile tire rims (straight side)	US	Klepper [362] / RethinkX [303]	24	1910–1933
Aviation tonne-kilometers	UK	CHAT [299]	47	1945–1991
BCG Vaccine	India	HATCH [299]	41	1981–2021
Beer Production	Japan	HATCH [299]	102	1896–1997
Biogas	Germany	HATCH [299]	22	2000–2021
Bone marrow transplants	World	CHAT [299]	28	1974–2001
Cable TV subscriptions	UK	CHAT [299]	17	1985–2001
Cadmium Refining	US	HATCH [299]	61	1906–1966

(continued on next page)

[15] We retain *DTP3 vaccine (India)* (adjusted $R^2 = 0.01$ higher; other metrics identical).

(continued from previous page)

Technology	Region	Source	N_i	Span
Canals	US	Comin & Hobijn [299]	42	1800–1854
Cane Sugar	Japan	HATCH [299]	58	1945–2002
Capture Fisheries	China	HATCH [299]	58	1960–2017
Caustic Soda	South Korea	HATCH [299]	43	1961–2003
Cement	US	HATCH [299]	120	1900–2019
Cigarette consumption per person	US	US DOT / RethinkX [303]	87	1881–1967
Coal (useful energy)	World	Way et al. [29] / OWID	15	1800–1940
Coal Production	China	HATCH [299]	41	1981–2021
Cobalt Mine Production	World	HATCH [299]	22	1995–2016
Coke (Prussia)	Germany	Landes [363] / RethinkX [303]	26	1825–1890
Color TV broadcasting (countries)	World	Wikipedia / RethinkX [303]	54	1950–2003
Commercial vehicles	US	Comin & Hobijn [299]	99	1904–2002
Containerised ports (cumulative)	World	Rua [364] / RethinkX [303]	28	1956–2003
Copper—Refining	India	HATCH [299]	22	1990–2011
Crop Harvester	South Korea	HATCH [299]	41	1961–2001
Crude oil (useful energy)	World	Way et al. [29] / OWID	24	1870–1978
DDT production	World	Schenker et al. [365] / RethinkX [303]	25	1941–1965
Diesel locomotives in service	US	Comin & Hobijn [299]	44	1925–1970
Dishwashers	US	HATCH [299]	40	1922–2011
Disk Brakes	US	HATCH [299]	19	1966–1984
DTP3 Vaccine	India	HATCH [299]	42	1980–2021
Electric steel production	US	USGS / RethinkX [303]	96	1896–1991
Electricity production	UK	CHAT [299]	98	1896–1993
Ethanol in motor fuel	US	US EIA / RethinkX [303]	39	1981–2019
Ethanol Production	World	HATCH [299]	51	1970–2020
Fax machine shipments (thousands)	World	NRC [366] / RethinkX [303]	15	1977–1991
Fertilizer (total)	World	CHAT [299]	42	1960–2001
Filter cigarettes (				
Flush Toilet	US	HATCH [299]	12	1860–1989
Gas (useful energy)	World	Way et al. [29] / OWID	67	1890–2023
GE corn (HT)	US	USDA / RethinkX [303]	28	1996–2023
Geothermal Energy	Germany	HATCH [299]	18	2004–2021
Gold	South Korea	HATCH [299]	50	1950–2003

(continued on next page)

(continued from previous page)

Technology	Region	Source	N_i	Span
Grain combines (thousands)	US	RethinkX [303]	24	1910–1960
Heart transplants	World	CHAT [299]	23	1978–2000
HIB3 Vaccine	World	HATCH [299]	31	1991–2021
Home Air Conditioning	US	HATCH [299]	55	1957–2011
Home dialysis (kidney)	UK	CHAT [299]	21	1970–1991
Household Internet Access	US	HATCH [299]	24	1993–2016
Hybrid corn (				
Hybrid corn acres	US	USDA / RethinkX [303]	26	1933–1958
Hybrid corn farmers (cumulative)	World	Rogers [40]; Ryan & Gross [367] / RethinkX [303]	15	1927–1941
Hydrochloric Acid	Japan	HATCH [299]	74	1930–2003
Hydropower (useful energy)	World	Way et al. [29] / OWID	67	1890–2023
Internet users	UK	FRED / RethinkX [303]	32	1990–2021
Kidney transplants	World	CHAT [299]	39	1963–2001
Lithotriptor units	World	CHAT [299]	20	1981–2000
Liver transplants	World	CHAT [299]	25	1977–2001
Long-term care beds	World	CHAT [299]	41	1960–2000
Lung transplants	World	CHAT [299]	18	1983–2000
Mammography units	World	CHAT [299]	30	1968–2000
Marine engine RPM	World	Mitchell shipping archive	111	1868–1999
Mechanical cotton harvesting	US	USDA / RethinkX [303]	24	1949–1972
Microcomputers	US	HATCH [299]	33	1984–2016
Milk Production	Japan	HATCH [299]	106	1899–2010
Monasteries (Europe)	Europe	Janauschek [334]	51	1113–1165
Motorship tonnage	World	CHAT [299]	69	1906–1974
MRI units	US	Comin & Hobijn [299]	18	1983–2000
Natural gas consumption	UK	RethinkX [303]	17	1962–1978
Nitric Acid	World	HATCH [299]	75	1929–2003
Nitrogen Fertilizer	India	HATCH [299]	60	1961–2020
Nuclear electricity (useful energy)	World	Way et al. [29] / OWID	38	1965–2002
Nuclear Weapons	World	HATCH [299]	66	1945–2010
Number of ships	World	Mitchell shipping archive	117	1856–1982
O-heath	US	Grübler [15, 35]	84	1874–1957
Oil Production	UK	HATCH [299]	23	1965–1987
Oil Refineries	World	HATCH [299]	68	1940–2007
Oxygen production	US	Grübler [15, 35]	26	1958–1983
Petroleum refining (catalytic cracking)	World	Linstone & Sahal [368] / RethinkX [303]	34	1938–1971
Phone subscriptions	World	World Bank	41	1980–2020
Phosphate Fertilizer	India	HATCH [299]	60	1961–2020

(continued on next page)

(continued from previous page)

Technology	Region	Source	N_i	Span
Pineapple production	Ivory Coast / Costa Rica	FAO / RethinkX [303]	61	1961–2021
Potash Fertilizer	China	HATCH [299]	59	1961–2020
Power steering (cars)	US	Grübler [15, 35]	34	1951–1984
Primary Aluminum Production	US	HATCH [299]	92	1900–1991
Primary Bauxite Production	US	HATCH [299]	59	1900–1961
Primary school enrollment	UK	CHAT [299]	54	1860–1913
Radial tire shipments (				
Radiation therapy equipment	World	CHAT [299]	41	1960–2000
Radio	South Korea	HATCH [299]	52	1948–1999
Railroad Tracks	US	HATCH [299]	123	1830–1952
RCV1 Vaccine	World	HATCH [299]	42	1980–2021
Retail locations accepting card	US	Comin & Hobijn [299]	16	1988–2003
ROTAC Vaccine	World	HATCH [299]	16	2006–2021
Sailing-ship tonnage	World	CHAT [299]	102	1788–1889
Salt production	US	HATCH [299]	118	1900–2017
Sand and Gravel Construction	US	HATCH [299]	91	1902–1992
Secondary school enrollment	UK	CHAT [299]	86	1886–1990
Sewing machines per person	US	Hounshell [369] / RethinkX [303]	43	1853–1895
Sexed dairy semen (				
Ship tonnage (all)	UK	CHAT [299]	178	1788–1967
Steamship tonnage	UK	CHAT [299]	111	1815–1927
Stoves	US	HATCH [299]	108	1900–2011
Sulphuric Acid	Canada	HATCH [299]	84	1919–2002
Synthetic filaments	World	HATCH [299]	81	1910–1990
Telegraphs	World	Grübler [15, 35]	87	1856–1942
Telephones	France	HATCH [299]	97	1889–1985
Television	China	HATCH [299]	41	1959–1999
Tractor adoption	Canada	Olmstead & Hayami / RethinkX [303]	63	1909–1971
Transatlantic passengers by air (share)	World	Gilbert (2008) / RethinkX [303]	27	1948–1974
Travel by car (share of total)	US	RethinkX [303]	66	1895–1960
Unleaded gasoline (share)	US	Kerr & Newell [370] / RethinkX [303]	26	1970–1995
Wet Flue Gas Desulfurization Systems	Germany	HATCH [299]	28	1972–1999
YFV Vaccine	World	HATCH [299]	25	1997–2021
Zinc	Canada	HATCH [299]	93	1911–2003

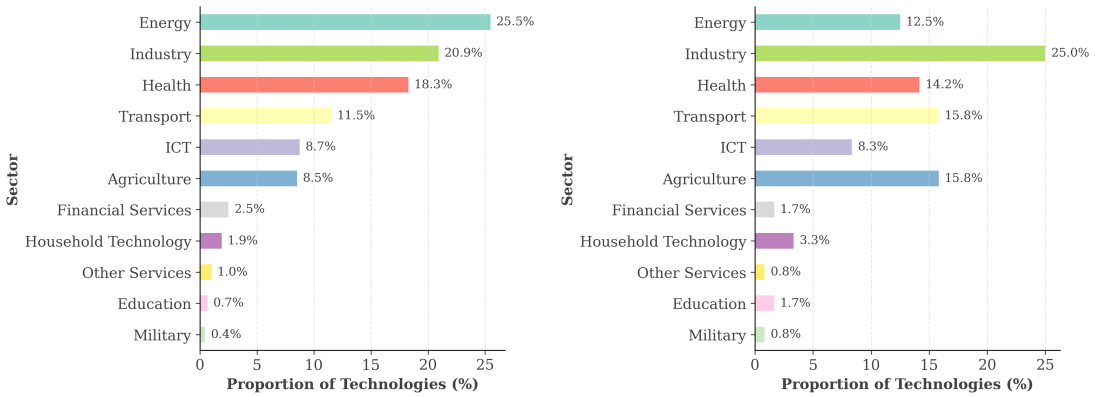
Augmenting the dataset with energy carriers in TWh. For the empirical-regularity diagnostics in Section 4.14 and the forecasting exercises in Sections 4.17–4.19, we additionally retain 11 useful-energy time-series in TWh covering the major fossil and low-carbon energy carriers (coal, oil, gas, nuclear, hydropower, biomass, solar PV, and wind, among others; the full list is in the data repository). These series are not part of the 120-technology set used for the calibration analysis in Section 4.16.

Multi-source construction for solar PV and wind. For solar PV and wind we combine several sources. Useful-energy values (PWh) come from Way et al. [29] and Our World in Data [210, 300]. This useful-energy series provides the historical data plotted in main paper Figure 4a. Installed capacity (TW) comes from IRENA [298], extended further back using Nemet [371] and Farmer and Lafond [30].

4.11.4 Behavior of the clean data

We present sector representation, normalized residuals, and R^2 values of cleaned time-series.

Sector classification in raw and cleaned datasets. Figure 4.12 shows sector representation in cleaned and raw datasets. We perform a manual classification into 11 sectors, based on the titles of the time-series. The sectors can be seen in Figure 4.12 and Table 4.5. The manual classification procedure can be found in the script, `data_cleaning_analysis.ipynb`. For inspiration on the classification, we use Refs. [335] for help with identifying *industry*, *infrastructure*, and *transport* technologies; [336] for *energy*; [337] for *health*; [338] for *agriculture*; [339] for ICT; and [340] for minerals in *industry*. We added *financial services*, *military*, and *other services* for miscellaneous technologies (e.g., monasteries).



(a) Raw dataset **(b)** Cleaned dataset

Figure 4.12: Sector representation in **(a)** raw dataset of 1362 technologies and **(b)** cleaned dataset of 120 technologies. *Industry* and *Health* are reasonably large sectors in both datasets. The cleaned dataset has no uncategorized technologies, while the raw dataset contains some. The differences in proportions are likely due to duplicate technologies being removed, uncategorized time-series (*other services*) in the raw dataset, and some sectors having many technologies that have not reached their saturation phase (*Energy*), while others have more mature technologies that properly saturate (*Transport*, *Agriculture*).

Table 4.5 shows technologies per sector in the final dataset.

Sector	**Number of technologies**	**Examples**
Industry	30	textile machinery, steel, and cement production
Transport	19	automobiles, railways, aircraft, shipping capacity
Agriculture	19	fertilizer, tractors, hybrid corn acreage
Health	17	vaccinations, surgeries, medical devices
Energy	15	gas, coal, oil, electricity production
ICT	10	mobile phones, internet users, television, radio
Household technologies	4	dishwashers, stove, flush toilet
Financial services	2	retail locations accepting card, ATMs
Education	2	primary school enrollment
Military	1	nuclear weapons
Other services	1	monasteries

Table 4.5: Sector representation in the cleaned dataset of 120 technologies.

The raw dataset contains 14 uncategorized technologies, which we place in *other services*, while the cleaned dataset has 1 (*Monasteries - Europe*). The uncategorized technologies in the raw dataset include *GDP*, *public investment*, *private investment*, and *population* for different countries. We removed these in the cleaning process as they do not adequately represent technology adoption.

The cleaned dataset has a more even sector distribution, with a higher proportion of *Industry*, *Transport*, and *Agriculture* technologies, likely because these sectors include more mature technologies. Some sectors, such as *Energy* and *ICT*, have a lower proportion in the cleaned dataset because many technologies have not reached saturation (e.g., solar, wind, CCS). Furthermore, many *ICT* technologies have extensive coverage in the raw dataset due to duplicates (e.g., mobile phones in different countries), which are removed (except for 1) in the cleaned dataset.

Adjusted R^2 and normalized residuals of cleaned dataset. Figure 4.13 shows the adjusted R^2 distribution for 120 technologies[16], the normalized residuals, and a comparison between observed residuals and normal and Student-t (df = 7) distributions[17]. Table 4.6 summarizes key statistics of normalized residuals for various cleaning steps.

Statistic	Step 2	Step 3	Cleaned Dataset	Std. Normal
Mean	0	0	0	0
Median	0.04	0.06	0.05	0
Std. Dev.	1	1	1	1
Skewness	-0.3	-0.3	-0.3	0
Excess kurtosis	2.4	2.9	1.7	0
KS Statistic	0.03	0.05	0.04	0
p-value	< 0.001	< 0.001	< 0.001	N/A

Table 4.6: Key statistics of the normalized residuals at different cleaning stages: after Step 2 (958 time-series), after Step 3 (457 time-series), and the cleaned dataset (120 time-series). Kurtosis is excess kurtosis, measured relative to the standard normal, which has excess kurtosis 0. The residuals have a slight negative skew and a positive excess kurtosis that reflects heavier tails.

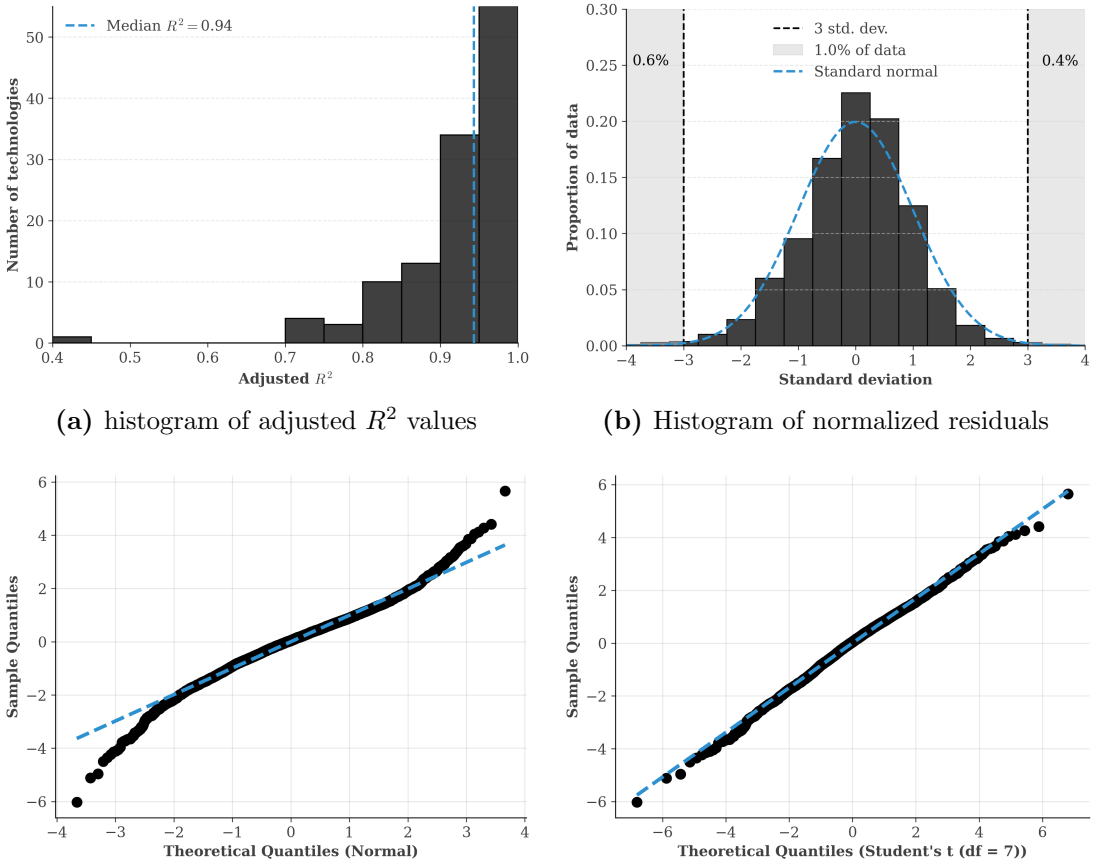
Figure 4.13a indicates that the curve explains $> 90\%$ of the variance for most technologies, with median adjusted $R^2 = 0.94$.[18] One technology has adjusted $R^2 < 0.5$ (*Gold reserves (South Korea)*, $R^2 = 0.45$).

Figure 4.13b shows a histogram of normalized residuals. Table 4.6 shows a slight

[16]For a comparison to other traditional S-curves (Logistic and Gompertz) see Section 4.15.

[17]df $= N_i - 4$, where $N_i = 11$, is the length of the shortest time-series in the dataset, and 4 is the number of estimated parameters, $\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\beta}_i$.

[18]Results remain unchanged when assuming universal $\beta = \frac{2}{3}$.



(a) histogram of adjusted R^2 values

(b) Histogram of normalized residuals

(c) Q-Q plot against normal distribution

(d) Q-Q plot against student t-distribution (df = 7)

Figure 4.13: **(a)** Histogram of adjusted R^2 values for the cleaned dataset of 120 technologies, showing most technologies have high explained variance. **(b)** Histogram of normalized residuals shows slight negative skew and heavier negative tails. **(c)** Q-Q plot comparing normalized residuals to standard normal distribution and **(d)** Student-t distribution with 7 degrees of freedom (df $= N_i - 4$, where $N_i = 11$, is the length of the shortest time-series in the dataset). The residuals deviate from normality in the tails, particularly on the negative side.

negative skew (-0.3) and a positive excess kurtosis (1.7). A positive excess kurtosis means the residuals have heavier tails than a normal distribution. This is common in real data. Consistent with this, 1% of the residuals lie past 3 standard deviations, where a normal distribution expects 0.3%. Most technologies track the B-R curve closely. A few show anomalous episodes, such as early rapid adoption or structural breaks, that produce occasional extreme residuals. The Q-Q plot against a normal distribution (Figure 4.13c) confirms this, with deviations from normality in the tails. The residuals match a Student-t distribution with 7 degrees of freedom (Figure 4.13d) more closely, which also has heavier

tails (excess kurtosis 2). The deviations are larger in the negative tail (black dots below the blue line at negative values in Figure 4.13d). These may stem from the systematic factors discussed earlier (Figure 4.7). Some technologies grow faster than the model early on. The model then overestimates the early growth rate and produces extreme negative residuals. It underestimates later diffusion points, which gives a median above 0. Section 4.14 discusses faster-than-expected early growth.

4.12 Limitations

Our data selection procedure faces several limitations.

Selection bias toward successful technologies. While we gather as much data as possible from secondary sources, the data collected in these sources tends to focus on successful technologies.

Subjective selection decisions. Several steps in the cleaning procedure involve subjective judgment: manual correction of death-phase breakpoints (Step 1), manual flagging of poor-quality series (Step 4b), and manual outlier removal (Step 5). While we validate subjective decisions with quantitative checks where possible, these choices introduce analyst degrees of freedom that could in principle affect the results. We have aimed to be transparent about each such decision and its rationale, and show that after each of these steps the change in the distribution of normalized residuals is minimal, and change in quality of fit (measured by adjusted R^2) is negligible.

Data-construction assumptions. When aggregating country-level data to a global proxy, missing values are treated as zeros (Section 4.11.1), which may understate early adoption levels if reporting coverage is incomplete. Similarly, converting flow data to cumulative stocks assumes zero initial stock, which can bias levels if a technology was already present before the sample begins. Both effects are most pronounced in the early

part of the time-series—the segment most relevant for the predictability exercise.

Subjective technology definitions. Determining when a time-series is or is not a technology may be difficult. We have not defined a strict criterion for what constitutes a technology, and we have made some subjective decisions in this regard. For example, we have removed *GDP* as it does not represent technology adoption, but we have retained *monasteries* as they represent a form of organizational technology. Other non-technology time-series are present, but were filtered out in earlier steps.

Conditioning on completed S-curves. Our model assumes an S-curve diffusion pattern, so we automatically exclude technologies that do not (or do not yet) feature a slowing down of diffusion. In other words, our dataset includes only technologies that have gone through a full diffusion cycle. The filtering criteria in Steps 2–4 (minimum data points, requiring $> 80\%$ of $\hat{L}$, requiring early-phase coverage) further condition the sample on well-behaved, already-mature technologies. As a consequence, the universality and backtest results reported in this paper are conditional on technologies whose diffusion is well-described by an S-curve. S-curve diffusion is therefore an *assumption* of our model. Conditional on this assumption, it is a *result* of our paper that a technology's diffusion can be predicted with a well-calibrated distributional forecast using the B-R model with technology-specific and universal parameters, even from early data alone. We note that this does not exclude technologies that were later substituted out (e.g., *telegrams*). That is, we model only the first three S-curve phases of net-stock growth, not the full birth-to-death life cycle, as the death phase is trimmed in Step 1 and the preceding diffusion is retained (Section 4.6).

4.13 Defining technology-invariant parameters

The full B-R model we consider here is a highly nonlinear model with a total of 7 free parameters: L, k, t_0, σ, β, ρ and ϵ_T. The terminal residual ϵ_T enters the model because the MA(1) noise carries the last observed shock forward into the first forecast step; it could in principle be estimated technology-specifically, but doing so adds a free parameter for each technology. Combined with the small number of data points (often as few as 10), fitting all these parameters may very likely result in overfitting. To improve this situation we take three parameters as technology-invariant, i.e. equal for all technologies: $\beta = 2/3$, $\rho = 0.80$, and $\epsilon_T = 0$. This introduces some bias (real technologies may differ in these parameters) but reduces variance and improves forecasts. We show below that observed variation in these parameters across our data set is small enough that real differences are hard to discern given the statistical error of short series. This section justifies these choices through theoretical considerations, empirical evidence, and support from the literature.

4.13.1 Asymmetry parameter

The asymmetry parameter β determines S-curve skewness: $\beta = 1$ yields symmetric curves (logistic), $\beta < 1$ yields right-skewed curves (inflection before 50% diffusion), $\beta > 1$ yields left-skewed curves (inflection after 50% diffusion). Here, the inflection point refers to the true inflection (maximum growth rate), which occurs at $Y_{\text{inf}}/L = (1+\beta)^{-1/\beta}$. Note that this differs from $Y(t_0)/L = 2^{-1/\beta}$, since t_0 is not the inflection point for $\beta \neq 1$. Both expressions equal 50% at $\beta = 1$.

Literature context. Mahajan et al. [41] review diffusion models, noting S-curve shape variations across technologies. Meade and Islam [20] conclude that the shape of the S-curve is one of the most important and distinctive features of the diffusion process. Bertalanffy [64] introduced asymmetric growth models, showing right-skewed S-curves fit

biological growth data. Rogers [40] and Moore [115] identify that near-exponential adoption occurs in the first 16% of adoption. Grubler [15] documents exponential growth patterns in infrastructure diffusion during initial market penetration phases. Despite available asymmetric models [14], many studies such as the first study of technology adoption using S-curves [8] default to symmetric logistic curves for simplicity.

4.13.1.1 In-sample fit and choice of asymmetry parameter

We first analyze the in-sample fit for various values of a technology-invariant β, and then check how well a DGP with technology-invariant β replicates the empirical distribution of $\hat{\beta}_i$ obtained when we estimate β_i separately for each technology. We use the empirical surrogate datasets defined in Section 4.9, with β fixed to a candidate technology-invariant value (rather than estimated per technology) and the remaining $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ inherited from the empirical fit. We then re-fit the B-R model with technology-specific β_i to each surrogate series and compare distributions. If the surrogate and empirical distributions of $\hat{\beta}_i$ are similar, the observed heterogeneity in $\hat{\beta}_i$ is consistent with estimation error rather than genuine cross-technology variation, and a technology-invariant β is justified for forecasting given the short data sets.

Goodness of fit for different values of β. Figure 4.14 shows fit quality across the 120-technology ensemble for four S-curve choices: logistic, Gompertz, B-R with technology-invariant $\beta = 2/3$, and B-R with technology-specific β_i. Figure 4.14a shows that all S-curve types yield similar adjusted R^2 distributions, with median values above 0.9, indicating high fit quality across models (though the Gompertz model does somewhat worse). B-R with varying β_i yields similar median R^2 to technology-invariant β models ($\beta = 1$ and $\beta = 2/3$), indicating that estimating technology-specific β_i does not substantially improve in-sample fit quality.

Figure 4.14b focuses on fit quality when using a technology-invariant β. Adjusted R^2

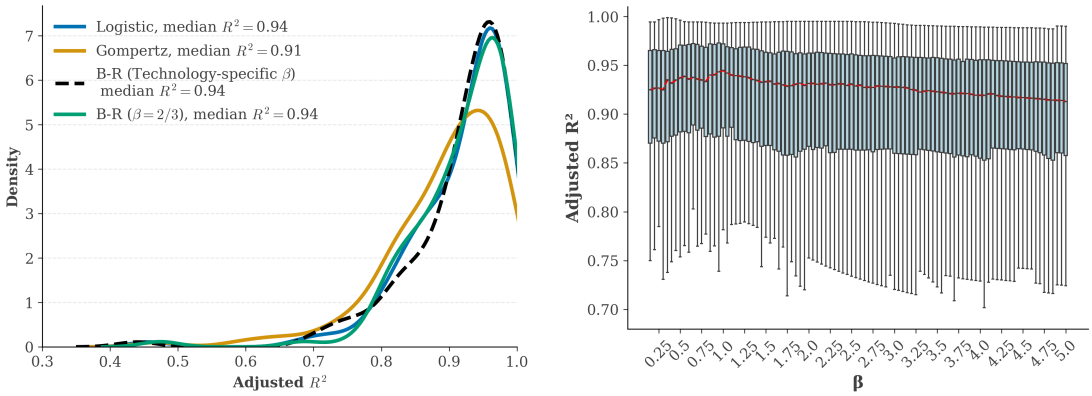


Figure 4.14: (a) Comparison of R^2 across Gompertz, logistic, and B-R curves (technology-specific β_i and technology-invariant $\beta = 2/3$). **(b)** Box-plots of adjusted R^2 of 120-technology ensemble for various values of a technology-invariant $\beta \in [0.05, 5]$.

is relatively stable across β, but very small ($\beta < 0.3$) and very large ($\beta \gg 1$) values yield slightly lower R^2. This implies that a wide range of S-curve models provide reasonable fits, and in-sample R^2 alone is insufficient for model selection. Section 4.15 compares Gompertz, logistic ($\beta = 1$), and Bertalanffy-Richards ($\beta = 2/3$) curves in more detail.

Replicating the distribution of estimated β_i. Given the short data sets, we expect substantial heterogeneity in *estimated* β_i even if all the technologies had the same true value of this parameter, i.e. if the *population* $\beta_i = \beta$, $\forall i$. We assess whether generating data using a technology-invariant $\beta_i = \beta$ can lead to an empirical distribution of estimated β_i similar to what we observe empirically, following Algorithm 4.

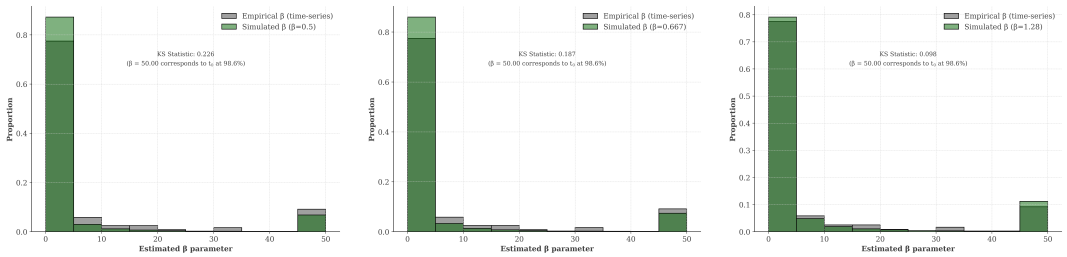
Algorithm 4: Assessing whether a DGP with technology-invariant β replicates the empirical distribution of β_i.

Input: Empirical technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Distribution of estimated β_i from surrogate data
Step 1: Estimate parameters from empirical data
  foreach *technology* $i = 1$ *to* 120 **do**
    Fit model with β to obtain:
    $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$
    ($\hat{L}_i$ via NLLS; apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ and regress on time to estimate other parameters)
  end
Step 2: Generate and fit surrogate data (M simulations)
  foreach *simulation* $s = 1$ *to* M **do**
    foreach *technology* $i = 1$ *to* 120 **do**
      Generate $Y_i^{(s)}(t)$ using DGP (Eq. 4.17) with $\hat{\theta}_i$, β, $\rho = 0.8$, and $\epsilon_T = 0$
      Fit Bertalanffy-Richards curve to $Y_i^{(s)}(t)$ to estimate $\hat{\beta}_i^{(s)}$
    end
  end
Step 3: Compare distributions
  Compute Kolmogorov-Smirnov distance:
$D_{KS}(\beta) = \sup_x |F_{\text{real}}(x) - F_{\text{surrogate}}(x)|$

Figure 4.15 compares $\hat{\beta}_i$ values estimated from surrogate data generated under a technology-invariant B-R model with $\beta \in \{0.5, 2/3, 1\}$. We show two probability histograms: (1) the full histogram with all estimated values, and (2) one cropped to $\beta_i < 2.41$. The clip at $\beta_i = 2.41$ corresponds to $Y(t_0)/L = 2^{-1/\beta} < 75\%$ (recall t_0 is not the inflection for $\beta \neq 1$; the true inflection at $Y_{\text{inf}}/L = (1+\beta)^{-1/\beta}$ corresponds to $\approx 58\%$ diffusion at $\beta = 2.41$). This is consistent with the 80% minimum-diffusion data-selection threshold (Section 4.11). In the fitting procedure, we set an upper bound $\beta = 100$.

Table 4.7 shows the Kolmogorov-Smirnov (K-S) distance $D_{KS}(\beta)$ between estimated β_i distributions from real and surrogate data for various values of β, as well as other statistics.

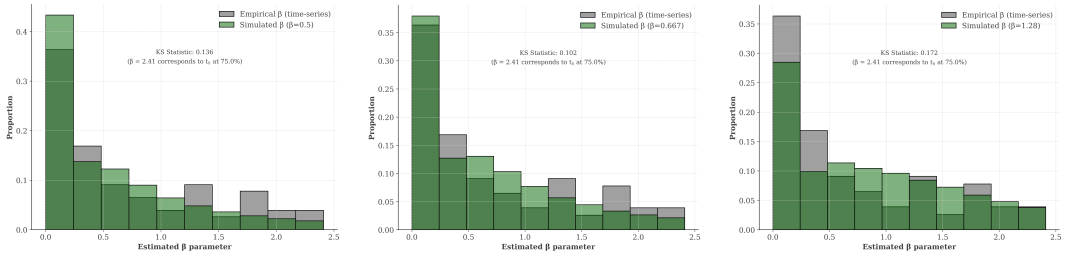
The empirical data shows substantial heterogeneity in estimated β_i values across technologies. The median estimated β_i is 1.28, with a mean of 11.6 due to a few high β_i outliers. Just over half of technologies have $\beta > 1$ (left-skewed S-curves; maximum growth after 50% diffusion) and the rest have $\beta < 1$ (right-skewed; maximum growth before 50%). This implies that there is no clear universal asymmetry pattern.



(a) Full histogram, technology-invariant $\beta = 0.5$

(b) Full histogram, technology-invariant $\beta = 2/3$

(c) Full histogram, technology-invariant $\beta = 1.28$



(d) Technology-invariant $\beta = 1/2$

(e) Technology-invariant $\beta = 2/3$

(f) Technology-invariant $\beta = 1.28$

Figure 4.15: Histograms comparing $\hat{\beta}_i$ estimated from surrogate data (green) with technology-invariant $\beta = 2/3$ based on 1000 simulations for 120 technologies versus β_i estimated from empirical data (gray). The top panels show the histogram with the x-axis cropped at 50 for visualization purposes. The bottom panels are the same, but they are cropped with $\beta_i \leq 2.41$, corresponding to $Y(t_0)/L < 75\%$. The left, middle and right panels correspond to three different values of the technology-invariant β used to create the surrogate data.

Data source	Mean	Median	D_{KS} (full)	p-value (full)	D_{KS} ($\beta_i < 2.41$)	p-value ($\beta_i < 2.41$)
Empirical data	11.6	1.28	–	–	–	–
Surrogate ($\beta = 1/2$)	7.9	0.57	0.23	$< 10^{-4}$	0.14	0.10
Surrogate ($\beta = 2/3$)	8.6	0.72	0.19	6×10^{-4}	0.10	0.38
Surrogate ($\beta = 1$)	10.7	1.04	0.12	0.06	0.14	0.08
Surrogate ($\beta = 1.28$)	12.7	1.32	0.11	0.16	0.17	0.02

Table 4.7: Mean and median of estimated $\hat{\beta}_i$ values from empirical data and surrogate data (technology-invariant $\beta = 0.5$, $\beta = 2/3$, $\beta = 1$, and $\beta = 1.28$), along with Kolmogorov-Smirnov distances D_{KS} and p-values comparing surrogate to empirical data distributions. Lower K-S distances and higher p-values indicate better replication of the empirical distribution.

Figure 4.15 shows the histogram of individually estimated β_i with a cutoff at $\beta_i < 50$ for visualization purposes). The green histogram shows estimated β_i values from surrogate data generated with technology-invariant value of $\beta \in \{1/2, 2/3, 1.28\}$.

The surrogate data with technology-invariant $\beta = 1.28$ (4.15c),[19] equivalent to the empirical median, has the lowest KS distance of those shown (Table 4.7), but it under-emphasizes the peak at low values of β ($\beta < 0.25$) in the empirical distribution (Figure 4.15f versus 4.15d and 4.15e). $\beta = 0.5$ and $\beta = 2/3$ (Figures 4.15d and 4.15e) both reasonably replicate the spread of estimated β_i values, the extreme outliers, and the peak at low β values. In both cases the surrogate slightly underestimates the empirical mean and median (Table 4.7), shown by the lighter green areas on the left of the histograms in Figures 4.15a and 4.15b.

Focusing on the distribution of estimated β_i values with $Y(t_0)/L < 75\%$ ($\beta_i < 2.41$, Figures 4.15d, 4.15e, and 4.15f), surrogate data with technology-invariant $\beta = 2/3$ best replicates the clipped empirical distribution (Table 4.7). $\beta = 1/2$ slightly over-emphasizes the peak at low β values ($\beta < 0.25$) in the empirical distribution compared to $\beta = 2/3$ (Figure 4.15d versus 4.15e).

We conclude that fixing β can reasonably replicate the empirical distribution of estimated β_i values, but no single value dominates across all criteria. When matching the full empirical distribution, $\beta = 1.28$ achieves the lowest KS distance (Table 4.7). However, the surrogate experiments demonstrate that β estimates are biased upward in the mean: surrogate data with true $\beta = 2/3$ produces mean $\hat{\beta} = 8.6$ (vs. median $\hat{\beta} = 0.72$), and true $\beta = 1.28$ produces mean $\hat{\beta} = 12.7$ (vs. median $\hat{\beta} = 1.3$), due to a heavy right tail of extreme outliers. Additionally, $\beta = 2/3$ best replicates the clipped distribution ($\beta_i < 2.41$, KS $p = 0.38$), while $\beta = 1.28$ is rejected under this clip ($p = 0.02$). $\beta = 0.5$ slightly over-emphasizes the peak at low β values, while $\beta = 1$ and $\beta = 1.28$ under-emphasize it.

[19] Similar conclusions apply to $\beta = 1$ (logistic). We do not show the graphs, but we show the statistics in Table 4.7.

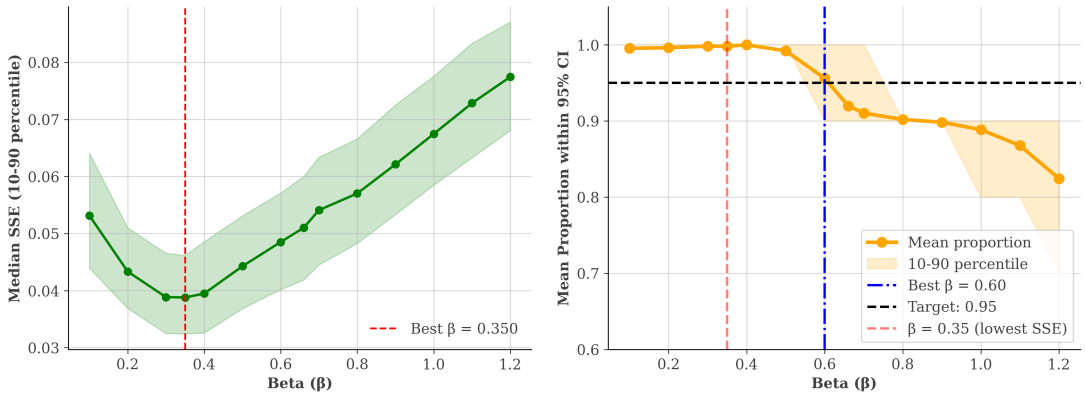
4.13.1.2 Growth replication

We assess the performance of models with technology-invariant β in replicating observed growth patterns. The chosen value $\beta = 2/3$ is shown against the data in main paper Figure 2d.

Replicating observed growth patterns. For each candidate technology-invariant β, we apply Algorithm 9 with the B-R DGP at that β in place of the default $\beta = 2/3$, generating 1,000 surrogate ensembles. From each ensemble and from the empirical pool we compute the cross-technology median growth trajectory (a sequence of medians indexed by diffusion level, not a smooth curve), and compare them via the sum of squared errors (SSE). We repeat the surrogate step 100 times and report the SSE distribution. We also report the proportion of empirical-median points falling within the 95% surrogate-median confidence interval; perfect replication gives 95%.

Figure 4.16 shows 100 SSEs as solid lines, with shaded regions indicating the $10^{\text{th}} - 90^{\text{th}}$ percentiles for each β. Figure 4.16a shows that $\beta = 0.35$ yields the lowest SSE, indicating the best replication of observed median growth patterns. SSE increases consistently for $\beta > 0.4$, with $\beta = 1$ (logistic) yielding a median SSE nearly 2 times higher than $\beta = 0.35$. Similarly for very low β values ($\beta < 0.2$), SSE increases. Figure 4.16b shows the proportion of points on the empirical median growth trajectory that fall within the 95% confidence interval of the surrogate median growth trajectory for each technology-invariant β value. Lower β yields higher coverage: $\beta < 0.4$ captures essentially all empirical points within the 95% CI, suggesting the surrogate CI is too wide. Higher β yields lower coverage: $\beta > 0.8$ captures fewer points, and $\beta = 1.5$ only 70–80%, indicating poorer replication of observed growth patterns. For $\beta \in [0.5, 0.8]$, the proportion fluctuates between $90 - 100\%$, with the median near 95%.[20] For $\beta = 0.6$ and $\beta = 0.66$, the mean proportion is 96% and 92%

[20]The empirical median growth trajectory has 10 points, so we expect 9-10 points within the 95% CI. Over 1000 simulations, this yields a mean coverage of approximately 95%.



(a) SSE of technology-invariant β

(b) Proportion of realized median within 95% CI

Figure 4.16: Sum of squared errors (SSE) between median growth trajectories from surrogate data and empirical data across varying technology-invariant β values. Lower SSE indicates better replication of observed growth patterns. **(a)** Median SSE across 100 repetitions of 1000 simulations of the 120-technology ensemble shown as solid line; shaded region indicates the $10^{\text{th}} - 90^{\text{th}}$ percentile. **(b)** Plot showing the proportion of empirical median growth trajectory points that fall within the 95% confidence interval of the surrogate median growth trajectory for each technology-invariant β value. Shaded region indicates $10^{\text{th}} - 90^{\text{th}}$ percentile across 100 repetitions. Perfect replication would yield 95% of points within the 95% CI.

respectively, indicating good replication of observed growth patterns.[21] Analyzing the 90% CI yields similar results, with $\beta \in \{0.6, 0.66\}$ both yielding mean proportions of 89%.

Conclusion: Justification for a technology-invariant β, with $\beta = 2/3$. Models with predetermined inflection points (Gompertz [11], Logistic [12]) and flexible inflection points (Bass [4], Richards [14]) all fit adoption data reasonably well [44, 49, 66, 69] (Section 4.15). However, the large variation in estimated β_i values indicates estimation uncertainty, which increases overfitting risk and degrades out-of-sample forecasts [31].

Multiple technology-invariant values of β yield similar fit quality (Figure 4.14b). When replicating the distribution of estimated β_i values (Figure 4.15), multiple technology-invariant β values work well, e.g., $\beta = 2/3$. Lower β slightly over-emphasizes the peak at low β values in the distribution. $\beta > 2/3$ have lower K-S distances for the full distribution, but under-emphasize the peak at low β values. Out-of-sample growth replication (Figure 4.16a)

[21] The last empirical growth data-point tends to fall slightly outside the 95% CI, indicating potential systematic model misspecification. See Figure 4.20b.

indicates $\beta \in [0.2, 0.4]$ best replicates observed growth patterns (lowest SSE) favoring $\beta < 1$ (right-skewed S-curves), where $\beta = 0.35$ yields the lowest SSE. $\beta \in [0.5, 0.8]$ tends to yield the expected coverage of 95% of points within the 95% CI (Figure 4.16b), across the 1000 simulations.

No single value of β is best on every criterion. The full-distribution KS test favors $\beta = 1.28$, the growth-replication SSE favors $\beta \approx 0.35$, and the clipped-distribution KS test favors $\beta = 2/3$. We choose $\beta = 2/3$ as a pragmatic compromise: (i) the upward estimation bias documented above means the empirical median 1.28 overstates the true β; (ii) $\beta = 2/3$ provides the best match to the clipped distribution; (iii) $\beta \in [0.5, 0.8]$ yields approximately the expected 95% coverage in the growth-replication exercise. Section 4.16 confirms that $\beta = 2/3$ yields reasonable out-of-sample forecasting performance: $\beta = 1/3$ and $\beta = 1.5$ are significantly worse, while $\beta = 0.5$ and $\beta = 1$ (logistic) perform similarly but slightly worse.

4.13.2 Moving average parameter

The MA(1) parameter ρ in our DGP (Eq. 4.17) controls temporal correlation in the noise: $\sigma_i(\rho\epsilon_{i,t-1} + \epsilon_{i,t})$, where $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$. This parametrization means the noise at time t depends on both the current white noise draw $\epsilon_{i,t}$ and the white noise draw from the previous time step $\epsilon_{i,t-1}$, weighted by ρ.

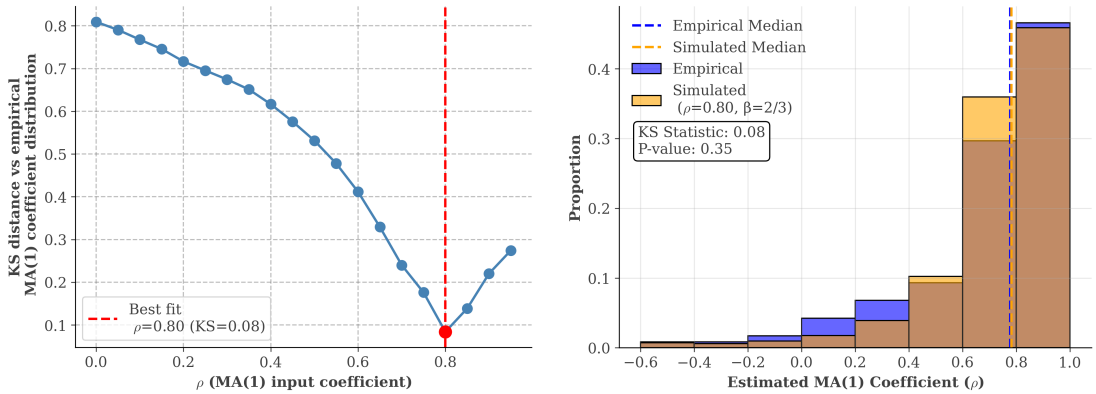
Literature support and justification. Baumgartner et al. [31] discuss positive autocorrelation ($\rho > 0$) in residuals of technology diffusion. Whether observed autocorrelation arises from unmodeled factors, model misspecification, or inherent adoption dynamics remains unclear. Autocorrelation is difficult to estimate accurately and is sometimes technology-invariant for parsimony [30]. We use a technology-invariant ρ rather than estimate technology-specific ρ_i values.

Estimating the MA(1) parameter from empirical data. To estimate ρ_i from empirical data, we fit the B-R model (Eq. 4.6) via NLLS to estimate $\hat{L}_i$ and $\hat{\beta}_i$, transform the coordinate to a linear model by applying the GFP transform $f(Y_{i,t}; \hat{L}_i, \hat{\beta}_i)$, regress the transformed data on time, and obtain residuals $S_i(t)$. We estimate ρ_i using Maximum Likelihood Estimation (MLE) via an ARIMA(0,0,1) model fitted to the residuals. We then assess how well a technology-invariant ρ replicates the empirical distribution of $\hat{\rho}_i$ using empirical surrogate datasets (Section 4.9), generated with the candidate technology-invariant ρ rather than the per-technology $\hat{\rho}_i$. Algorithm 5 summarises the procedure.

Algorithm 5: Assessing whether a technology-invariant DGP parameter ρ replicates the empirical distribution of estimated MA(1) parameters.

Input: Empirical technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Distribution of estimated ρ_i values from surrogate data
Step 1: Estimate parameters from empirical data
  foreach *technology* $i = 1$ *to* 120 **do**
    Fit B-R curve to obtain $\hat{\rho}_i$:
    Fit $\hat{L}_i$ and $\hat{\beta}_i$ via NLLS; apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \hat{\beta}_i)$ and estimate other parameters via OLS
    Calculate residuals $S_i(t)$
    Estimate MA(1) parameter $\hat{\rho}_i$ using MLE via ARIMA(0,0,1) model fitted to residuals
  end
Step 2a: Generate surrogate data (M simulations)
  foreach *simulation* $s = 1$ *to* M **do**
    foreach *technology* $i = 1$ *to* 120 **do**
      Generate $Y_i^{(s)}(t)$ using the DGP (Eq. 4.17) with parameters $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ (estimated using $\beta = 2/3$ and technology-invariant $\rho = \tilde{\rho}$), $\beta = 2/3$, technology-invariant ρ, and $\epsilon_T = 0$
    end
  end
Step 2b: Estimate $\hat{\rho}_i^{(s)}$ from surrogate data
  foreach *simulation* $s = 1$ *to* M **do**
    foreach *technology* $i = 1$ *to* 120 **do**
      Estimate $\hat{L}_i^{(s)}$ and $\hat{\beta}_i^{(s)}$ from NLLS regression on $Y_i^{(s)}(t)$
      Apply $f(Y_i^{(s)}, \hat{L}_i^{(s)}, \hat{\beta}_i^{(s)})$ (GFP transform) and regress on time
      Calculate residuals $S_i^{(s)}(t)$
      Estimate MA(1) parameter $\hat{\rho}_i^{(s)}$ using MLE via ARIMA(0,0,1) model fitted to residuals
    end
  end
Step 3: Collect estimated $\hat{\rho}_i^{(s)}$ values
  foreach *simulation* $s = 1$ *to* M **do**
    Store $\hat{\rho}^{(s)} = \{\hat{\rho}_i^{(s)}\}_{i=1}^{120}$
  end
Step 4: Collect estimated $\hat{\rho}^{(s)}$ values across simulations
Store $\hat{\rho}_M = \{\hat{\rho}^{(s)}\}_{s=1}^{M}$
Step 5: Compare MA(1) parameter distributions
  Compute Kolmogorov-Smirnov distance:
$D_{KS}(\rho) = \sup_x |F_{\text{empirical}}(x) - F_{\text{surrogate}}(x)|$

Results: replicating estimated MA(1) parameter distribution. Figure 4.17 presents the results of applying Algorithm 5 for $M = 1000$ simulations across technology-invariant ρ values ranging from 0 to 0.95 in steps of 0.05. Figure 4.17a shows the KS distance between the empirical distribution of estimated MA(1) parameters $\hat{\rho}_i$ from empirical



(a) KS distance for varying ρ **(b)** Histogram comparison for $\rho = 0.8$

Figure 4.17: (a) KS distance between estimated MA(1) parameter distributions from surrogate data versus empirical data for varying technology-invariant ρ values. Lower KS distances indicate more similar distributions. **(b)** Histograms comparing estimated MA(1) parameter distributions from surrogate data (orange) versus empirical data (blue) for technology-invariant $\rho = 0.8$.

data and that from surrogate data for varying technology-invariant ρ values. Using technology-invariant $\rho = 0.8$ (Figures 4.17a, 4.17b) reasonably replicates the estimated MA(1) parameter distribution, with Kolmogorov-Smirnov distance $D_{KS} = 0.08$ and p-value 0.35, making it the only tested value not significantly different from the empirical distribution. The empirical histogram shows more mass between -0.2 and 0.4, which is not well captured by any tested ρ value.[22]

Impact of ρ on forecasts. Autocorrelation and stochastic shocks increase the uncertainty of forecasts, but play a relatively minor role in forecast variance compared to parameter uncertainty, as discussed in Section 4.16. The choice of ρ therefore has limited impact on the overall forecast distribution.

[22]Results remain consistent when using $\hat{L}_i$ from NLLS fitting using technology-varying $\hat{\beta}_i$, which has a minimal effect on the $\hat{L}_i$ estimate. In that case $\rho = 0.8$ has lowest KS-distance, with only $\rho = 0.8$ and $\rho = 0.85$ yielding non-significant p-values (p-value > 0.05).

4.13.3 Terminal residual

The terminal residual $\epsilon_{i,T}$ is the standardised noise at the final observation time T. In the MA(1) process, the noise at time T is $\sigma_i(\rho\epsilon_{i,T-1} + \epsilon_{i,T})$, which propagates into the forecast at $T+1$ through $\sigma_i(\rho\epsilon_{i,T} + \epsilon_{i,T+1})$. One could estimate technology-specific $\epsilon_{i,T}$ values as additional parameters during Bayesian inference, but this adds a free parameter per technology and increases overfitting risk.

By construction of our noise model, $\epsilon_{i,T} \sim \mathcal{N}(0,1)$, so $\mathbb{E}[\epsilon_{i,T}] = 0$. We therefore set $\epsilon_{i,T} = 0$ universally: for each technology, we take the realized noise at the final observation to be at its expected value.

In practice, the forecast distribution is also anchored to the last observed value via a level-shift applied at the forecast origin (Section 4.7). This is distinct from the MA(1) terminal residual and does not contradict the choice $\epsilon_{i,T} = 0$.

4.13.4 Summary and limitations

We use three technology-invariant parameters to reduce overfitting and simplify the model: (1) $\beta = 2/3$, (2) $\rho = 0.80$, and (3) $\epsilon_T = 0$. Empirical data shows substantial heterogeneity in estimated β_i and ρ_i values. However, surrogate data generated with technology-invariant parameters reasonably replicates the empirical distributions of estimated β_i and ρ_i values (Figures 4.15 and 4.17). This suggests that observed heterogeneity in parameter estimates may reflect estimation uncertainty rather than true technological heterogeneity, though the technology-invariant parameters chosen here may not fully capture all observed heterogeneity. Section 4.16 compares forecasting performance using these technology-invariant values versus technology-specific estimates.

4.14 Testing for empirical regularities

This section examines regularities in the 120 time-series dataset. First, we test whether technologies grow exponentially at early diffusion stages, as reported in [15, 23, 29, 31]. Second, we compare empirical growth-rate behavior against the empirical surrogate datasets of Section 4.9 generated with $\beta = 2/3$. Third, we compare the stochastic variation in the empirical data to the same surrogates across diffusion levels. Finally, we test three additional regularities noted in the literature.

Note on box plots. All box plots in this section (and the main text) are binned for equal time intervals or diffusion intervals. Therefore, box plots do not necessarily contain the same number of data points. We make this choice to account for the non-linear diffusion process. Binning similar time or diffusion points is more informative than binning equal numbers of points, which may span large ranges of time or diffusion levels and obfuscate trends.

Box plots follow the standard definition: the box spans the interquartile range (IQR, 25th to 75th percentile), a horizontal line within the box indicates the median, and whiskers extend to the smallest and largest values within $1.5 \times$ IQR of Q1 and Q3, unless otherwise stated. We exclude points beyond the whiskers.

4.14.1 Do technologies grow exponentially?

The B-R family of models has exponential growth at early stages. Approximate exponential growth has been observed in many cases [23], but there are also cases in which this assumption is violated [91]. Here we examine the extent to which the technologies in our database exhibit approximate exponential behavior in early stages and flag the technologies for which this seems to be a poor assumption.

Setup. We define the early diffusion stage as the first 10% of diffusion. We identify this by estimating the asymptote L_i via NLLS and keeping all observations $Y_{i,t}$ such that $Y_{i,t}/L_i \leq 0.10$. Results are similar when using $L_i := \max_t Y_{i,t}$. When fewer than 6 observations exist in the first 10%, we use the first 5 observations of the series.

For each technology, we then calculate yearly growth rates as the annualized log-differences:

$$\Delta \log Y_{i,h}(t) = \frac{\log Y_i(t) - \log Y_i(t-h)}{h}, \tag{4.47}$$

where h is the number of years between consecutive observations (typically $h = 1$; we use larger h only when data gaps prevent adjacent-year pairs). Under deterministic exponential growth, $\Delta \log Y_{i,t}$ is constant. To test this, we regress $\Delta \log Y_{i,t}$ on time as

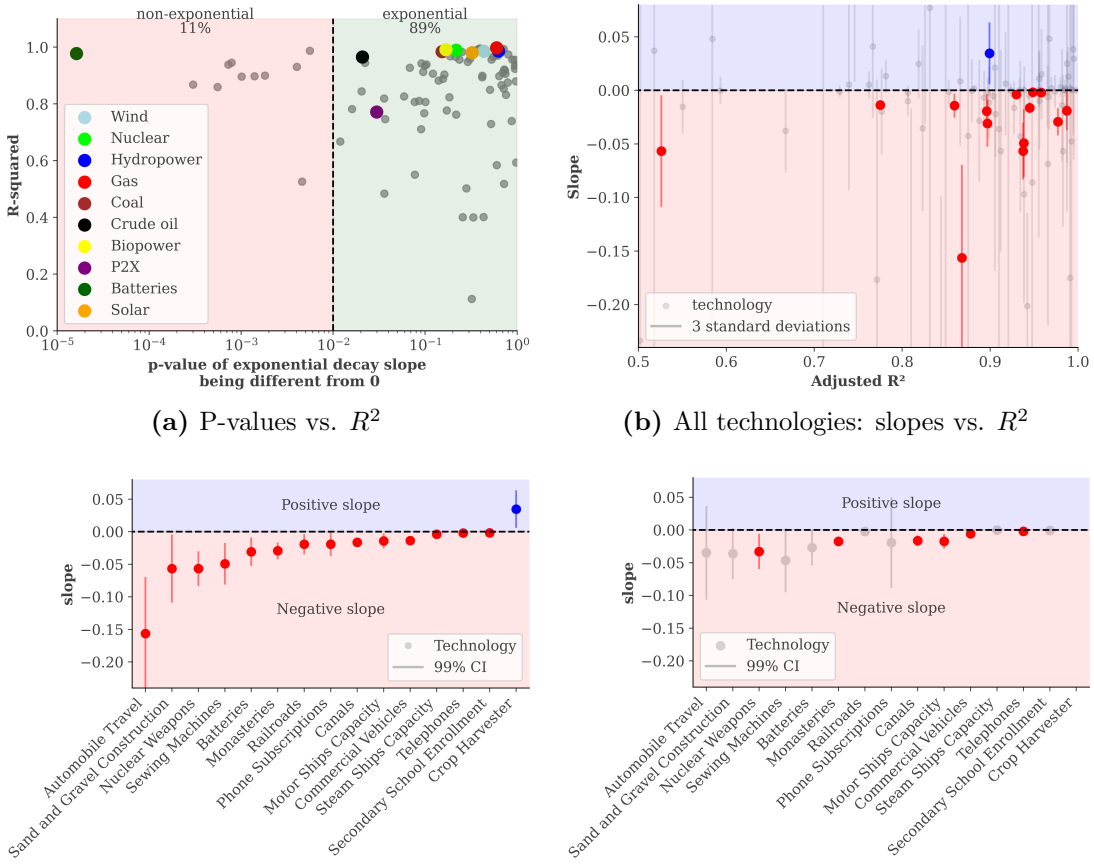
$$\Delta \log Y_{i,t} = \alpha_i + \gamma_i t + \epsilon_{i,t}, \tag{4.48}$$

and test $H_0 : \gamma_i = 0$ using the t-statistic. We use $p < 0.01$ as the threshold for rejecting exponential growth. The sign of significant γ_i values is informative: negative γ_i would indicate the rapid-then-stabilising pattern reported by Comin et al. [91].

Results. Figure 4.18a shows p-values for the slope from regressing $\Delta \log Y_i(t)$ on time, as per (4.48), against R^2.[23] That is, we test whether growth rates are constant (exponential growth) at early diffusion stages. Testing at the 1% significance level, 89% of the 131 technologies have slopes not significantly different from zero. This means 15 technologies have significant slopes (14 from the 120-technology set plus *Batteries* from the TWh energy data). Figure 4.18b shows all the estimated slopes plotted against R^2. All but one significant slope are negative. This is in line with Comin et al. [91], who report fast initial growth that levels off to exponential growth. Figure 4.18c shows that the significant

[23] All analysis uses the 120-technology set. For Figure 4.18a, we additionally include the 11 useful-energy technologies in TWh introduced in Section 4.11 (131 series total).

technologies span multiple industries, with slopes predominantly negative and near 0. Figure 4.18d shows the same significant technologies after excluding the first 0.5% of diffusion, revealing that most become non-significant.[24] This suggests that the initial rapid growth occurs at very low diffusion levels ($< 0.5\%$), after which growth rates stabilize.



(a) P-values vs. R^2 **(b)** All technologies: slopes vs. R^2

(c) Significant technologies (0–10% diffusion) **(d)** Significant technologies (0.5–10% diffusion)

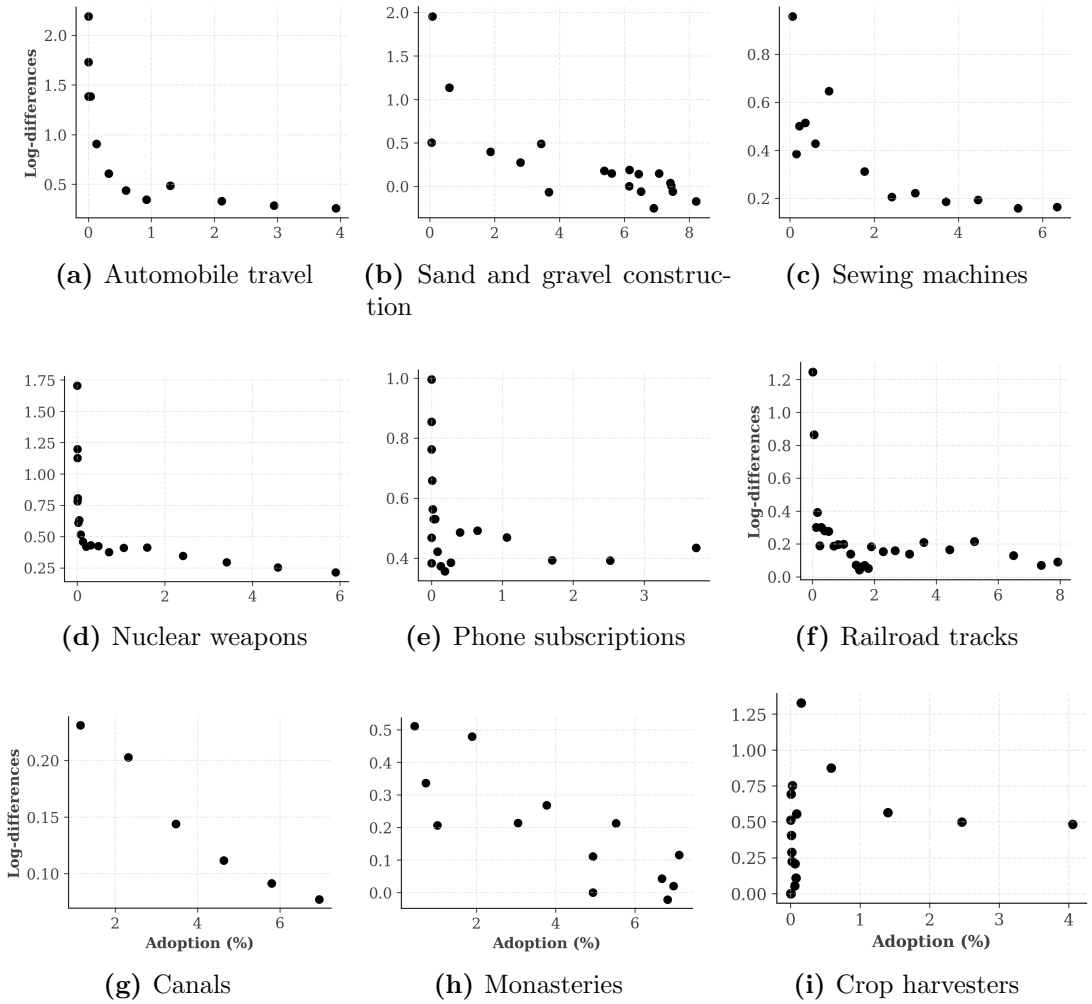
Figure 4.18: Testing whether technologies exhibit exponential growth at early diffusion stages (0–10%). **(a)** P-values vs. R^2 for 131 technologies (120-technology set plus 11 TWh energy technologies, highlighted). **(b)** Slopes vs. R^2 for all technologies; 89% have non-significant slopes (gray). **(c)** Slopes for the 15 significant technologies. **(d)** Same technologies after excluding first 0.5% of diffusion; most become non-significant. Those with fewer than 5 data points after exclusion are omitted.

Figure 4.19 further zooms in on 9 of the 15 technologies with significant slopes,[25] showing the annual growth rates against the adoption level.

Figure 4.19 shows that technologies (a–f) grow rapidly at $< 0.5\%$ diffusion and slow to

[24] Those technologies with less than 5 data points after excluding the first 0.5% are omitted.

[25] The other 6 technologies (including *Batteries* from the TWh energy data) have slopes near 0 or behave similarly to those shown.



(a) Automobile travel **(b)** Sand and gravel construction **(c)** Sewing machines

(d) Nuclear weapons **(e)** Phone subscriptions **(f)** Railroad tracks

(g) Canals **(h)** Monasteries **(i)** Crop harvesters

Figure 4.19: Log-differences for technologies with significant slopes. 9 of 15 technologies shown (14 from the 120-technology set plus *Batteries* from TWh energy data; *Batteries* not shown). **(a)**–**(f)** have negative slopes under regression. Growth is rapid at $< 0.5\%$ diffusion, then decreases toward constant growth (exponential). **(g)** *Canals* and **(h)** *Monasteries* have decreasing growth (marginally significant). **(i)** *Crop harvesters* has positive slope: accelerating growth at $< 0.2\%$ diffusion, decelerating from 0.5% to 1.5%, then approximately exponential growth.

exponential by $\sim 1\%$ diffusion, consistent with [91]. *Canals* and *Monasteries* (Figures 4.19g, 4.19h) show marginally significant decreasing growth throughout. *Crop harvesters* (Figure 4.19i) shows a different pattern: accelerating growth, then decelerating, before reaching exponential behavior.

Limitations of the model at very early diffusion stages. The B-R curve assumes exponential growth at early diffusion stages. Figure 4.19 shows that 10% of technologies exhibit accelerated growth at $< 1\%$ diffusion, decaying to exponential. The

B-R model does not capture this behaviour, but the period is short and occurs at very low diffusion, so the impact on model fit and forecasts is minimal. Comin et al. [91] suggest the pattern may occur more broadly. Whether our 10% figure reflects reality or systematic data omission at very early stages is unclear.

4.14.2 Replicating observed growth behavior

This subsection provides the SI methodology and evidence behind the in-sample universality of growth paths shown in main paper Figure 2. Here we test whether the B-R curve with $\beta = 2/3$ approximates observed growth behavior by normalizing log-differences and time to pool technologies, and compare the empirical pool to the surrogate ensemble generated under the same DGP (Section 4.9).

Normalizing growth rates and time. At early diffusion ($t \ll t_0$), the B-R curve has $\Delta \log Y_{i,t} \approx k_i$ independently of β. We normalize log-differences by $\hat{k}_i$ (median growth rate from first 10% diffusion) and time by $\hat{k}_i(t - \hat{t}_{0,i})$:

$$\Delta \log^{(n)} Y_{i,t} = \frac{\Delta \log Y_{i,t}}{\hat{k}_i}, \quad t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$$

This collapses all technologies onto a universal curve: $\Delta \log^{(n)} Y_{i,t} \approx 1$ at early diffusion, ≈ 0 at late diffusion.

Mathematical justification. All B-R curves with the same β have identical shape in GFP-coordinates but differ in three technology-specific parameters: growth rate k_i, inflection time $t_{0,i}$, and asymptote L_i. At early diffusion ($t \ll t_0$, approximately $Y_{i,t} < 0.1L_i$), the GFP-transformed data satisfies $\log(Y_{i,t}/L_i) \approx k_i(t - t_{0,i}) + \epsilon_{i,t}$, giving $\Delta \log Y_{i,t} \approx$

$k_i + \epsilon_{i,t+1} - \epsilon_{i,t}$. Thus:

$$\mathbb{P}_{50}[\Delta \log Y_{i,t}] \approx k_i \quad \text{for } Y_{i,t} \leq 0.1 L_i$$

This provides a consistent estimator $\hat{k}_i = \mathbb{P}_{50}[\Delta \log(Y_{i,t}/L_i)]$ at early diffusion.

For time alignment, rearranging the early-diffusion approximation gives $k_i t_{0,i} = k_i t - \log(Y_{i,t}/L_i) + \epsilon_{i,t}$. Taking medians:

$$\hat{k}_i \hat{t}_{0,i} = \hat{k}_i \mathbb{P}_{50}[t] - \mathbb{P}_{50}[\log(Y_{i,t}/L_i)], \quad \text{so} \quad \hat{t}_{0,i} = \mathbb{P}_{50}[t] - \frac{\mathbb{P}_{50}[\log(Y_{i,t}/L_i)]}{\hat{k}_i}$$

Normalized time $t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$ aligns all technologies at their inflection points, removing rate and timing differences. Normalizing growth rates by $\hat{k}_i$ removes rate scaling, collapsing all curves onto a universal shape determined solely by β.

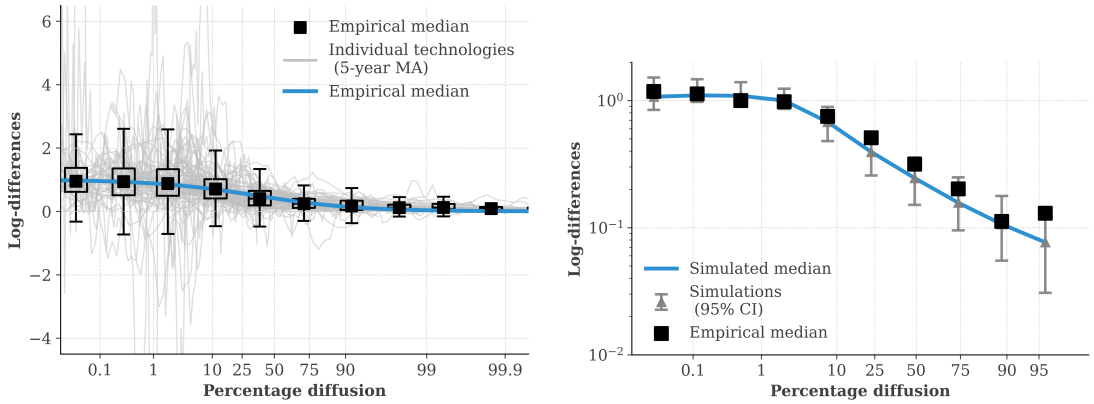
We use medians rather than OLS for robustness to outliers and deviations from exponential growth at very early diffusion. Comin et al. [91] report rapid initial growth that decays toward exponential behavior. Figure 4.19 confirms this pattern: $\sim 10\%$ of technologies exhibit non-exponential growth at < 0.5–1% diffusion before stabilizing. Median estimation is robust to these transient dynamics.

Algorithm 6: Replicating growth behavior with universal β

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Comparison of empirical vs. simulated growth behavior
Step 1: Normalize empirical data:
  foreach *technology* $i = 1$ *to* 120 **do**
    Calculate $\Delta \log Y_{i,t}$ using Eq. 4.47; define $L_i = \max_t Y_{i,t}$
    Estimate $\hat{k}_i = \mathbb{P}_{50}\left[\Delta \log(Y_{i,t}/L_i) : Y_{i,t} \leq 0.1 L_i\right]$
    Estimate $\hat{t}_{0,i}$ from median: $\hat{t}_{0,i} = \mathbb{P}_{50}[t] - \mathbb{P}_{50}[\log(Y_{i,t}/L_i)]/\hat{k}_i$
    Normalize: $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t}/\hat{k}_i$, $t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$
  end
Calculate median growth trajectory: $\mathbb{P}_{50}[\Delta \log^{(n)} Y_{i,t}]$
Step 2: Generate and normalize surrogate data:
  foreach *simulation* $s = 1$ *to* $M = 1{,}000$ **do**
    Generate $\{Y_i^{(s)}(t)\}_{i=1}^{120}$ using DGP (Eq. 4.17) with $\hat{\theta}_i$, $\beta = 2/3$, $\rho = 0.8$
    Apply Step 1 normalization to obtain $\{\Delta \log^{(n)} Y_i^{(s)}(t)\}_{i=1}^{120}$
    Calculate surrogate median: $\mathbb{P}_{50}[\Delta \log^{(n)} Y_i^{(s)}(t)]$
  end
Step 3: Compare distributions:
  Compare empirical median $\mathbb{P}_{50}[\Delta \log^{(n)} Y_{i,t}]$ to distribution of surrogate medians $\{\mathbb{P}_{50}[\Delta \log^{(n)} Y_i^{(s)}(t)]\}_s$

Algorithm 6 compares median growth behavior from empirical data to surrogate data generated with $\beta = \frac{2}{3}$. The algorithm outputs the distribution of median growth trajectories from $M = 1{,}000$ surrogate datasets for comparison to the empirical median growth trajectory.

Results. Figure 4.20 shows the representative growth trajectory (median time-binned log-differences of empirical data) compared to the Bertalanffy-Richards growth-rate profile with $\beta = \frac{2}{3}$ and the distribution of median normalized log-differences from 1,000 surrogate data simulations (as per Figure 2d in the main text). We also show the 5-year moving average of log-differences for all technologies. Figure 4.20a shows that growth behaviour is



(a) Collapse plot of log-differences (empirical data)

(b) Collapse plot of log-differences (simulated data)

Figure 4.20: Collapse plot of log-differences against diffusion level. **(a)** Collapse plot of log-differences of 120 technologies (empirical data). Grey lines are 5-year moving averages of log-differences for each technology. Box-and-whisker plot shows binned median and interquartile range across technologies. Blue line shows median. Black dashed line shows Bertalanffy-Richards curve with $\beta = \frac{2}{3}$. **(b)** Collapse plot of log-differences (log-scale) from 1,000 surrogate datasets generated with $\beta = \frac{2}{3}$. Vertical lines show the $2.5 - 97.5^{\text{th}}$ percentiles. Blue line (and triangles) shows median across simulations. Black dots show median from empirical data (from **(a)**).

volatile across technologies, even with 5-year moving averages. The median log-differences (blue line) closely follow the B-R curve with $\beta = \frac{2}{3}$, with small deviations at high diffusion levels. Figure 4.20b compares the empirical median to surrogate data, showing minimal differences.

Figure 4.20b shows that all empirical median log-differences lie within the $2.5 - 97.5^{\text{th}}$

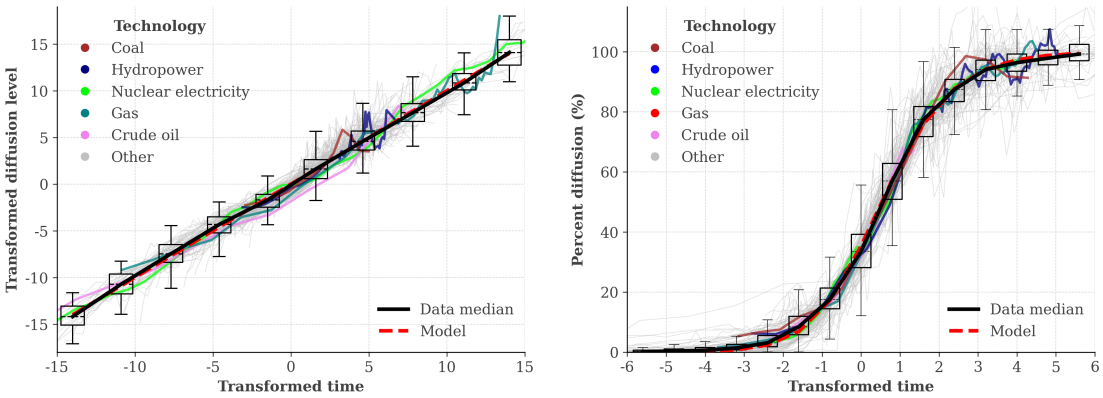
percentiles: any point outside would be marked as a red square, but none are present. Notably, after 10% adoption, empirical median boxes consistently lie above the simulated median triangles. In particular, although it still lies within the error bars, the last empirical median (rightmost black square) lies in the upper tail of the surrogate distribution. This pattern is consistent with the saturation-noise misspecification of Section 4.9: the surrogate band tightens toward zero at high d while empirical log-differences retain variability, biasing their cross-technology median into the surrogate's upper tail. Section 4.16 confirms the same pattern in forecast errors. We use the full-series estimate of $\hat{\sigma}_i$ here, as is appropriate for diagnostic replication tests (Section 4.9).

4.14.2.1 Collapse plot of technology adoption curves

We collapse technology time-series onto a single plot with $L = 1$, $k = 1$, $t_0 = 0$ using Eqs. 4.19, 4.20, and 4.21. We then create collapse plots for both levels of adoption, and their GFP transformation (Figures 2a and 2b in main text).

We estimate parameters as follows. For β, we assume $\beta = \frac{2}{3}$ in both cases. For the collapse of the levels, we use NLLS estimates of $\hat{L}_i$, $\hat{k}_i$, and $\hat{t}_{0,i}$. For the GFP collapse, we first estimate $\hat{L}_i$ by NLLS, then apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ (Eq. 4.16) and regress $f(Y_{i,t}; \hat{L}_i, \beta)$ on time to estimate $\hat{k}_i$, $\hat{t}_{0,i}$, and $\hat{\sigma}_i$. We then apply normalization per Eqs. 4.19, 4.20, and 4.21. For the GFP collapse plot, we additionally normalize time by dividing by $\hat{\sigma}_i$. That is, the GFP-transformed data satisfies $y_{i,t} = k_i(t - t_{0,i}) + \sigma_i \eta_{i,t}$, and after setting $k = 1$ and $t_0 = 0$ (via the standard normalization), the residual scale σ_i still varies across technologies. Dividing time by $\hat{\sigma}_i$ removes this last technology-specific quantity, ensuring complete collapse.

Figure 4.21 shows collapse plots for GFP transformation and levels of adoption, with selected technologies highlighted. Figure 4.21 shows that the data collapses onto a single plot. The median of the ensemble (black line) tracks the model (red dashed line) closely,



(a) Collapse plot of the data under the GFP transformation

(b) Collapse plot of the data in levels

Figure 4.21: Collapse and binned box plots for Bertalanffy-Richards ($\beta = 2/3$) curve fits. Black line: median of ensemble. Red dashed line: model. **(a)** Data under the generalized Fisher-Pry transformation with selected technologies highlighted. **(b)** Untransformed data with energy technologies highlighted.

and technologies across sectors (highlighted) collapse onto similar diffusion paths.

Test to replicate observed technology S-curves. We test whether the DGP (Eq. 4.17) approximates the stochastic variation observed in the data, comparing the normalized empirical ensemble to the normalized empirical surrogate ensemble (Section 4.9) via Algorithm 7.

Algorithm 7: Testing homoskedasticity and normality of noise

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Assessment of homoskedasticity and normality of residuals
Step 1: Calculate normalized residuals from empirical data:
  foreach *technology* $i = 1$ *to* 120 **do**
    Fit Bertalanffy-Richards curve via NLLS to obtain $\hat{L}_i$ for each technology
    Apply GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ to estimate remaining $\hat{\theta}_i$
    Normalize data per Eqs. 4.19, 4.20, 4.21
    Calculate standard deviation of residuals $\text{std}(\hat{\epsilon}_i) = \sqrt{\frac{1}{T_i - 1} \sum_t (\hat{\epsilon}_{i,t} - \bar{\epsilon}_i)^2}$
    Normalize data: $f_i^{(n)}(t) = f(Y_{i,t}; \hat{L}_i, \beta)/\text{std}(\hat{\epsilon}_i)$
  end
  Define set of normalized data: $f^{(n)}(t) = \{f_i^{(n)}(t)\}_{i=1}^{120}$
Step 2: Generate surrogate data and calculate residuals:
  foreach *simulation* $s = 1$ *to* $M = 1{,}000$ **do**
    foreach *technology* $i = 1$ *to* 120 **do**
      Generate $Y_i^{(s)}(t)$ using DGP (Eq. 4.17)
      with $\hat{\theta}_i$, $\beta = \frac{2}{3}$, and $\rho = 0.8$
    end
    Calculate normalized data for surrogate data as in Step 1
    Define the set of normalized data: $f_s^{(n)}(t) = \{f_i^{(n)(s)}(t)\}_{i=1}^{120}$
  end
  Define set of simulations: $\{f_s^{(n)}(t)\}_{s=1}^{M}$
Step 3: Compare distributions:
  Compare ensemble normalized data (Step 1)
  to distribution of normalized ensemble from surrogate data (Step 2)
  across all normalized time units

Under the DGP, the distributions of normalised data from real and surrogate data should match across all diffusion levels. We calculate residuals as

$$\hat{\epsilon}_{i,t} = f(Y_{i,t}; \hat{L}_i, \beta) - \hat{k}_i(t - \hat{t}_{0,i}), \tag{4.49}$$

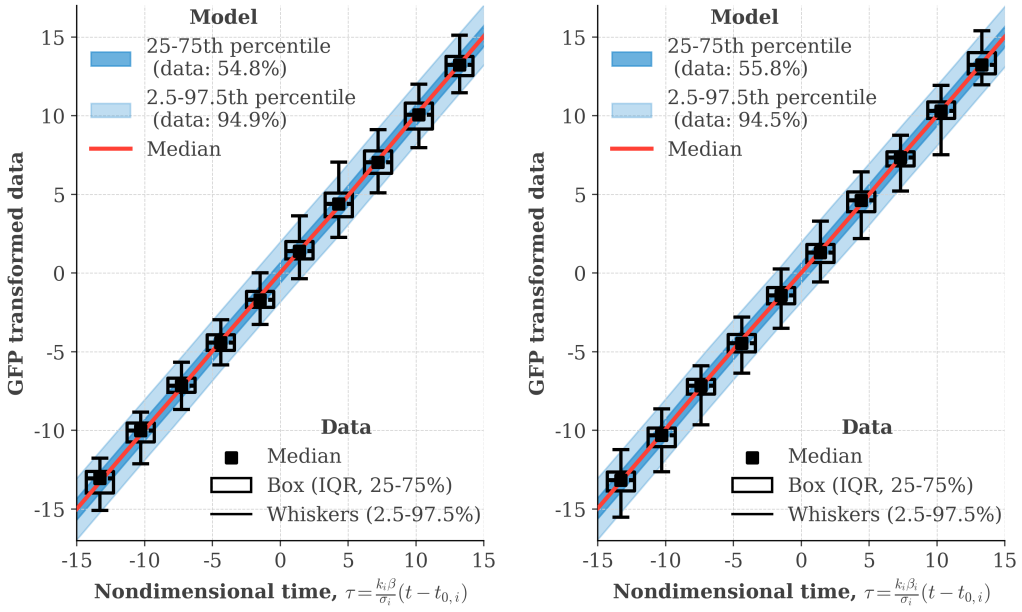
and normalised residuals as

$$\hat{\epsilon}_{i,t}^{(n)} = \frac{\hat{\epsilon}_{i,t}}{\text{std}(\hat{\epsilon}_i)}, \tag{4.50}$$

where $\text{std}(\hat{\epsilon}_i)$ is the standard deviation of residuals for technology i under the GFP

transform.

Replicating ensemble behavior. Figure 4.22 shows box-plots of $f^{(n)}(t)$ (normalized GFP-transformed data from Step 1 of Algorithm 7) with surrogate data $\{f_s^{(n)}(t)\}_{s=1}^{M}$ shaded. This is Figure 2c in main text. Further, we complement this with a similar collapse using technology-specific β_i. That is, we apply Algorithm 7, but use β_i instead of $\beta = 2/3$ in step 1. We keep $\beta = 2/3$ in step 2 to generate surrogate data (i.e., estimate parameters of empirical data using $\beta = \frac{2}{3}$ for surrogate generation), since we want to test whether the DGP with a universal β can approximate the stochastic variation observed in the data. We run 1000 simulations for 120 surrogate technology time-series and plot the distribution of normalized data from surrogate data against the normalized data from empirical data. The median of the ensemble (black dots) is highly consistent with the



(a) Collapse assuming $\beta = \frac{2}{3}$ **(b)** Collapse using technology-specific β_i

Figure 4.22: Collapse plot of normalized GFP-transformed data $f^{(n)}$ against normalized time, overlaid with 1,000 ensembles of 120 technology surrogate data from DGP (Algorithm 7). Box-plots: distribution of normalized data from empirical data ensemble. Whiskers are $2.5 - 97.5^{\text{th}}$ percentile. Shaded regions: $25 - 75^{\text{th}}$ and $2.5 - 97.5^{\text{th}}$ percentile ranges of surrogate data. Proportions shown indicate empirical data within each percentile range. Black dots: median of ensemble in each box-plot. Red line: median of surrogate data. **(a)** Collapse assuming $\beta = \frac{2}{3}$. **(b)** Collapse using technology-specific β_i.

median of surrogate data (red line) in both Figure 4.22a and Figure 4.22b. In both cases, $\approx$ 95% of the data lie within the 2.5-97.5th percentile range of surrogate data (94.9% when assuming $\beta = \frac{2}{3}$, 94.5% when using technology-specific β_i). However, the proportions of empirical data outside the 25–75th percentile range are higher than expected. This suggests that the data exhibit heavier tails than the DGP, which assumes normality after the generalized Fisher-Pry transformation. These deviations (heavier tails, leptokurtic distribution) represent limitations of the normality assumption.

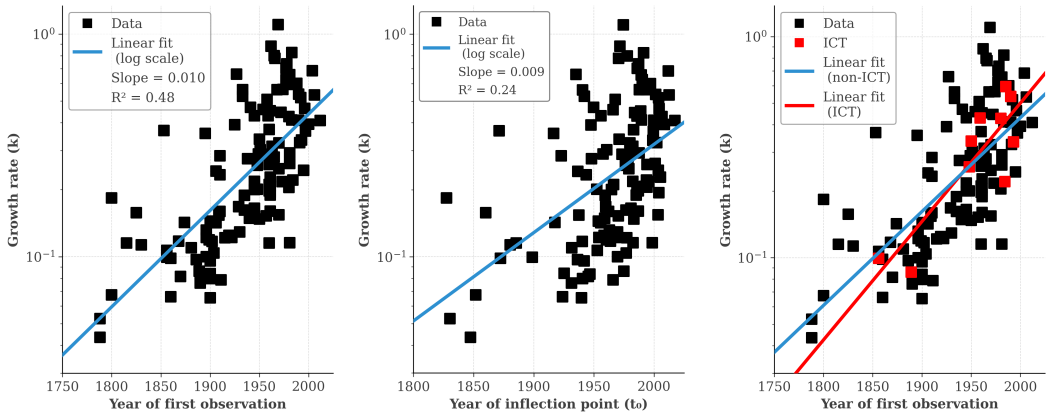
The results are similar when using technology-specific β_i for collapse or assuming $\beta = \frac{2}{3}$. Note that while Figure 4.22b uses technology-specific β_i for the empirical collapse, we still generate surrogate data with $\beta = 2/3$ to test whether a universal β can approximate the observed stochastic variation. The similarity in collapse indicates that it can. Despite the heavy-tailed deviations from the normality assumption, the replication shown in Figure 4.22 indicates that the DGP is a reasonable approximation for forecasting, and we use it in subsequent sections.

4.14.3 Other regularities in technology diffusion from literature

Here we test three additional regularities reported in previous studies: (1) Are technologies diffusing faster over time [16]? (2) Are ICT technologies diffusing faster than other technologies [204]? (3) Do technologies with higher growth rates have more uncertainty?

Are technologies diffusing faster over time? Grubler [16] reports technologies diffuse faster over time. Figure 4.23 shows[26] estimated $\hat{k}_i$ against year of first observation (Figure 2f in main text) and against $\hat{t}_{0,i}$. Red squares: ICT technologies. Colored lines: linear regression lines. $\log \hat{k}_i$ and the year of first observation are positively correlated

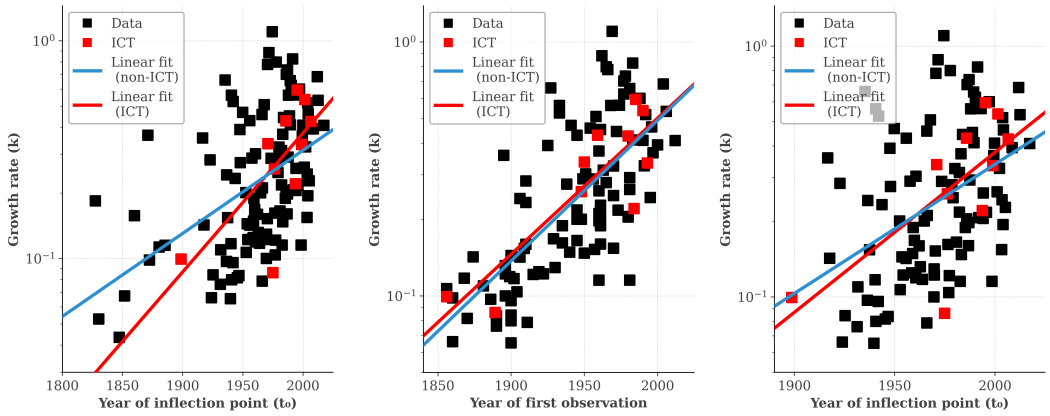
[26] Monasteries (1100s) excluded.



(a) $\log k_i$ vs. year of first observation

(b) $\log k_i$ vs. $t_{0,i}$

(c) k_i vs. year of first observation (ICT vs non-ICT)



(d) k_i vs. $t_{0,i}$ (ICT vs non-ICT)

(e) k_i vs. year of first observation (ICT vs non-ICT, excl. early diffusers)

(f) k_i vs. $t_{0,i}$ (ICT vs non-ICT, excl. early diffusers)

Figure 4.23: Estimated growth rate $\hat{k}_i$ against time (log-scale). Red dots: ICT technologies. Colored lines: regression lines. **(a)** $\hat{k}_i$ vs. year of first observation. **(b)** $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$. Blue line: linear regression line with positive slope. **(c)** $\hat{k}_i$ vs. year of first observation, ICT (red) vs non-ICT (blue). **(d)** $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$, ICT (red) vs non-ICT (blue). **(e)** $\hat{k}_i$ vs. year of first observation, ICT (red) vs non-ICT (blue), excluding technologies before 1900. **(f)** $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$, ICT (red) vs non-ICT (blue), excluding technologies before 1900.

($R^2 = 0.48$, slope $= 0.01$; Figure 4.23a). The correlation between $\log \hat{k}_i$ and $\hat{t}_{0,i}$ is positive ($R^2 = 0.24$, slope $= 0.009$; Figure 4.23b). Newer technologies diffuse faster in this dataset, consistent with [16]. This result may reflect selection bias: the dataset includes only technologies that diffused sufficiently. Slow-diffusing recent technologies may not yet have reached required thresholds, skewing the sample toward faster-growing technologies in recent years.

Are ICT technologies diffusing faster? Tankwa et al. [204] report that ICT growth rates exceed non-ICT when analyzing country-level data, attributing this to modularity, consistent with [211]. Figures 4.23c–4.23f compare ICT and non-ICT growth rates. Table 4.8 shows statistical tests. Comparisons: ICT vs. all non-ICT, and ICT vs. non-ICT in similar time-period (after first ICT technology observation). Table 4.8 indicates

Metric	Non-ICT (all)	p-value (vs ICT)	ICT	p-value (vs ICT)	Non-ICT (same time-period)
Median k_i	0.21	–	0.34	–	0.22
Mann-Whitney U test	–	0.29	–	0.40	–
K-S test	–	0.21	–	0.32	–
Slope of OLS ($\log k_i$ *vs. year of first observation*)	0.010	–	0.012	–	0.012
Slope of OLS ($\log k_i$ *vs.* $t_{0,i}$)	0.009	–	0.015	–	0.009

Table 4.8: Statistical tests comparing growth rates of ICT vs. non-ICT technologies, generated by `make_tab_S21.py` in the replication package. Technologies first observed before 1700 (one technology, monasteries) are excluded; the same time-period subset comprises non-ICT technologies first observed in 1850 or later (the earliest ICT technology, telegraphs, appears in 1856). ICT median $k_i = 0.34$, non-ICT median $k_i = 0.21$, non-ICT (same time-period) median $k_i = 0.22$. Mann-Whitney U test p-values: 0.29 (ICT vs. all non-ICT), 0.40 (ICT vs. non-ICT same period). K-S test p-values: 0.21 (ICT vs. all non-ICT), 0.32 (ICT vs. non-ICT same period). OLS slopes ($\log k_i$ vs. year of first observation): 0.010 (non-ICT all), 0.012 (ICT), 0.012 (non-ICT same period). OLS slopes ($\log k_i$ vs. $t_{0,i}$): 0.009 (non-ICT all), 0.015 (ICT), 0.009 (non-ICT same period).

that ICT technologies have higher median growth rates than non-ICT technologies (both when compared to all non-ICT and to the same time-period subset). The difference is not statistically significant (Mann-Whitney U p-values: 0.29, 0.40). The small ICT sample size ($n = 10$) may limit statistical power.

Table 4.8 and Figures 4.23c, 4.23d: ICT growth rates increase faster over time than non-ICT growth rates. The slope difference is smaller when comparing $\log k_i$ vs. year of first observation (0.012 vs. 0.010), but larger when comparing $\log k_i$ vs. $t_{0,i}$ (0.015 vs. 0.009).

Excluding technologies that diffused before 1900, the slope difference decreases (Fig-

ures 4.23e, 4.23f). For k_i vs. year of first observation (Figure 4.23e), non-ICT growth rates increase marginally faster, though not significantly. The difference observed when including all technologies may be driven by varied growth patterns among early diffusers. When controlling for time period, ICT and non-ICT growth rates may increase at similar rates.

Notably, these results are subject to dataset limitations. The dataset only includes technologies that have surpassed certain diffusion thresholds, which may bias the sample toward faster diffusing technologies, especially in time-periods closer to the present day. Thus, we should interpret conclusions regarding ICT vs. non-ICT growth rates with caution.

Do technologies with higher growth rates have more uncertainty? Farmer and Lafond [30] report that technologies with faster cost improvements have higher volatility. Here we test a similar phenomenon, but for technology diffusion. Figure 4.24 shows estimated $\hat{\sigma}_i$ against estimated $\hat{k}_i$.

Figure 4.24 shows that the correlation between $\hat{k}_i$ and $\hat{\sigma}_i$ is only slightly positive, with $R^2 = 0.06$ and a slope of 0.33. The data are noisy, suggesting the relationship may be spurious. This dataset shows a weak positive relationship between growth rate and uncertainty, weaker than the learning rate-volatility relationship reported in [30].

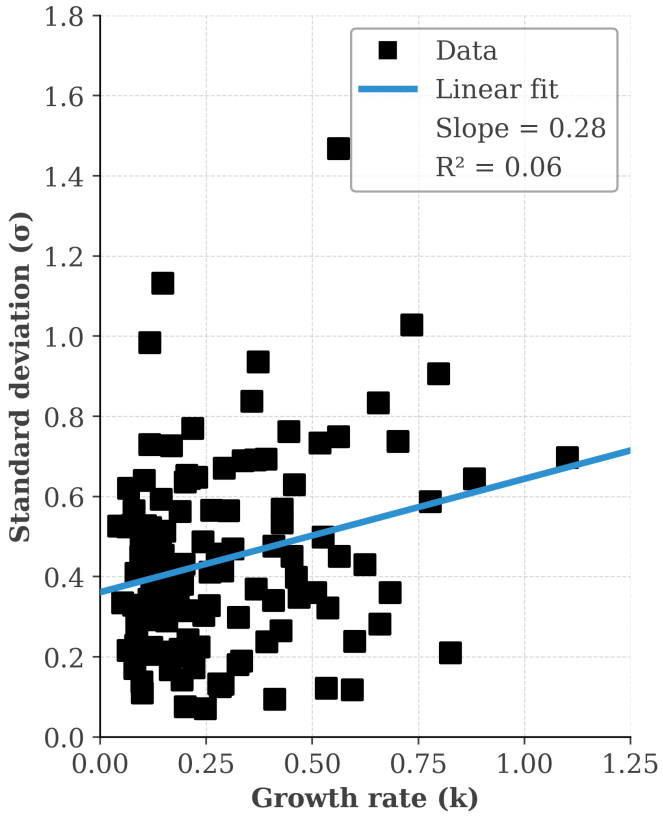


Figure 4.24: Estimated growth rate $\hat{k}_i$ against estimated noise parameter $\hat{\sigma}_i$. Correlation: $R^2 = 0.06$, slope $= 0.33$.

4.15 Comparison of S-curve models

The literature is unclear on S-curve model selection [41, 69, 84, 86]. We show that Gompertz, logistic, and Bertalanffy-Richards models produce reasonable in-sample fits ($R^2 \approx 0.93$) despite assuming different S-curve functional forms. In this section, we analyze which model best captures empirical growth dynamics. We then show that ensemble averaging across technologies with heterogeneous parameters (i.e. k_i and $t_{0,i}$ drawn from a distribution) creates apparent Gompertz-like decay even when individual series follow logistic or B-R dynamics. This helps explain why empirical studies sometimes favor Gompertz and sometimes logistic, while B-R is often not considered (Section 4.6).

We perform three analyses: (1) Compare R^2 values to confirm similar in-sample fit quality. (2) Compare the growth collapse plot and show that ensemble averaging across time-series can make the Gompertz curve appear accurate even when individual technologies

exhibit logistic- or B-R-like exponential growth at early stages. (3) All three models fail to varying degrees to describe early diffusion dynamics, consistent with Comin et al. [91]. We conclude that Gompertz exhibits systematic early-stage misspecification. Ensemble averaging explains why pooled data may indicate a good fit, even though the underlying growth is actually better explained by another model. The B-R model (with $\beta = 2/3$) is the least misspecified, but the logistic model ($\beta = 1$) is also reasonable. The backtesting comparison in Section 4.16.1 provides the out-of-sample evidence underpinning the choice of B-R as the default forecasting model.

4.15.1 Theoretical differences between Gompertz and logistic curves

We compare growth behaviour under Gompertz and logistic models. Both provide reasonable fits despite prescribing fundamentally different early-stage dynamics: Gompertz assumes exponentially decaying growth, while the logistic assumes approximately constant exponential growth at early stages.

As derived in Section 4.6, the Gompertz curve features exponentially decaying growth at all diffusion levels:

$$\frac{d \log Y}{dt} = k \exp(-k(t - t_0))$$

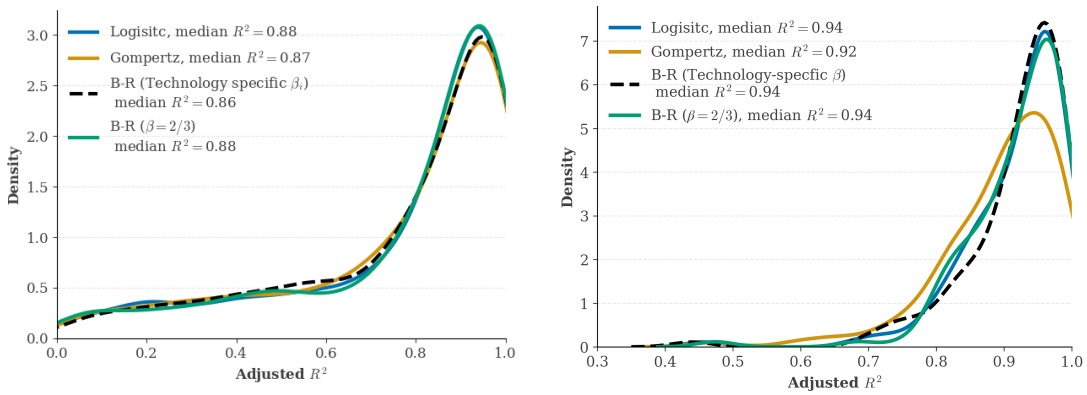
In contrast, the logistic curve exhibits approximately constant exponential growth at early diffusion, then exponentially decaying growth at late diffusion:

$$\begin{aligned} \frac{d \log Y}{dt} &= \frac{k}{1 + \exp(k(t - t_0))} \\ &\approx k \quad \text{for } t \ll t_0 \quad \text{(early diffusion: exponential growth)} \\ &\approx k \exp(-k(t - t_0)) \quad \text{for } t \gg t_0 \quad \text{(late diffusion: Gompertz-like decay)} \end{aligned} \tag{4.51}$$

Under the Gompertz model, growth rates decay exponentially throughout diffusion, implying the logarithm of log-differences (log-log-differences) should be linear in time. Under the logistic (and Bertalanffy-Richards) model, growth rates remain approximately constant at early stages, implying log-differences should be roughly constant initially before decaying later.

4.15.1.1 Goodness of fit

Figure 4.25 compares adjusted R^2 values for Gompertz, logistic, and B-R curves ($\hat{\beta}_i$ and universal $\beta = 2/3$) across both datasets. All four model variants — Gompertz, logistic,



(a) Adjusted R^2 values for raw data

(b) Adjusted R^2 values for clean data

Figure 4.25: Comparison of R^2 values across Gompertz, logistic, and Bertalanffy-Richards models. **(a)** Comparison for raw data and **(b)** comparison for final dataset. All curves yield similar R^2 distributions with similar median R^2 values.

B-R with fixed $\beta = 2/3$, and B-R with estimated $\hat{\beta}_i$ — provide similar ad-hoc fits on both the raw (Figure 4.25a) and clean data (Figure 4.25b), with Gompertz having the lowest median adjusted R^2 of 0.91.

4.15.2 Ensemble growth behavior under different S-curve assumptions

We compare how well the Gompertz and logistic models capture ensemble growth behavior using collapse plots of log-differences (cf. B-R collapse in Figure 4.20b). Algorithm 8

details Gompertz normalization; see Section 4.14 for logistic and B-R curves.

Mathematical derivation of normalization factors. For the Gompertz curve with growth rate $\frac{d \log Y}{dt} = k \exp(-k(t - t_0))$, integrating over discrete time step $h = 1$ yields:

$$\Delta \log Y_{i,t} = [1 - e^{-k_i}] \exp(-k_i(t - t_{0,i})). \tag{4.52}$$

Taking logarithms yields

$$\log \Delta \log Y_{i,t} = \log(1 - e^{-k_i}) - k_i(t - t_{0,i}).$$

This expression is linear in time throughout the entire diffusion curve, allowing parameter estimation on the full sample. We regress $\log \Delta \log Y_{i,t}$ on time t using OLS to obtain slope $m_i = -k_i$ and intercept $b_i = k_i t_{0,i} + \log(1 - e^{-k_i})$, giving:

$$k_i = -m_i, \quad t_{0,i} = \frac{b_i - \log(1 - e^{-k_i})}{k_i}.$$

The normalization factor $(1 - e^{-\hat{k}_i})$ arises from the discrete-time integration. Setting time $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$ and then $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t}/(1 - e^{-\hat{k}_i})$ collapses all technologies onto the universal curve $\log(\Delta \log^{(n)} Y_{i,t}) = -\tilde{t}_i$.

Algorithm 8: Gompertz collapse plot construction

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Gompertz collapse plot parameters $\{k_i, t_{0,i}\}$
Step 1: Calculate log-log-differences:
foreach *technology* $i = 1$ *to* 120 **do**
  Calculate log-log-differences:
  $\log \Delta \log Y_{i,t} = \log\left(\frac{\log Y_i(t+h) - \log Y_{i,t}}{h}\right)$
end
Step 2: Estimate parameters via OLS:
foreach *technology* i **do**
  Regress $\log \Delta \log Y_{i,t}$ on time t using OLS: $\log \Delta \log Y_{i,t} = m_i t + b_i$
  Extract parameters:
  $\hat{k}_i = -m_i$
  $\hat{t}_{0,i} = \frac{b_i - \log(1 - e^{-\hat{k}_i})}{\hat{k}_i}$
end
Step 3: Normalize and construct collapse plot:
foreach *technology* i **do**
  Normalize time: $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$
  Normalize log-differences: $\Delta \log^{(n)} Y_{i,t} = \frac{\Delta \log Y_{i,t}}{1 - e^{-\hat{k}_i}}$
end
Plot $\Delta \log^{(n)} Y_{i,t}$ vs $\tilde{t}_i$ on log scale

4.15.2.1 Replicating ensemble growth behavior via simulation

We construct empirical surrogate ensembles (Section 4.9) under each of the three candidate DGPs (B-R with $\beta = 2/3$, Gompertz, logistic) and compare median normalized growth trajectories to the empirical pool. The procedure follows the standard fit-then-resample protocol; the only deviation is the substitution of the candidate DGP at the generation step.

Parameter estimation procedures and normalization factors. For the B-R curve, the GFP-transform at early diffusion ($t \ll t_{0,i}$) gives:

$$-\frac{1}{\beta} \log \left[\left(\frac{L_i}{Y_{i,t}} \right)^{\beta} - 1 \right] \approx k_i (t - t_{0,i})$$

At early diffusion, $\frac{d \log Y}{dt} \approx k$, so:

$$\Delta \log Y_{i,t} = \log Y_{i,t+1} - \log Y_{i,t} \approx k_i$$

Normalizing by k_i gives $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t}/k_i \approx 1$ at early diffusion.

We estimate $\hat{L}_i$ via NLLS fitting of the B-R model, then regress the GFP-transformed variable on time t using OLS on the first 10% of data ($Y_{i,t} < \hat{L}_i/10$) to obtain slope m_i and intercept b_i:

$$\hat{k}_i = m_i, \quad \hat{t}_{0,i} = -b_i/m_i$$

For the **Gompertz model**, we apply OLS to the full sample, regressing $\log \Delta \log Y_{i,t}$ on time t. As derived above, this yields $\hat{k}_i = -m_i$ and $\hat{t}_{0,i} = (b_i - \log(1 - e^{-\hat{k}_i}))/\hat{k}_i$. The normalization factor $(1 - e^{-\hat{k}_i})$ arises from integrating the exponential decay. Unlike logistic and B-R curves where $\Delta \log Y \approx k$ only at early diffusion, the Gompertz curve has $\log \Delta \log Y$ linear in time, allowing full-sample estimation.

Thus, normalized log-differences are defined as

$$\Delta \log Y_{i,t}/\hat{k}_i \quad \text{(logistic and B-R)}, \qquad \Delta \log Y_{i,t}/(1 - e^{-\hat{k}_i}) \quad \text{(Gompertz)}$$

The normalized time is $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$ for all models.

Algorithm 9: Surrogate data generation under three candidate DGPs (Gompertz, logistic, and B-R with $\beta = 2/3$) for replicating empirical ensemble growth behaviour. Per-technology parameters are estimated on the early-data window—first 10% for logistic and B-R, full sample for Gompertz (which has analytic OLS-on-log-log form). $M = 1000$ surrogate ensembles are drawn per DGP and compared to the empirical pool via the cross-technology median growth trajectory.

Input: Real technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: $M = 1000$ surrogate ensembles under both DGPs
Step 1: Estimate parameters from empirical data:
  foreach *technology* $i = 1$ *to* 120 **do**
    Estimate $\hat{L}_i$ via NLLS
    Regress $f(\hat{L}_i)$ on time t to estimate $\hat{k}_i, \hat{t}_{0,i}$
    Estimate noise: $\sigma_i = \sigma_{\epsilon,i}/\sqrt{1+\rho^2}$ with $\rho = 0.8$,
    where $\sigma_{\epsilon,i}$ is OLS residual standard deviation
  end
Step 2: Generate surrogate data:
  foreach *simulation* $s = 1$ *to* $M = 1000$ **do**
    foreach *technology* $i = 1$ *to* 120 **do**
      Generate time-series under each DGP (B-R, Gompertz, logistic) with parameters $\{L_i, k_i, t_{0,i}, \sigma_i, \rho = 0.8\}$, using $\beta = 2/3$ for B-R and $\beta = 1$ for logistic
    end
  end
Step 3: Calculate ensemble median growth rates:
  foreach *simulation* s **do**
    Calculate log-differences
    Logistic and B-R (technology-invariant β): Estimate parameters on first 10% using $\hat{L}_i = \max_t Y_{i,t}$ and normalize
    Gompertz: Estimate parameters on log-log-differences and normalize
    Calculate median across 120 technologies at each normalized time
  end
  Calculate median of medians across $M = 1000$ simulations

Figure 4.26 shows log-differences collapse plots under Gompertz, logistic, and B-R ($\beta \in \{1/2, 2/3\}$) curves. It shows that the Gompertz DGP captures ensemble growth behavior when fitting the full sample, with empirical median marginally below the 2.5$^{\text{th}}$ percentile at high diffusion levels (red squares). The logistic DGP captures early diffusion growth when fitting the first 10% but underestimates growth in the last 5–10% (red square). Otherwise it is comparable to the B-R $\beta = 2/3$ model, but assuming a slightly faster decay in growth rate at later diffusion. The B-R curve with $\beta = 0.5$ performs comparably to $\beta = 2/3$, but slightly better captures growth at later 90-95% diffusion (the black square is further within the 95% CI of simulation). The logistic and two B-R models tend to

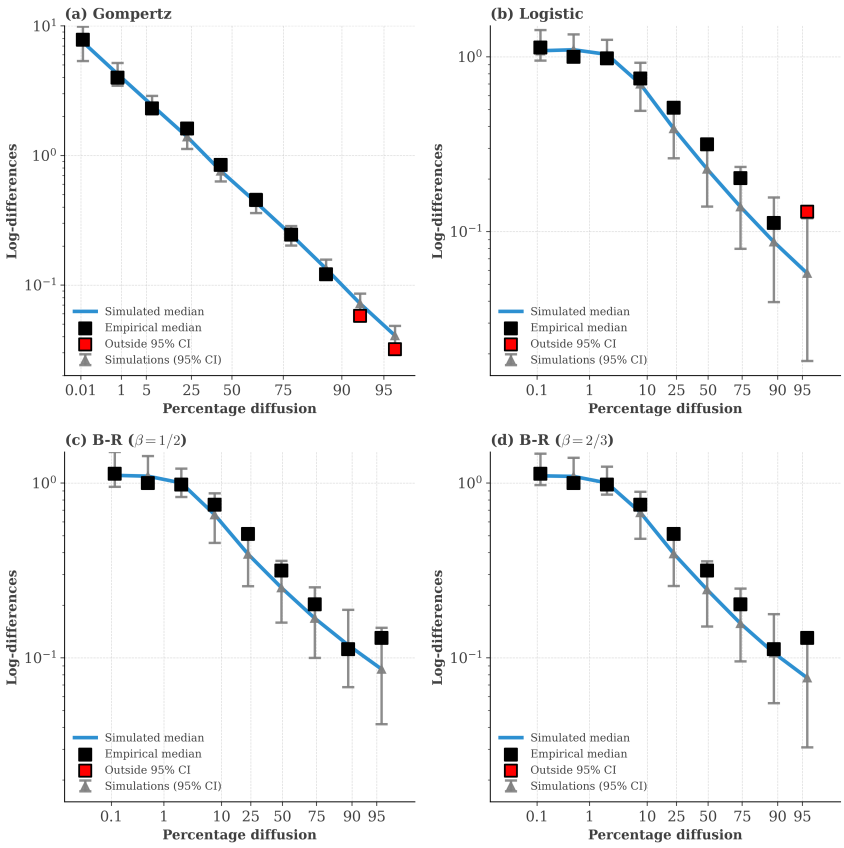


Figure 4.26: Collapse plots of log-differences under **(a)** Gompertz (parameter estimates on full sample), **(b)** logistic (parameter estimates on 10% of the data), **(c)** B-R ($\beta = 0.5$, parameter estimates on 10% of the data), and **(d)** B-R ($\beta = 2/3$, parameter estimates on 10% of the data). Black squares: median observed growth rates. Blue lines (and gray triangles): simulation ensemble medians. Gray whiskers: $2.5 - 97.5^{\text{th}}$ percentiles. Red squares: empirical median outside simulated $2.5 - 97.5^{\text{th}}$ percentiles.

underestimate growth rates, observed by the empirical medians (black squares) often being above the surrogate medians (blue lines), from 15% diffusion onward. The same direction of bias appears across these three GFP-additive-Gaussian-noise DGPs despite their different functional forms, which points to the saturation-noise misspecification of Section 4.9 (shared across them) rather than to functional form; the Gompertz model shows the opposite-direction deviation, reflecting its exponentially decaying growth law rather than the noise specification. As in Section 4.14, $\hat{\sigma}_i$ is estimated on the full series for this diagnostic.

These results illustrate the difficulty of identifying the true DGP from ensemble growth behaviour alone. In-sample fit quality (adjusted R^2) is high for all models, indicating that

each captures the overall growth pattern. The Gompertz DGP fits the full sample but slightly overestimates the median growth at high diffusion; the logistic and B-R DGPs fit the first 10% well but tend to underestimate growth thereafter. Ensemble growth behaviour alone therefore does not conclusively identify the true DGP.

4.15.3 Ensemble averaging effects

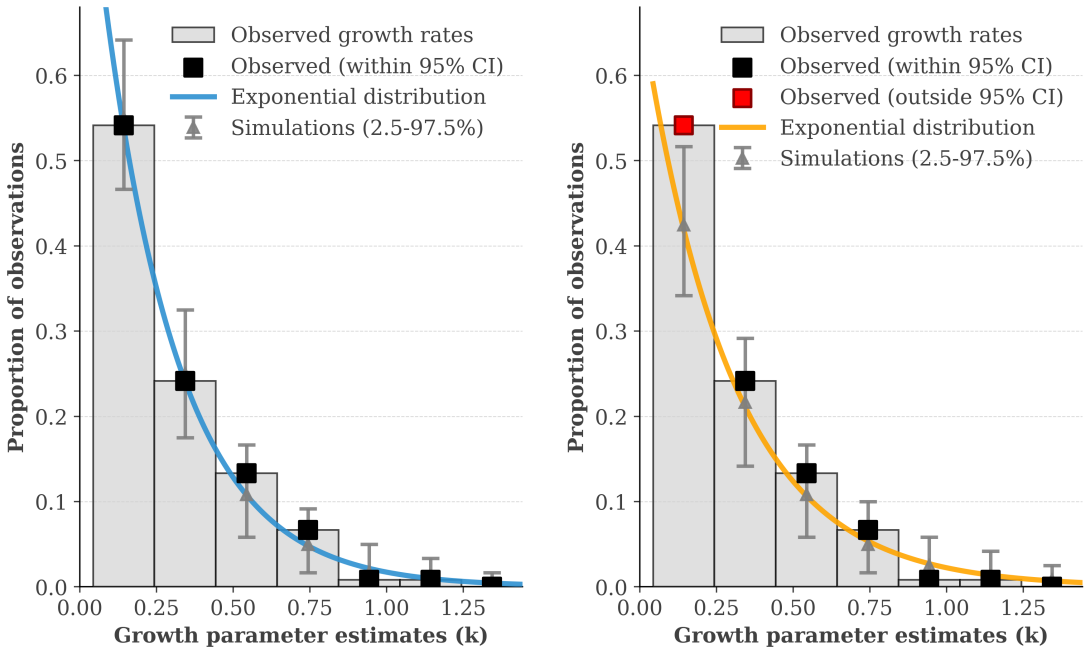
The conclusion that Gompertz captures ensemble growth may be misleading. Heterogeneous growth rates k_i [8, 40, 292], combined with pooling, could create apparent exponential decay even when individual technologies grow exponentially, which is called an *ensemble averaging artifact*.

Heterogeneous growth rates. We test whether k_i follows a (shifted) exponential distribution (Algorithm 10). A shifted exponential is defined as $k_i = \xi + \text{Exp}(s)$, where ξ is the location shift and s is the rate parameter.

Algorithm 10: Testing exponential distribution of growth rates

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Test if growth rates follow exponential distribution
B-R curve parameter estimates
  foreach *technology* $i = 1$ *to* 120 **do**
    Fit Bertalanffy-Richards curve via NLLS with $\beta = \frac{2}{3}$ to obtain $\hat{L}_i$
    Apply GFP transformation to $Y_{i,t}$ to get $y_i(t) = f(Y_{i,t}; \hat{L}_i, \beta)$
    Estimate $\hat{k}_i$ by regressing $y_i(t)$ on time t using OLS
  end
  Fit shifted exponential distribution to $\{\hat{k}_i\}_{i=1}^{120}$
  Draw 10,000 samples from fitted exponential distribution
  Compare empirical distribution to ensemble distributions

Figure 4.27 compares shifted versus non-shifted exponential distributions fitted to empirical growth rates $\{\hat{k}_i\}_{i=1}^{120}$. The left panel shows a shifted exponential distribution, $k_i = \xi + \text{Exp}(s)$ (as per Figure 2e in main text), where ξ is the location shift and s is the rate parameter, fitted via maximum likelihood. The shifted exponential is theoretically justified because S-curves cannot have zero growth rates (i.e., $k_i > 0$ by construction). The location shift ξ captures the minimum growth rate. The right panel shows a non-shifted



(a) Shifted exponential **(b)** Non-shifted exponential (independent)

Figure 4.27: Comparison of **(a)** the shifted exponential distribution, $k_i = \xi + \text{Exp}(s)$, and **(b)** the non-shifted exponential distribution, $k_i = \text{Exp}(s')$, both fitted to empirical growth rates. Black squares: observed proportions. Red squares: observed proportions outside the surrogate distribution. Solid curves: theoretical distributions. Triangles with error bars: median and 95% confidence intervals from 10,000 Monte Carlo simulations.

exponential distribution, $k_i = \text{Exp}(s')$, with independently fitted rate parameter s'. The shifted exponential shows close visual agreement with the empirical data, with all observed proportions falling within the 95% confidence intervals from 10,000 Monte Carlo simulations. The non-shifted exponential underestimates the lowest growth-rate bin, whose observed proportion falls outside the interval (red square).

Table 4.9 presents comprehensive statistical tests comparing three exponential models. The shifted exponential generally outperforms both non-shifted alternatives across most statistical tests (KS test, log-likelihood, AIC, BIC). The sole exception is the chi-square test, where the non-shifted model using the rate parameter from the shifted fit (nested model) exhibits the lowest chi-square statistic ($p = 0.12$). Note that S-curve time-series must have $k_i > 0$ by construction. The observed minimum growth rate $\min_i \hat{k}_i = 0.04$ confirms this, and closely matches the fitted location shift $\xi = 0.04$. We adopt the shifted

Table 4.9: Comparison of three exponential models for the growth-rate distribution. Figure 4.27 plots the shifted exponential and the non-shifted (independent) models; the non-shifted (same rate) is a nested model included here only for the statistical comparison.

Metric	**Shifted exponential**	**Non-shifted (same rate)**	**Non-shifted (independent)**
Location shift c	0.04	0.00	0.00
Rate λ	4.02	4.02	3.42
Goodness-of-fit tests			
KS statistic	0.09†	0.22***	0.19***
KS p-value	0.269	0.000	0.000
χ^2 statistic	43.71***	8.80†	16.29**
χ^2 p-value	0.000	0.117	0.006
Model comparison criteria			
Log-likelihood	47.00†	26.06	27.69
AIC	-89.99†	-50.11	-53.39
BIC	-84.42†	-47.32	-50.60

Note: † indicates best value. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Lower KS/chi-square statistics and lower AIC/BIC indicate better fit. Shifted exponential: $k_i = c + \text{Exp}(\lambda)$ ($c = 0.04$ matches the observed minimum growth rate $\min_i \hat{k}_i = 0.043$). Non-shifted (same rate): nested model with $c = 0$, same λ as shifted. Non-shifted (independent): $c = 0$, λ fitted independently.

exponential distribution $k_i = \xi + \text{Exp}(s)$ for subsequent analysis.

Ensemble averaging artifact. We use the approximate distribution of the empirical growth rates (shifted exponential) to demonstrate how ensemble averaging across technologies with heterogeneous growth parameters can create apparent Gompertz-like decay patterns even when individual technologies follow logistic or Bertalanffy-Richards dynamics with early-stage exponential growth. That is, pooling the log of log-differences across technologies with heterogeneous k_i and $t_{0,i}$ can distort ensemble growth behavior.

Mechanism. When pooling log-log-differences at common calendar times t, two factors create systematic bias:

- **Technologies are at different diffusion stages:** Pooling poses the risk of mixing different diffusion stages, which can bias the ensemble average.
- **Measurement noise adds bias (Jensen's inequality):** Taking logarithms of noisy growth rates systematically underestimates the true average because the logarithm is a concave function—technologies with high noise relative to their growth rate contribute disproportionately to downward bias [372].

Pooling $\log[\Delta \log Y]$ across technologies at common time coordinates creates systematic bias in the ensemble average.

Simulation demonstration. We simulate an ensemble where the true DGP is B-R ($\beta = 2/3$, $\rho = 0.8$), then demonstrate that pooling produces apparent Gompertz-like decay (Algorithm 11).

Algorithm 11: Demonstrating ensemble averaging artifact. The noise parameter $\sigma = 0.4$ is technology-invariant in this simulation.

Input: simulation parameters
Output: Demonstration of ensemble averaging artifact
Setup: $N = 120$, $k_i \sim 0.02 + \text{Exp}(2)$, $t_{0,i} \sim \Gamma(2, 10, 10)$, $L_i = 100$, $\sigma = 0.4$, $t \in [0, 50]$[a]
Step 1: Generate time-series: foreach *technology* $i = 1$ *to* N **do**
  $Y_{i,t} = \frac{L_i}{\left[1+\exp(-\beta k_i(t-t_{0,i})+\beta\sigma(\epsilon_t+\rho\epsilon_{t-1}))\right]^{1/\beta}}$, $\beta = 2/3$, $\rho = 0.8$
end
Step 2: Calculate transformations:
Calculate collapse plot transformations per algorithm 8

[a]Other parameter choices yield similar results, including heterogeneous σ_i drawn from a Gamma distribution.

Results. Figure 4.28 shows Gompertz-curve growth collapse plots (Algorithm 8) for empirical data and 120 simulated B-R time-series (Algorithm 11). Pooling across B-R processes with heterogeneous k_i and $t_{0,i}$ creates apparent exponential decay, mimicking Gompertz behavior. Figure 4.28a shows the collapse plot of log-differences for Gompertz.

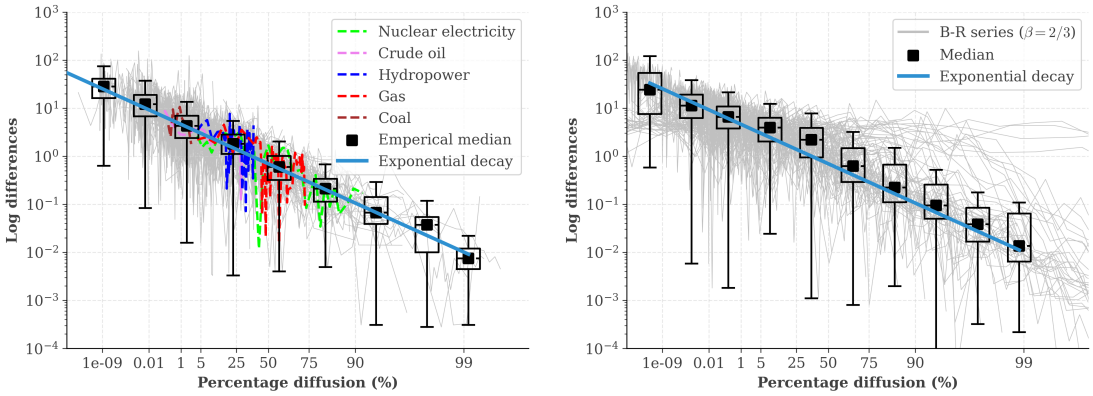


Figure 4.28: Effects of ensemble averaging assuming Gompertz growth when individual time-series follow logistic or Bertalanffy-Richards curves. Individual time-series: gray, black squares: median, blue line: Exponential decay. **(a)** Collapse plot of log-differences under Gompertz normalization. We highlight energy technologies. Individual technology clustering at specific diffusion stages indicates potential ensemble averaging bias. **(b)** Collapse plot across 120 Bertalanffy-Richards series ($\beta = 2/3$) simulated via Algorithm 11. Pooling creates apparent exponential decay, mimicking Gompertz-like behavior.

While ensemble behaviour appears Gompertz-like, individual technologies cluster at specific time and diffusion stages, indicating potential ensemble-averaging bias.

Figure 4.28b shows the same averaging effect using B-R ($\beta = 2/3$) collapsed via

Algorithm 8. The ensemble median (black dots) lies slightly above the true Gompertz curve (green line), unlike empirical data (Figure 4.28a). Some individual time-series (gray lines) exhibit exponential growth (horizontal lines at low diffusion). Ensemble averaging across heterogeneous B-R processes creates Gompertz-like decay patterns. The artifact is robust to changes in simulation size and parameters: it persists for different β values (e.g. $\beta = 1$), different ρ, and heterogeneous σ_i drawn from a Gamma distribution rather than a fixed σ, as long as noise is sufficiently large relative to growth rates. We do not attempt to delineate the exact parameter region where the artifact occurs.

4.15.4 Comparing Gompertz, logistic, and B-R collapse plots

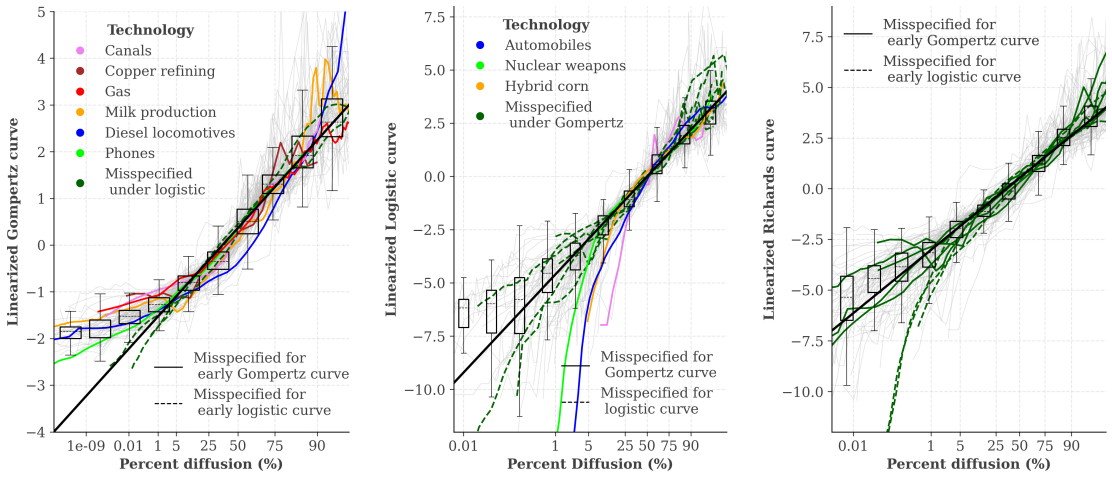
We test for model misspecification using FP-transform collapse plots. Under correct specification, parameters estimated on post-5% data should predict pre-5% trajectories. Comin et al. [91] show S-curves fail to capture very early growth. We omit the first 5% of diffusion, estimate parameters on the remainder, then assess whether omitted observations align with predictions. Systematic deviations indicate fundamental misspecification.

Defining early diffusion threshold. The 5% threshold is defined analytically. For B-R with parameters $\{L_i, k_i, t_{0,i}, \beta_i\}$, solve $Y_{i,t} = 0.05L_i$:

$$t_{5\%,i} = t_{0,i} - \frac{1}{\beta_i k_i} \log\left[\left(\frac{1}{0.05}\right)^{\beta_i} - 1\right] \tag{4.53}$$

Similarly, for Gompertz and logistic curves, we solve $Y_{i,t} = 0.05L_i$ to find $t_{5\%,i}$.

Results and comparison. Figure 4.29 presents FP-transform collapse plots for all three models. Figure 4.29 shows that all models exhibit early-stage deviations, revealing misspecification. Gompertz shows the most systematic deviations (Figure 4.29a), with many technologies consistently deviating from predictions. Logistic (Figure 4.29b) shows



(a) Gompertz fitted to $> 5\%$ diffusion **(b)** Logistic fitted to $> 5\%$ diffusion **(c)** B-R ($\beta = \frac{2}{3}$) fitted to $> 5\%$ diffusion

Figure 4.29: (G)FP-transform collapse plots comparing Gompertz, logistic, and B-R ($\beta = 2/3$) curves fitted to $> 5\%$ diffusion (i.e., omitting early $< 5\%$ diffusion). **(a)** Gompertz curve. We highlight technologies that deviate from the model description. Green dashed lines show technologies fitted better by logistic than by Gompertz. **(b)** Logistic curve. We highlight technologies that deviate from the model description. Green dashed lines show technologies fitted better by Gompertz than by logistic. **(c)** B-R curve ($\beta = 2/3$) omitting first 5% diffusion. Green solid lines fit poorly under Gompertz; green dashed lines fit poorly under logistic.

fewer deviations; specific technologies (*automobiles*, *nuclear weapons*) deviate before 5–10% diffusion. Technologies deviating under Gompertz often fit better under logistic (e.g., *steam turbines*, *air conditioners*), and vice versa. B-R (Figure 4.29c) exhibits the least systematic deviations.[27]

Limitations. All models provide reasonable ad-hoc fits (high R^2), but this test ranks them: B-R first, logistic second, Gompertz third. No single model captures early adoption ($< 5\%$ diffusion) for all technologies. Technologies that fit poorly under Gompertz often fit well under logistic and vice versa, suggesting technology-specific idiosyncrasies that none of the models investigated here fully capture. This quasi-out-of-sample test (estimating on post-5% data, testing on pre-5% data) reveals fundamental heterogeneity in early diffusion dynamics across technologies. This represents a limitation of assuming universal functional

[27] *Automobiles* and *nuclear weapons* deviate before $\approx 1\%$ diffusion.

forms. However, our backtesting analysis (Section 4.16) demonstrates that the B-R model with $\beta = 2/3$ produces robust out-of-sample forecasts: despite early-stage heterogeneity, the model captures sufficient structure for medium- to long-term prediction.

4.16 Testing forecast performance on empirical data

This section gives the detail behind the cross-technology backtest of main paper Figure 3, in which we train the model up to a specified diffusion level and forecast the rest of the adoption trajectory. We first set out the forecasting setup and the accuracy and calibration metrics (Section 4.16.1), then report the backtest results and explain the shape of the error curve. We also benchmark against the Bass model (Section 4.16.3). We calibrate the widening transform that corrects the prediction intervals (Section 4.16.4), and check robustness to the subjective data-cleaning step (Section 4.16.5), to the prior, and to the technology-invariant β and ρ (Section 4.16.6). We close by bracketing the prior between an uninformed and a diffusion-aware specification, to bound the achievable forecast quality (Section 4.16.7).

Throughout, we use the *shifted* forecast variant (Section 4.7.4). This variant adjusts the fitted model so the residual is zero at the last observation. Results are robust to the choice of shifted versus non-shifted forecast. Surrogate medians, confidence bands, and PIT distributions in this section come from the empirical surrogate datasets of Section 4.9, using the truncated $\hat{\sigma}_i$ described there.

4.16.1 Methods underlying main paper Figure 3

4.16.1.1 Priors

All models use technology-specific parameters $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$. We use the universal prior defined in Eq. (4.35) (Section 4.7.4), shown in Figure 4.30. The same hyperparameters apply to every technology. Only the t_0 location is anchored to each series, through the preliminary growth-rate estimate $\hat{k}_{\text{ssr},i}$. This section explains how we obtain that estimate, justifies the hyperparameter choices, and reports robustness checks.

Growth-rate estimate. We obtain $\hat{k}_{\text{ssr},i}$ by minimising the sum of squared residuals of the GFP transform (Eq. 4.16) over L_i:

$$\hat{L}_{\text{ssr}} = \arg\min_{L_i} \sum_t \left[f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i}) \right]^2, \qquad L_i \in [y_{i,\max},\ 10{,}000].$$

The OLS slope at $L_i = \hat{L}_{\text{ssr}}$ gives $\hat{k}_{\text{ssr},i}$. Conditional on L_i, this slope is approximately unbiased, so the estimate is a fast and reliable anchor for the t_0 location. If the fit returns a NaN value we use a fallback $\hat{k}_{\text{ssr}} = 0.15$.

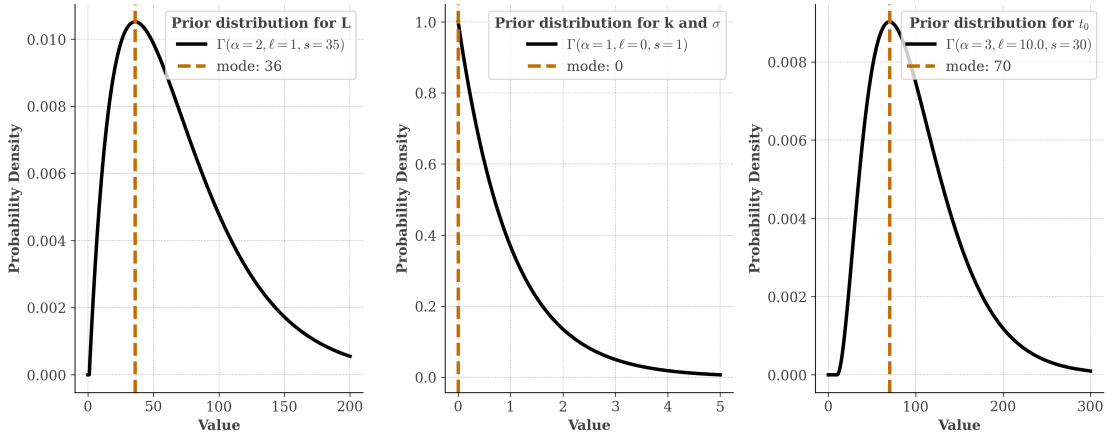


Figure 4.30: Priors used in the main text (universal prior). The t_0 prior is shown for a technology with $\hat{k}_{\text{ssr},i} = 0.3$ and $T_{\max} = 20$, so that $\ell_{t_0} = \max(5, 20 - 3/0.3) = 10$.

Design rationale. *L prior ($\alpha_L = 2$, $\ell_L = 1$, $s_L = 35$).* The lower bound $\ell_L = 1$ forces $\tilde{L}_i > 1$, so the technology has not yet saturated. The shape $\alpha_L = 2$ places the mode at 36 times current adoption. This encodes that, for a technology observed early, the saturation level usually lies well above current adoption. Because $\alpha_L = 2$ rather than 1, the prior peaks at this mode rather than at its lower bound, so it does not assume the technology is already near saturation. The scale $s_L = 35$ keeps the prior wide but retains mass at low values, so it still admits technologies that saturate early. Forecasts stay well-calibrated at a 50% origin, where the true asymptote is twice the current value and far below the mode (Section 4.16.6).

t_0 prior ($\alpha_{t_0} = 3$, $\ell_{t_0} = \max(5, T_{\max} - 3/\hat{k}_{\text{ssr},i})$, $s_{t_0} = 30$). The location anchors the inflection to the data. A fixed small location would be unreasonable for a long series, placing most mass far before the observed window. The offset $3/\hat{k}_{\text{ssr},i}$ scales with the inverse growth rate, so faster-growing technologies expect the inflection closer to $T_{\max}$. The floor at 5 keeps the location positive. The shape $\alpha_{t_0} = 3$ places the mode 60 steps ahead of the location, encoding that at early diffusion the inflection lies well ahead. The scale $s_{t_0} = 30$ keeps the prior wide.

k and σ priors ($\alpha = 1$, $\ell = 0$, $s = 1$). Both are unit exponentials (Section 4.7.4). Unlike L and t_0, they are well identified even at early diffusion, so we do not anchor them to the data.

The numerical choices ($s_L = 35$, $s_{t_0} = 30$, and the offset $3/\hat{k}$) are validated by backtesting on the full cross-section (Section 4.16.6). The prior is well-calibrated across a diverse set of technologies. We therefore take it as a reasonable default for technologies the model has not seen, such as solar PV and wind.

When the t_0 location is valid. The location ℓ_{t_0} is valid when it lies at or below the true t_0. We derive the diffusion level at which this holds. In the deterministic B-R

model (4.6) with $\beta = 2/3$, the noise-free generalised Fisher–Pry transform (Eq. 4.16) is

$$\frac{1}{\beta}\log\bigl(d^{-\beta} - 1\bigr) = -k\,(t - t_0),$$

where $d = Y/L$ is the diffusion level. The time from t_0 to diffusion d is

$$T_{\max} - t_0 = -\frac{1}{\beta\,k}\log\bigl(d^{-\beta} - 1\bigr).$$

The condition $\ell_{t_0} \le t_0$ (with $\hat{k} \approx k$) requires $T_{\max} - t_0 \le c/k$ for offset c. Both sides scale with $1/k$, so the growth rate cancels and the threshold depends only on c and β:

$$d \;\le\; \bigl(1 + e^{-\beta\,c}\bigr)^{-1/\beta}.$$

For $\beta = 2/3$ and our default $c = 3$, this gives $d \le (1 + e^{-2})^{-3/2} \approx 83\%$. So ℓ_{t_0} lies at or below the true t_0 whenever observed diffusion is below about 83%. Above that level the data is highly informative about t_0, and in-sample diagnostics readily reveal the inflection, so the prior can be adjusted by hand. We keep $c = 3$. Results are robust across $c \in \{2, 3, 4\}$, which shift the threshold to about 70%, 83%, and 90% (Section 4.16.6).

Empirical Bayes. The universal prior is an empirical Bayes prior. Two of its hyperparameters depend on the series. First, the t_0 location uses the preliminary estimate $\hat{k}_{\mathrm{ssr},i}$. Second, the normalisation by $y_{T_{\max}}$ sets the scale of the L prior, so the saturation level is a multiple of current adoption. After this the sampler treats the prior as fixed. The prior does not depend on the forecast origin, so the forecasts are out-of-sample. However, the functional form and the technology-invariant parameters $\beta = 2/3$ and $\rho = 0.8$ were calibrated on the full cross-section (Section 4.13), so the backtests are not fully independent of the test data.

4.16.1.2 Logistic and Gompertz models

Here we provide the data-generating processes and likelihoods for the two competing S-curve models. Both use the same priors as B-R, with technology-invariant $\rho = 0.8$ and $\epsilon_{i,T} = 0$.

The logistic data-generating process is

$$Y_{i,t}^{(l)} = \frac{L_i}{1 + \exp\big(-k_i(t - t_{0,i}) + \sigma_i(\rho\,\epsilon_{i,t-1} + \epsilon_{i,t})\big)}. \tag{4.54}$$

Under the Fisher–Pry (FP) transformation, adding $k_i(t - t_{0,i})$ to both sides yields

$$x_{i,t}^{(l)} = \log\!\left(\frac{L_i - Y_{i,t}}{Y_{i,t}}\right) + k_i(t - t_{0,i}) = \rho\sigma_i\epsilon_{i,t-1} + \sigma_i\epsilon_{i,t}, \tag{4.55}$$

with $\epsilon_{i,t} \sim \mathcal{N}(0,1)$. The Jacobian is diagonal with log-determinant

$$\log\det\!\left(\frac{\partial f}{\partial Y_i}\right) = \sum_{t=1}^{T}\Big[\log L_i - \log\big[Y_{i,t}(L_i - Y_{i,t})\big]\Big].$$

The τ-step forecast for posterior draw m is

$$\mathbf{Y}_{m,i}^{(l)}(\tau) = \frac{L_{m,i}}{1 + \exp\big(-k_{m,i}(\boldsymbol{\tau} - t_{0,m,i}) + \boldsymbol{\epsilon}_{T+1:T+\tau,m}^{(l)}\big)}.$$

The Gompertz data-generating process is

$$Y_{i,t}^{(g)} = L_i \exp\big(-\exp\big(-k_i(t - t_{0,i}) + \sigma_i(\rho\,\epsilon_{i,t-1} + \epsilon_{i,t})\big)\big). \tag{4.56}$$

Adding $k_i(t - t_{0,i})$ to the GFP-transformed observations yields

$$x_{i,t}^{(g)} = \log\log\!\left(\frac{L_i}{Y_{i,t}}\right) + k_i(t - t_{0,i}) = \rho\sigma_i\epsilon_{i,t-1} + \sigma_i\epsilon_{i,t}, \tag{4.57}$$

with $\epsilon_{i,t} \sim \mathcal{N}(0,1)$. The Jacobian log-determinant is

$$\log\det\left(\frac{\partial f}{\partial Y_i}\right) = -\sum_{t=1}^{T} \log\left[Y_{i,t}\left(\log L_i - \log Y_{i,t}\right)\right].$$

The τ-step forecast for posterior draw m is

$$\mathbf{Y}^{(g)}_{m,i}(\tau) = L_{m,i}\exp\Big(-\exp\Big(-k_{m,i}(\boldsymbol{\tau} - t_{0,m,i}) + \boldsymbol{\epsilon}^{(g)}_{T+1:T+\tau,m}\Big)\Big).$$

4.16.1.3 Measuring point forecast errors

Let $\{Y_{i,t}\}_{t=1,\ldots,T_i}$ be technology i's observed time-series. We evaluate forecast accuracy at every diffusion target $d^* \in \{1\%, 2\%, \ldots\}$ as follows. The fractional year $t_i^*(d^*)$ at which the observed series crosses $d^* \cdot L_i$ is obtained by linear interpolation between the first observation below and the first observation above this level,

$$t_i^*(d^*) = t_k + \frac{d^*\cdot L_i - Y_{i,t_k}}{Y_{i,t_{k+1}} - Y_{i,t_k}}\,(t_{k+1} - t_k), \qquad Y_{i,t_k} \le d^*\cdot L_i \le Y_{i,t_{k+1}}. \tag{4.58}$$

The observed value at t_i^* equals $d^*\cdot L_i$ by construction. The predicted value $\hat{Y}_i(t_i^*)$ is obtained as follows. For each posterior MCMC sample $(L_s, k_s, t_{0,s})$ we evaluate the deterministic shifted S-curve (Section 4.7.4) at $t = t_i^*$, and take the cross-sample median. From this we build two error statistics.

Pointwise absolute log error (ALE) and its cross-technology median (MedALE). Primary metric, used in main paper Fig. 3a, b and Fig. 4.31 of this SI. Per technology, the absolute log error at diffusion target d^* is

$$\mathrm{ALE}_i(d^*) = \left|\log\left(\frac{\hat{Y}_i(t_i^*)}{Y_i(t_i^*)}\right)\right|. \tag{4.59}$$

We aggregate across technologies by the cross-technology median, denoted MedALE,

$$\mathrm{MedALE}(d^*) = \underset{i \in \mathcal{I}(d^*)}{\mathrm{median}}\ \mathrm{ALE}_i(d^*), \tag{4.60}$$

where $\mathcal{I}(d^*)$ is the set of technologies whose diffusion reaches d^* within the observed window. The cross-technology median is robust to a small number of outlier-fit technologies whose per-horizon log error sits substantially above or below the typical level. It pairs naturally with the median posterior forecast $\hat{Y}_i$ to give a *typical technology, typical forecast* summary. We also report the cross-technology *mean* as a robustness check (Table 4.11).

The noise floor near the forecast origin. A correctly-specified model still makes errors, because the data carry irreducible noise. The expected absolute log error of a single forecast is bounded below by this noise floor (Eq. (4.45), Section 4.9). At $d = 0.05$ the floor is about 0.25, the best accuracy any model can reach there. The surrogate MedALE sits at the floor for short forecast horizons.

From about $d^* = 15\%$, where the noise floor has dropped enough to no longer dominate, the empirical MedALE for B-R rises above the surrogate 5–95% band, and the gap grows through the middle of the diffusion range. This gap is the model's error beyond noise: parameter-estimation error and structural departures of the data from the B-R process. We estimate $\hat{\sigma}_i^{1\text{–}95}$ on the diffusion range where the B-R model is well specified, excluding the early super-exponential regime and the saturation tail (Section 4.9). Re-estimating it on the wider 0–95% or tighter 5–95% window shifts the band only slightly and leaves the empirical curve above it from $d^* \approx 15\%$ until it re-enters near saturation (Section 4.16.6).

4.16.1.4 Distributional calibration at each diffusion target

We assess distributional calibration with Q-Q plots of the probability integral transform (PIT), following the methodology of Section 4.8. For technology i at diffusion target d^*, let

F_i be the forecast CDF of the posterior predictive distribution at the crossing year $t_i^*(d^*)$ given by the interpolation method of Section 4.16.1.3, and $Y_i(t_i^*)$ the realised value there. The PIT value is

$$u_i(d^*) = F_i\big(Y_i(t_i^*)\big).$$

This interpolation yields one PIT value per technology per diffusion target. We treat distinct technologies as independent (pairs such as *sailing ships/steamships* or *telegrams/telephones* are a small minority of the cross-sector dataset), so at each d^* the PIT values form an i.i.d. sample of size n equal to the number of technologies whose data crosses d^*. This n sets the Beta-order-statistic confidence band of Section 4.8. A Q-Q curve on the 45° line is calibrated.

4.16.2 Backtest results

4.16.2.1 Point-forecast accuracy

Main paper Fig. 3a, b reports the MedALE for the B-R, Logistic, and Gompertz models on the 120-technology calibration set, peaking near $d^* \approx 0.5$ and declining toward saturation. Here we add three views of point-forecast accuracy: the same metric resolved by forecast horizon in years (Fig. 4.31), the MedALE and bias by forecast origin (Table 4.10), and the pointwise error by diffusion target (Table 4.11).

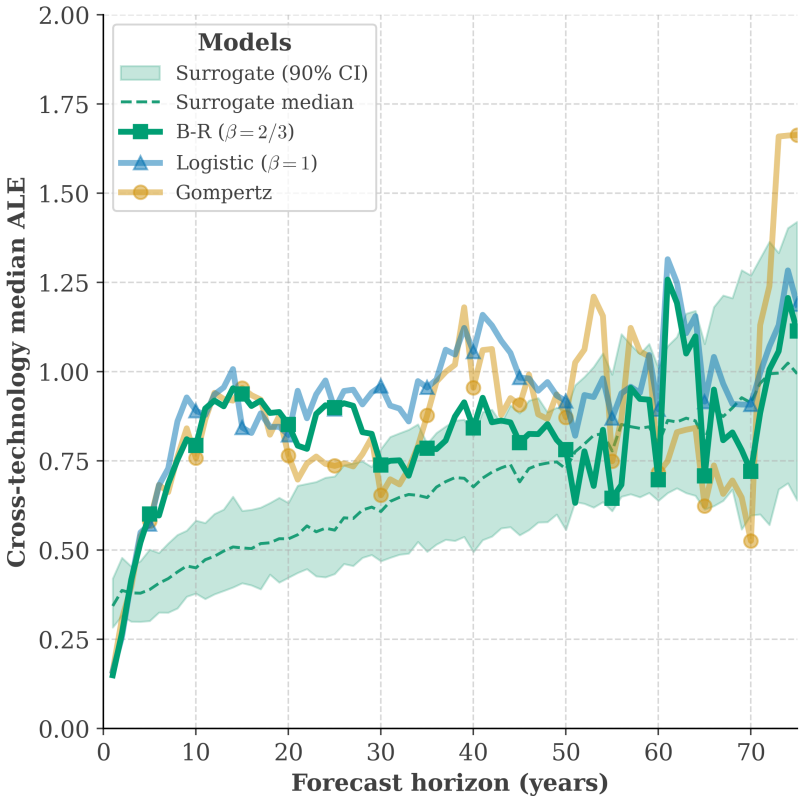


Figure 4.31: MedALE on the calibration set as a function of forecast horizon in years, at the 5% forecast origin. The teal band is the 5–95% surrogate envelope under the same cross-technology median aggregation. The empirical B-R curve exits the surrogate upper bound at about year 5. It oscillates around 0.7 to 0.9 through year 30. It then re-enters the surrogate band from about year 30 to 35 onward. The band widens at long horizons because fewer technologies remain in sample. Markers are placed at horizons $5, 10, 15, \ldots$ for readability.

Table 4.10 reports, for each model and forecast origin, the MedALE averaged over diffusion targets above the origin and the pooled median signed log error (a measure of bias). B-R has the lowest averaged MedALE at the 5% origin. Gompertz is lowest at the 10%, 25%, and 50% origins, but only through a large negative bias (-0.64 at 5%, shrinking at later origins). That is, under-prediction in level space partially cancels noise-driven over-prediction, lowering the absolute log error in a way a bias-free metric cannot. The pointwise ALE confirms this (Table 4.11): B-R is lowest at the high targets (50%, 70%, 100%), and Gompertz wins only at 10% and 25%. B-R's and Logistic's pooled bias are closest to zero at the 5% origin. B-R's bias remains low at the 10% and 25% origins and trends more positive at the 50% origin.

Table 4.10: MedALE, averaged across diffusion targets d^* on a 1% grid from the training origin up to 100%, and pooled median signed log error by model and forecast origin. This is the same metric reported per d^* in main paper Figure 3a, b, c, summarised across diffusion targets. The signed log error is the median of $\ln(\hat{Y}/Y)$ pooled across (technology, integer year horizon) pairs at each origin. Lower averaged median ALE is better. Bias closer to zero is better. The default universal prior of Eq. (4.35) is used throughout.

	Averaged cross-tech median ALE			Median signed log error		
Forecast origin d_{train}	B-R	Logistic	Gompertz	B-R	Logistic	Gompertz
5%	0.90	1.05	1.01	-0.07	-0.02	-0.64
10%	0.72	0.77	0.63	+0.07	+0.20	-0.33
25%	0.56	0.63	0.38	+0.11	+0.15	-0.13
50%	0.34	0.48	0.26	+0.30	+0.24	+0.12

Table 4.11 reports the pointwise ALE at five diffusion targets ($d_{\text{train}} = 5\%$) under both cross-technology median (MedALE) and mean (MALE) aggregations. The model rankings are the same under both. Absolute values are higher under the mean by about 0.1 to 0.2, reflecting moderate right skew, where a few technologies are harder to fit and pull the mean above the median, with no single outlier dominating. The gap is largest at $d^* = 100\%$, where N is smallest.

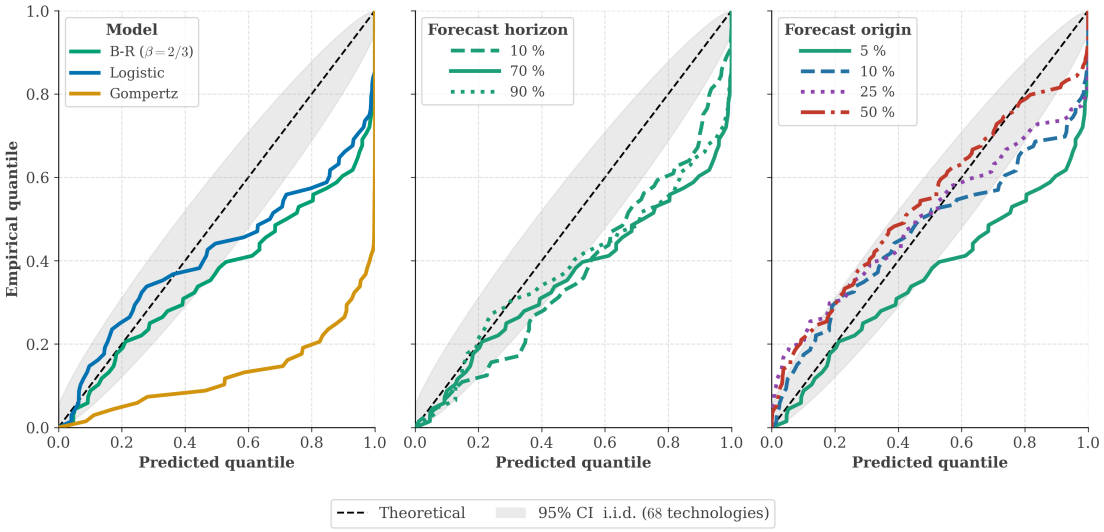
Table 4.11: Pointwise absolute log error of the median forecast at forecast origin $d_{\text{train}} = 5\%$ under the universal prior of Eq. (4.35), evaluated at five fractional diffusion targets d^*. We report the cross-technology median as the primary aggregation (used in main paper Figure 3a, b and throughout this SI) and the cross-technology mean as a robustness check. Lower is better. Bold marks the best model at each d^* within each aggregation.

Aggregation	Model	$d^* = 10\%$	25%	50%	70%	100%
Cross-tech median	B-R ($\beta = 2/3$)	0.374	0.907	**1.023**	**1.005**	**0.868**
Cross-tech median	Logistic	0.457	1.032	1.220	1.209	0.948
Cross-tech median	Gompertz	**0.345**	**0.804**	1.092	1.285	1.171
Cross-tech mean	B-R ($\beta = 2/3$)	0.518	0.931	**1.139**	**1.155**	**1.091**
Cross-tech mean	Logistic	0.551	1.023	1.232	1.235	1.166
Cross-tech mean	Gompertz	**0.489**	**0.890**	1.180	1.257	1.240

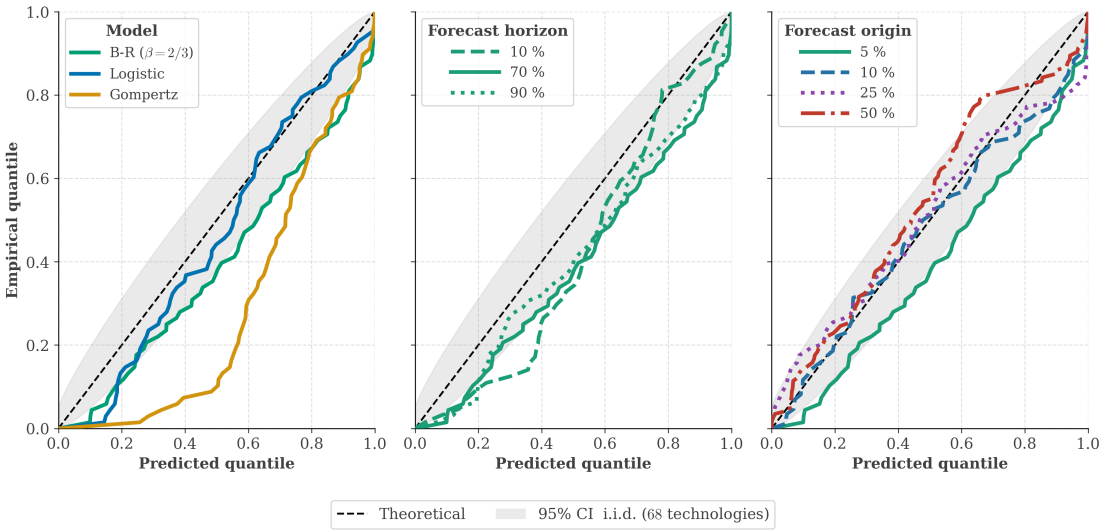
4.16.2.2 Calibration results

We now turn from point accuracy to distributional calibration. Figure 4.32 shows the PIT Q-Q plots without and with the post-hoc widening transform.

Figure 4.32a shows the un-widened PIT. The B-R curve sits below the diagonal across



(a) Baseline (un-widened)



(b) Widened (asymmetric)

Figure 4.32: Forecast calibration under the universal prior on the 120-technology set, with FP-shifted forecasts. PIT values are evaluated at the fractional year where each technology's diffusion crosses the target. **(a)** Baseline calibration, without the post-hoc widening of Section 4.7.5. **(b)** The same comparisons after the asymmetric widening of Eq. 4.36, with anchors calibrated in Section 4.16.4. Within each row, the left panel compares the B-R, Logistic and Gompertz models at a 5% forecast origin and a 70% forecast horizon. The middle panel shows B-R coverage at a 5% forecast origin across three forecast horizons (10%, 70% and 90% diffusion). The right panel shows B-R coverage at a 70% forecast horizon across four forecast origins (5%, 10%, 25% and 50% diffusion). The forecast origin is the diffusion level at which a forecast is made and the forecast horizon the diffusion level at which it is evaluated. Same layout as main paper Figure 3(d-f). The grey band is the 95% pointwise i.i.d. confidence envelope for a uniform draw of $K = 68$ technologies.

most of the range, most strongly in the middle, with two implications. At high percentiles

too many observations land in the upper tail, so the forecast upper bound is too narrow

and B-R is overconfident there. At low percentiles too few land in the lower tail, so the lower bound is mildly conservative. The asymmetric widening transform of Section 4.7.5 stretches the upper tail more than the lower, correcting the upper-tail overconfidence while keeping the lower bound conservative. After widening, the B-R curve tracks the diagonal better at every forecast origin and diffusion target (Fig. 4.32b). We calibrate the anchors in Section 4.16.4 and validate them by leave-one-technology-out cross-validation (Table 4.15).

B-R. The Q-Q curve tracks the diagonal at low percentiles and falls below it at high percentiles, mostly within the confidence band. Its baseline KS against the uniform averages 0.23 across origins, and the widening reduces this to a mean of 0.13 (Table 4.15).

Logistic ($\beta = 1$). Logistic's pooled bias at the 5% origin is slightly smaller than B-R's (-0.02 vs -0.07, Table 4.10). After widening it is only marginally better calibrated than B-R at the 5%, 10%, and 25% origins and slightly worse at the 50% origin (Table 4.14). It has worse MedALE than B-R at every origin (Table 4.10), and needs more widening to achieve the same calibration. The widening exponent needed for the Logistic model is not robust across the 120-technology and the less-filtered 154-technology sets (Table 4.17), so it is not a reliable choice for unseen technologies.

Gompertz. Gompertz has the largest pooled bias, negative at every origin below 50% (Table 4.10). Its widening exponent exceeds B-R's at every origin (Table 4.14), most at low diffusion where the bias is strongest. Note that a median-preserving widening cannot move the biased median, so Gompertz stays miscalibrated.

4.16.2.3 Interpreting the results

The B-R MedALE rises from the early-diffusion noise floor, plateaus between $d^* \approx 30\%$ and $d^* \approx 80\%$, and declines toward saturation (main paper Fig. 3a, b). We explain this shape, link it to the PIT calibration, and end with our choice of model.

Downward bias in $\hat{k}$. The posterior median $\hat{k}$ sits below the full-sample maximum-likelihood estimate of k on the calibration set. On the 68-technology subset whose data crosses $d = 5\%$ (the set used in main paper Fig. 3a), the cross-technology median ratio $\hat{k}/k_{\text{full}}$ is approximately 0.75 at $d_{\text{train}} = 5\%$. The bias decreases with more training data (i.e. later forecast origins). The bias is much smaller on the empirical surrogate of Section 4.9 ($\hat{k}/k_{\text{true}} \approx 0.94$ at $d_{\text{train}} = 5\%$). A plausible source is misspecification of the noise model at early diffusion. The surrogate is generated under the model's noise structure by construction, so this source is largely absent there.

Heterogeneity in the true asymmetry β_i. The B-R model is fit with a universal $\beta = 2/3$. Technology-specific β_i are observed in the data (Section 4.13, Fig. 4.22b). Modest cross-technology departures from $\beta = 2/3$ introduce a parameter-fit mismatch between the universal model and the technology-specific truth, contributing to the empirical MedALE shape on top of the $\hat{k}$ bias above.

Reproducing the shape. A deterministic simulation isolates the contribution of each candidate mechanism. We draw $N = 120$ technologies with heterogeneous k_i, make shifted forecasts that pass through the training endpoint at $d_{\text{train}} = 5\%$, and take L as known. A 25% downward shift in $\hat{k}$ alone (Table 4.12, column B) reproduces the plateau-and-decay shape. Adding cross-technology heterogeneity in the true β_i around 2/3 while the model assumes $\hat{\beta} = 2/3$ (column C) slightly widens the plateau. The empirical surrogate of Section 4.9, where neither condition applies, shows no plateau-and-decay (main paper Fig. 3a band).

Implication for PIT calibration. The downward $\hat{k}$ bias also explains the upper-tail overconfidence of the un-widened PIT seen above. Adoption grows faster than the median forecast expects, so too many observations fall in the upper tail, which the asymmetric

Table 4.12: Cross-technology MedALE at $d_{\text{train}} = 5\%$ on $N = 120$ synthetic technologies, L known, FP-shifted forecasts. Column B: downward $\hat{k}$ bias ($\hat{k} = 0.75\,k$, $\hat{\beta} = \beta = 2/3$). Column C: B with heterogeneous true $\beta_i \sim \mathcal{N}(2/3, 0.08^2)$ and model $\hat{\beta} = 2/3$.

d^*	B: $\hat{k}$ bias	C: B + β_i het
0.10	0.167	0.168
0.30	0.371	0.375
0.50	0.395	0.401
0.70	0.341	0.347
0.95	0.129	0.133

widening then corrects more aggressively than the lower tail.

Saturation forecasts from $d_{\text{train}} = 5\%$ versus $d_{\text{train}} = 10\%, 25\%$. B-R forecasts made from the 5% origin are more accurate near saturation than those made from the 10% or 25% origins (main paper Fig. 3b). This is counterintuitive, since more training data should help. For diffusion targets $d^* \leq 70\%$ the usual ordering holds, where more training data gives smaller error. Near saturation the error is dominated by the asymptote L, and the wide universal L prior is the likely culprit. The diffusion-aware prior of Section 4.16.7 tightens the L prior by diffusion stage and materially lowers the high-diffusion MedALE, pointing to a prior issue rather than a fundamental limit.

Model selection. No model absolutely dominates on every metric, but B-R is the best choice across forecast origins. Gompertz attains the lowest averaged MedALE at the 10%, 25%, and 50% origins (Table 4.10), but only because a large negative bias cancels noise in the median forecast. The same bias leaves it badly calibrated, and a median-preserving widening cannot remove it. Logistic matches B-R's calibration after widening but has higher MedALE and needs more widening. B-R alone is accurate (lowest MedALE at the 5% origin and at almost every diffusion target), nearly unbiased, and well calibrated after widening (mean widened KS 0.13). Our central application forecasts from early diffusion, where solar PV sits at about 2% of its estimated saturation and wind at about 40 to 50%. At higher diffusion origins, model performance is more similar, and the choice of model is

less consequential for forecast quality. We therefore select B-R as our primary model.

4.16.3 The Bass model

We benchmark a fourth model class against B-R: the Bass diffusion model [4], which explicitly separates innovation (p) and imitation (q) effects. The Bass model is often presented as a differential equation, but it has a closed-form solution for the cumulative diffusion $Y(t)$, given by

$$Y(t) = L\,\frac{1 - e^{-k(t-t_0)}}{1 + r\,e^{-k(t-t_0)}}, \qquad k = p + q, \quad r = q/p, \tag{4.61}$$

where L is the market potential, r is the imitation-to-innovation ratio, and $Y(t_0) = 0$. Setting $r = 14$ places the inflection at $Y^* = L(r-1)/(2r) = 46.4\%$ of L, virtually identical to the B-R model with $\beta = \frac{2}{3}$ (46.5%), enabling a like-for-like comparison. We apply the same MA(1) noise structure in FP-transformed space with $\rho = 0.80$ and $\epsilon_{i,T} = 0$, and the same universal prior (Eq. (4.35)).

Applying this framework to the 120-technology dataset gives substantially worse backtests than B-R. The median $\log(\hat{Y}/Y)$ is strongly negative (i.e. forecasts are biased downward), against near-zero bias for B-R. The distributional forecasts are very poorly calibrated. We did not recalibrate the prior for Bass, so the comparison is not fully controlled, and a dedicated recalibration might narrow the gap. We therefore draw no strong conclusion about the Bass model class, beyond noting that under our prior and noise assumptions it is clearly less accurate than B-R on this dataset.

4.16.4 Calibrating the widening transform on backtesting data

The adaptive widening transform introduced in Section 4.7.5 requires the anchor values $w^*(d)$ that enter Eq. (4.38). Here we calibrate those anchors on the 120-technology backtesting set, validate the asymmetric form (Eq. (4.36)) against a symmetric variant and against the empirical surrogates, and report the resulting calibration improvement.

Fitting the widening exponent. We calibrate one exponent w per diffusion origin $d \in \{0.05, 0.10, 0.25, 0.50\}$. For each technology we compute the predictive distribution at the fractional year where the data cross the target diffusion level $d^* \in \{5, 10, \ldots\}\%$. We then pool PIT values across technologies and target levels and choose w to minimise the Kolmogorov–Smirnov (KS) statistic

$$D(F_1, F_2) = \sup_x \big|F_1(x) - F_2(x)\big|, \tag{4.62}$$

where F_1 and F_2 are any two CDFs and x ranges over their common support. Here we use the one-sample form with $F_1 = \hat{F}_{\text{PIT}}$ (the empirical CDF of the pooled PIT values) and $F_2(u) = u$ (the $\mathcal{U}(0,1)$ CDF), so $x = u \in [0,1]$. We minimise using a bounded scalar search on $w \in [0.5, 8]$.

Table 4.13 reports w^* at each forecast origin for the asymmetric form (Eq. (4.36), used throughout the paper) and the symmetric variant in which the same exponent is applied above and below the median (without the $\sqrt{w}$ in the lower-tail branch). The symmetric exponents are smaller because they would over-widen the already-calibrated lower tail. The asymmetric form yields a lower widened KS at $d = 0.05$ (0.14 vs 0.16), $d = 0.10$ (0.10 vs 0.13), and $d = 0.25$ (0.14 vs 0.15). The two forms are essentially tied at $d = 0.50$ (0.12 vs 0.12). The square-root lower-tail exponent was itself chosen from a small sweep across

$w^{1/2}$, $w^{1/3}$, $w^{1/4}$, $w^{1/5}$, and no lower-tail widening. $\sqrt{w}$ gave the best average calibration across the forecast origins.

Table 4.13: Widening exponent w^* by forecast origin d on the 120-technology backtesting set, fitted under the KS objective. **Asymmetric** applies w above the median and $\sqrt{w}$ below (Eq. 4.36). **Symmetric** applies w on both sides. The asymmetric form is used throughout the paper. The widened KS is the deployed value, after applying the anchors adaptively per technology as $w^*(1/\tilde{L}^{\text{med}})$.

	Asymmetric		Symmetric	
Forecast origin d	w^*	widened KS	w^*	widened KS
0.05	1.93	0.142	1.76	0.157
0.10	2.72	0.101	2.03	0.126
0.25	3.58	0.137	2.84	0.151
0.50	1.73	0.122	1.34	0.119

Per-model anchors. The widening exponents above are fitted on the B-R PIT distribution. To check the other models, we re-fit w^* independently on the Logistic and Gompertz distributions at every origin (Table 4.14). B-R needs the least widening overall. Logistic needs more, especially at the 5% origin, but reaches comparable calibration once widened. Gompertz needs the most at every origin and stays the worst calibrated, because a median-preserving widening cannot remove its negative bias. Figure 4.33 shows each model widened with its own exponent: B-R reaches the diagonal at every origin, Logistic as well (slight exception at 25% forecast origin), and Gompertz stays off the diagonal below the 50% origin. As more of the series is observed, the data dominate the prior and the models' growth descriptions converge, so the calibration gap between them narrows (as observed in Figure 4.33 panel d, where all three models lie nearer the diagonal post widening.).

Held-out cross-validation. We fit the widening anchors of Table 4.13 on the backtesting set, so the reported widened KS is an in-sample quantity. We therefore cross-validate by leaving one technology out. We refit the anchors on the remaining technologies, then apply

Table 4.14: Per-model widening on the 120-technology backtesting set. For each model we fit the anchor exponent w^* per forecast origin on the model own un-widened PIT distribution, with the asymmetric widening of Eq. 4.36. Baseline KS measures the un-widened forecasts against uniformity. Widened KS measures the same after the adaptive widening.

Model	Forecast origin d	Baseline KS	w^*	Widened KS
B-R	0.05	0.296	1.93	0.142
	0.10	0.216	2.72	0.101
	0.25	0.215	3.58	0.137
	0.50	0.173	1.73	0.122
Logistic	0.05	0.252	3.37	0.095
	0.10	0.177	2.69	0.091
	0.25	0.240	3.40	0.119
	0.50	0.246	3.82	0.143
Gompertz	0.05	0.599	5.14	0.378
	0.10	0.488	4.28	0.310
	0.25	0.373	6.02	0.198
	0.50	0.176	2.86	0.144

the adaptive widening $w^*(1/\tilde{L}^{\text{med}})$ to the held-out one. We repeat this for every technology and pool the held-out PIT values. The held-out KS matches the in-sample KS to within 0.005 at every origin, and w^* is stable across folds (Table 4.15). The widening is therefore a genuine out-of-sample recalibration of interval width, not an artefact of in-sample fitting. We use the 120-technology in-sample anchors for all forecasts in the paper.

Table 4.15: Leave-one-technology-out cross-validation of the adaptive widening on the 120-technology backtesting set. The model is B-R, the objective is KS, and the transform is the asymmetric sqrt-lower form of Eq. 4.36. KS (raw) is the un-widened PIT KS against the uniform. KS (in-sample) is the adaptive widened KS under the deployed anchors. KS (LOO-CV) refits the anchors on the other $N - 1$ technologies, applies the adaptive exponent to the held-out technology, and pools the held-out PITs. The w^* LOO range is the refit anchor range across folds.

Forecast origin d	w^*	KS (raw)	KS (in-sample)	KS (LOO-CV)	w^* LOO range
0.05	1.93	0.296	0.142	0.145	[1.91, 2.68]
0.10	2.72	0.216	0.101	0.102	[2.27, 2.89]
0.25	3.58	0.215	0.137	0.139	[3.24, 3.65]
0.50	1.73	0.173	0.122	0.121	[1.60, 1.96]

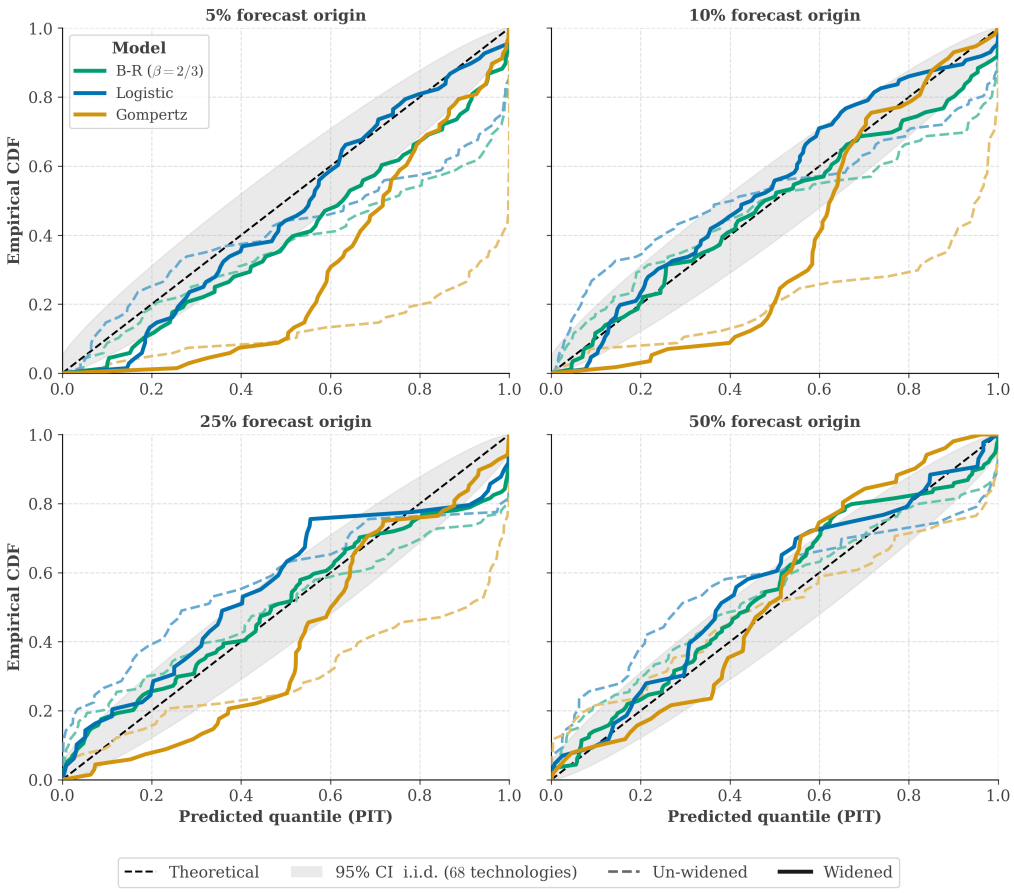


Figure 4.33: Un-widened and Widened PIT calibration of the three models at the four forecast origins, at the 70% diffusion target. PIT computation and layout follow Section 4.16.1.4 and main paper Fig. 3(d). Each model's un-widened curve is the lighter dashed line of the same colour, and its widened curve is the solid line. B-R and Logistic reach the diagonal at every origin.

Validation on the empirical surrogates. For a correctly-specified model the PIT is approximately uniform by construction, so re-fitting w^* on surrogate data should return values close to 1. We tested this on the empirical surrogate at the 5% forecast origin, using the same universal prior as the main analysis. The fitted exponent is $w^* \approx 1.1$, far below the empirical $w^* \approx 1.9$ at the same origin (Table 4.13). It stays near one and is leave-one-out stable whether we estimate the per-technology noise scale $\hat{\sigma}_i$ on the primary 1–95% diffusion window or on the wider 0–95% window that keeps all early-diffusion data (both defined in Section 4.9; Table 4.16), so the 1% early-diffusion cut does not drive the result. The small residual above 1 likely reflects mild prior misspecification, which can bias the posterior [373, 374], rather than miscalibration of the model form.

Table 4.16: Widening exponent w^* on the empirical surrogate at the 5% forecast origin, fitted under the KS objective, for the primary 1–95% and the wider 0–95% noise windows (Section 4.9). The last column gives the leave-one-technology-out range of w^*.

Surrogate	Widening	w^*	w^* (LOO range)
$\hat{\sigma}_i^{1\text{–}95}$ (≈ 0.28)	Asymmetric	1.13	[1.10, 1.14]
$\hat{\sigma}_i^{1\text{–}95}$ (≈ 0.28)	Symmetric	1.10	[1.08, 1.10]
$\hat{\sigma}_i^{0\text{–}95}$ (≈ 0.33)	Asymmetric	1.19	[1.17, 1.22]
$\hat{\sigma}_i^{0\text{–}95}$ (≈ 0.33)	Symmetric	1.13	[1.11, 1.14]

Section 4.8 shows that under a correctly specified model the PIT calibration depends on whether the prior matches the true parameter distribution. The surrogate is correctly specified by construction, so its slightly off-diagonal PITs reflect prior misspecification rather than model form. On real data, prior misspecification may offset model misspecification, consistent with the approximately well-calibrated real-data PITs.

4.16.5 Robustness to subjective data cleaning

Step 4(b) of the data cleaning procedure (Section 4.11.1) removes 34 time-series through partly subjective judgement. Step 6 removes one further series for high correlation with another. To check that the main results do not depend on these removals, we repeat the full backtesting on a set that retains all 35 of these series.[28] The pipeline is identical to the 120-technology analysis. We use the same universal prior of Eq. (4.35). We fit the widening exponent w^* on this set independently. We look up the per-technology widening through the same adaptive map $\hat{d}_i = 1/\tilde{L}_i^{\text{med}}$ as in the main analysis, with no future knowledge of the saturation level.

Figure 4.34 shows the MedALE on the 154-technology set. The model ranking matches the 120-technology set. B-R has generally lower MedALE and the peak in MedALE remains. Gompertz has the largest negative bias (median ≈ -0.54). We do not overlay an empirical

[28]This gives 155 series in total. One of them, annual space launches, declines sharply in its final years. All of its observations then fall within the training window at every backtest origin. It yields no out-of-sample forecast point, so it contributes to no calibration curve. This leaves the 154 technologies analysed here.

surrogate band here, because the surrogate is calibrated on the 120-technology parameter set and would mix datasets.

Baseline calibration is shown in Figure 4.35a and widened calibration in Figure 4.35b. Both visually follow the 120-technology version.

Table 4.17 compares the widening exponent w^* re-fitted on each set, per model. B-R is stable: its exponents change by at most 0.32 across origins (13% in relative terms), and its baseline KS, the distance of the un-widened PIT from the uniform, changes by at most 0.03. The benchmark models move more. Logistic is far more sensitive at the 5% origin, where its exponent falls from 3.37 to 2.47 (difference of $\approx 27\%$) while B-R's barely moves (1.93 to 1.85). The subjective cleaning step therefore does not materially change B-R's calibration, point-forecast accuracy, or widening.

Table 4.17: Widening exponent w^* and baseline KS for the 120-technology and 154-technology backtesting sets, per model, under the KS objective with the asymmetric widening of Eq. 4.36. The 154-technology set retains the 34 subjectively excluded series from Step 4(b). Baseline KS is the distance of the un-widened PIT from the uniform. B-R is shown at every forecast origin; Logistic and Gompertz at the 5% origin only.

	w^*			Baseline KS		
d_{train}	120	154	Δw^*	120	154	ΔKS
B-R ($\beta = 2/3$)						
0.05	1.93	1.85	−0.08	0.295	0.278	−0.017
0.10	2.72	3.04	+0.32	0.217	0.236	+0.019
0.25	3.58	3.57	−0.01	0.216	0.211	−0.005
0.50	1.73	1.50	−0.23	0.174	0.145	−0.029
Logistic						
0.05	3.37	2.47	−0.89	0.251	0.243	−0.007
Gompertz						
0.05	5.14	5.45	+0.31	0.601	0.564	−0.037

The leave-one-technology-out cross-validation of Table 4.15 also holds on the 154-technology set (Table 4.18).

Table 4.18: Leave-one-technology-out cross-validation of the adaptive widening on the 154-technology set. This set adds the 34 subjectively excluded series to the 120 (cf. Table 4.17). Columns are as in Table 4.15. The held-out and in-sample KS again agree to within 0.01 at every origin. The widening calibration is robust to the data-selection step.

Forecast origin d	w^*	KS (raw)	KS (in-sample)	KS (LOO-CV)	w^* LOO range
0.05	1.85	0.278	0.156	0.156	[1.71, 1.88]
0.10	3.04	0.236	0.123	0.129	[2.88, 3.37]
0.25	3.57	0.212	0.143	0.144	[3.43, 3.88]
0.50	1.50	0.144	0.125	0.123	[1.42, 1.52]

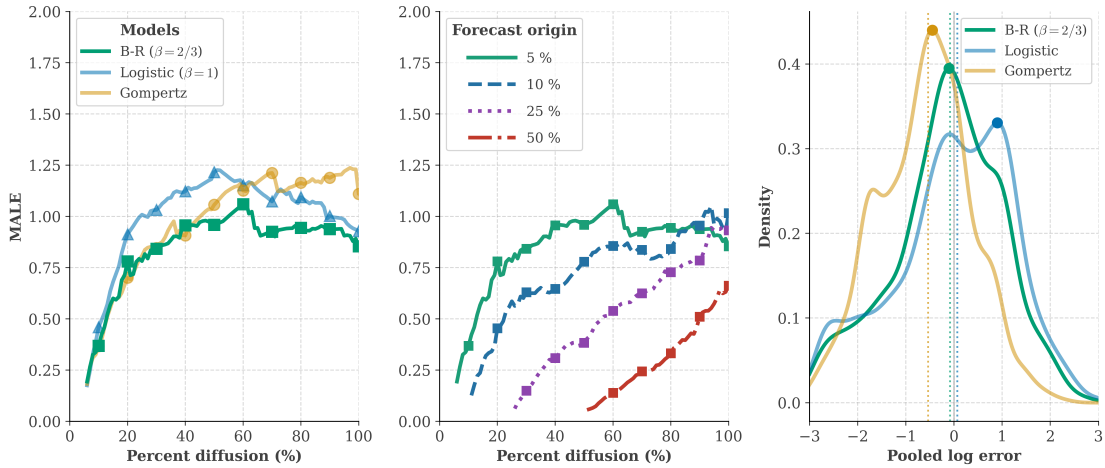
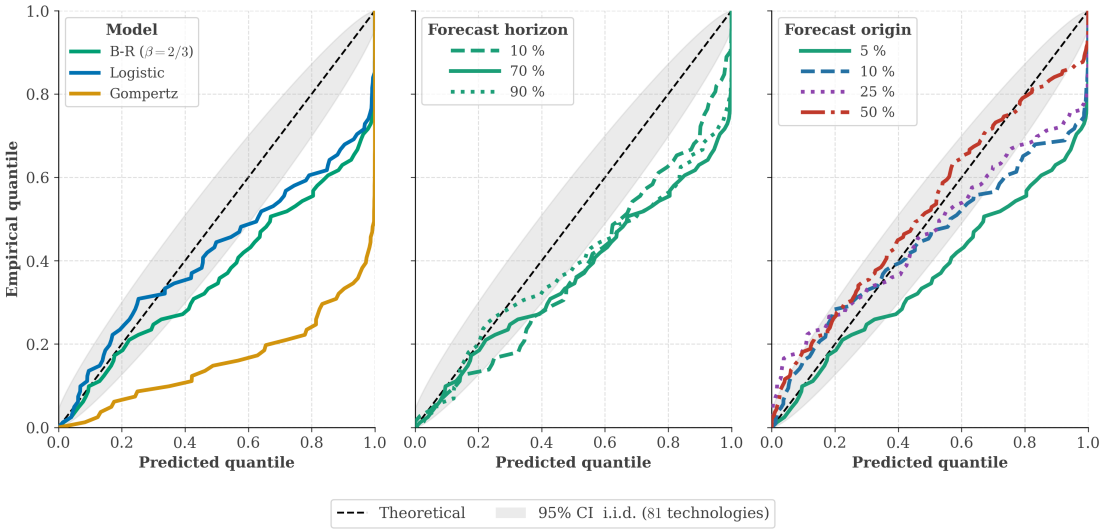
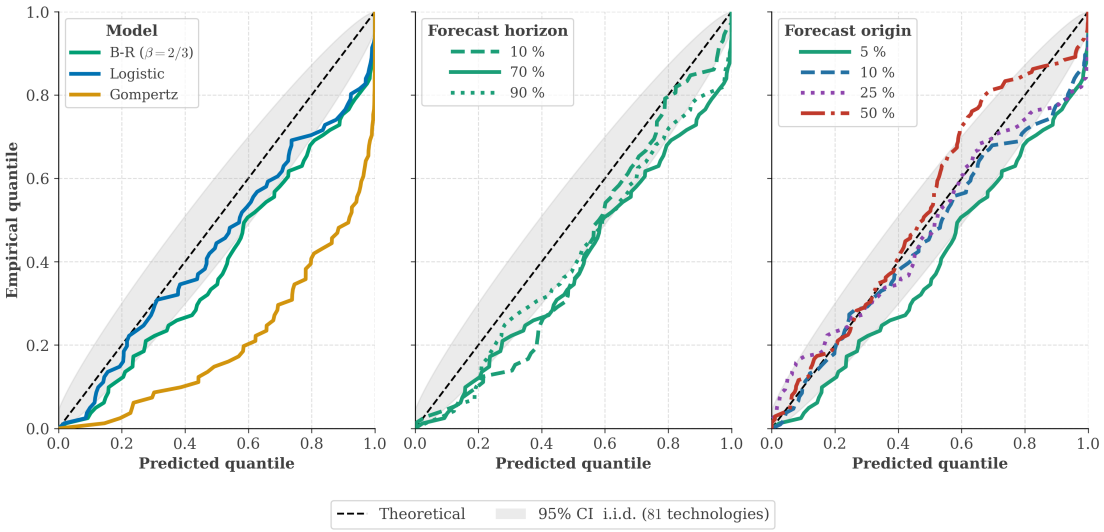


Figure 4.34: Forecast accuracy on the 154-technology dataset using FP-shifted forecasts. Same layout as main paper Figure 3a, b, c. As there, the y-axis median$_i$ is the median across technologies. **(a)** MedALE at each diffusion target for B-R, Logistic, Gompertz at the 5% forecast origin. **(b)** The same for B-R at four forecast origins (5%, 10%, 25%, 50%). **(c)** Pooled per-(technology, horizon) log error $\ln(\hat{Y}/Y)$ density at the 5% origin.



(a) Baseline (un-widened)



(b) Widened (asymmetric)

Figure 4.35: Calibration on the 154-technology dataset. **Top:** baseline (un-widened) PIT. **Bottom:** after asymmetric widening of Eq. 4.36 with anchors re-calibrated on this set (Table 4.17). Same layout as main paper Figure 3(d-f) and Figure 4.32a: within each row the left panel compares the B-R, Logistic and Gompertz models at a 5% forecast origin and 70% forecast horizon, the middle panel shows B-R coverage at a 5% forecast origin across three forecast horizons (10%, 70% and 90% diffusion), and the right panel shows B-R coverage at a 70% forecast horizon across four forecast origins (5%, 10%, 25% and 50% diffusion).

4.16.6 Further robustness checks

Here we present a brief overview of robustness and sensitivity analyses not included in the main text. With the exception of the prior sensitivity checks, we keep the same priors as in main paper Fig. 3. We show sensitivity analyses for prior specification, for varying β and

ρ, and for estimating technology-specific β_i and ρ_i, and discuss how forecast performance improves for more *well-behaved* technologies (i.e. those more similar to our null model in terms of R^2).

Sensitivity to priors. Our results are robust to moderate changes in the prior hyper-parameters, though extreme misspecification does degrade performance. A narrow prior centred near the true L and t_0 improves forecasts. But narrow priors may understate parameter uncertainty and become overconfident, and technologies that fit the narrow prior poorly forecast substantially worse. We therefore keep the prior as uninformative as possible, while using information from data available to us.

We also varied the t_0 offset, using $2/\hat{k}_{\text{ssr}}$ or $4/\hat{k}_{\text{ssr}}$ in place of $3/\hat{k}_{\text{ssr}}$, and found the results not hugely sensitive. A tighter offset ($\ell_{t_0} > T_{\max} - 2/\hat{k}_{\text{ssr}}$) helps marginally at low diffusion but performs poorly at high diffusion. A fixed universal location $\ell_{t_0} = 5$ still works reasonably but is less accurate at low diffusion and worse calibrated than the data-informed location. Setting the prior by diffusion stage helps more, which we examine in Section 4.16.7. Under the tighter offset, the MedALE rankings (Table 4.10) are preserved, with similar widening and calibration, further evidence that B-R is a reasonable default.

Sensitivity to constant parameters. We analyze sensitivity to the choice of technology-invariant β and ρ, varying each independently while keeping the other fixed at the baseline value. Varying ρ has a minimal effect on MedALE and calibration. Varying β has a small effect on MedALE but can strongly affect calibration. For example, $\beta = 1/3$ leads to underdispersed forecasts that underestimate upward deviations, while $\beta = 3/2$ leads to overdispersed forecasts at the lower quantiles of the forecast distribution. For $\beta = 1/2$, forecast performance is similar to the baseline, albeit marginally worse, indicating that the choice of β is not critical as long as it remains within a reasonable range of around $[0.5, 1]$.

We also estimate technology-specific β_i and ρ_i, and find that the resulting forecasts have similar point-forecast accuracy but worse probabilistic calibration than the baseline with constant β and ρ (less uniform PIT distributions and visible deviations from the 45° line in the Q-Q plot). This suggests that technology-specific β_i and ρ_i are not well identified, and that estimating them leads to overfitting and worse out-of-sample performance.

Forecast performance as a function of in-sample fit. Forecast performance is better for technologies with better in-sample fits, as measured by the R^2 of the full-sample fit. Technologies that behave more similarly to the B-R model in-sample are forecast more accurately out-of-sample. This is expected: technologies better described by the assumed DGP in-sample are likely to yield better forecasts out-of-sample.

4.16.7 The role of the prior: bounding forecast quality

The universal empirical-Bayes prior (Eq. (4.35)) occupies a middle ground between two extremes. To put its performance in perspective, we evaluate the B-R model under two bracketing configurations:

- **Wide (agnostic) prior, floor.** $\tilde{L}_i \sim \Gamma(2,\, 1,\, 100)$ with a fixed, non-data-informed t_0 location ($\ell_{t_0} = 5$, $s_{t_0} = 30$). This prior encodes almost no information about the technology's saturation level or timing.

- **Diffusion-aware prior, ceiling.** Priors for L and t_0 are set conditional on the observed diffusion band at the forecast origin, using the five-band specification in Table 4.19. This encodes the qualitative diffusion-stage knowledge that a practitioner would typically infer from visual inspection of the adoption curve. It does not require knowledge of the true asymptote.

The three configurations thereby span the range from prior ignorance (wide prior),

through our default cross-sectionally calibrated prior (universal, $s_L = 35$), to the best case available with observable domain knowledge (diffusion-aware).

Wide prior. Under a wide prior ($s_L = 100$), point accuracy deteriorates. The MedALE is higher than under the universal prior across the whole diffusion range (main paper Fig. 3a, b), most at mid-to-high diffusion where the L posterior is more diffuse. Calibration worsens only slightly. The ordering of B-R, Logistic, and Gompertz is unchanged.

Diffusion-aware prior. The diffusion-aware prior conditions on the approximate diffusion band observable at the time of forecasting. An analyst can distinguish near-exponential early growth from a trajectory with visible curvature without knowing the true asymptote, using only information available at forecast time. In that sense, diffusion-stage awareness is the kind of qualitative prior knowledge a practitioner would typically have.

The prior specification is

$$\pi(\theta_i \mid \text{diffusion band}) = \begin{cases} \tilde{L}_i \sim \Gamma(\alpha_L, \ell_L, s_L), \\ t_{0,i} \sim \Gamma(\alpha_{t_0}, \ell_{t_0}, s_{t_0}), \\ k_i \sim \Gamma(1, 0, 1), \\ \sigma_i \sim \Gamma(1, 0, 1), \end{cases}$$

with hyperparameters varying by diffusion band as in Table 4.19. The state-dependence reflects structural identification differences. At early diffusion the asymptote L_i and inflection timing $t_{0,i}$ are weakly identified from near-exponential growth, so the prior must remain wide. As curvature becomes visible the likelihood becomes more informative about both parameters, and tighter priors are warranted. Because L_i is normalised by the last observed value, the same numerical scale implies very different absolute asymptotes at 5% versus 30% diffusion, so both the scale and location of the prior are adjusted across bands.

Table 4.19: Hyperparameters of the diffusion-aware prior by forecast-origin band. All bands: $k_i \sim \Gamma(1, 0, 1)$, $\sigma_i \sim \Gamma(1, 0, 1)$. Here $\tilde{t}_T$ denotes the index of the last observation (first observation at $t = 0$).

Diffusion band	α_L	ℓ_L	s_L	α_{t_0}	ℓ_{t_0}	s_{t_0}
<2%	4	5	30	3	$\tilde{t}_T$	35
2–9%	2	1	40	2	$\max(5, \tilde{t}_T - 5)$	20
9–20%	2	1	25	1.5	$\max(5, \tilde{t}_T - 5)$	25
20–45%	2	1	5	1.5	$\tilde{t}_T/2$	15
>45%	1.5	1	3	1.5	$\tilde{t}_T/4$	20

Under the diffusion-aware prior, point-forecast accuracy improves substantially at a 5% forecast origin. The B-R MedALE is materially lower than under the universal prior across the diffusion range, with the largest gains at intermediate diffusion. Probabilistic calibration is also preserved. The PIT Q-Q plot remains well within the confidence bands, so the tighter diffusion-stage priors do not introduce overconfidence.

Implication. Across the three prior configurations, the ordering of forecast error is consistent. The wide prior is worst, the universal prior is in the middle, and the diffusion-aware prior is best. Our universal prior already achieves a substantial improvement over the wide baseline by exploiting two observable quantities. These are the series-specific growth rate $\hat{k}_{\text{ssr},i}$ and the last observed value $y_{T_{\max}}$. The diffusion-aware results show that analysts who can judge a technology's diffusion stage from the shape of the adoption curve can expect considerably better forecast quality. We therefore encourage practitioners to exploit available prior information about the technology when applying the B-R model.

4.17 Backtest on energy technologies

Primary energy demand increased by about 2% annually over 75 years, increasing useful energy generation in TWh despite constant or decreasing percentage shares for many technologies.

In this section, we compare backtesting performance of three approaches: the **B-**

R model (standard S-curve applied to TWh), the **growing-market B-R model** (S-curve with an exponentially growing asymptote, applied to TWh), and the **shares B-R model** (standard S-curve applied to market shares). The motivation is that many energy technologies have grown in absolute terms alongside total energy demand even after reaching maturity in share terms. Ignoring this can cause the fixed-asymptote model to dramatically underestimate future levels.

The three cases examined here are deliberately adversarial and should not be interpreted as representative of typical model performance. They are chosen precisely because they expose specific failure modes: (1) gas and hydropower reached share-saturation but continued growing in TWh as aggregate energy demand grew; (2) oil experienced a large, unpredictable production shock (a sharp swing in oil production around the 1979 peak) that altered its near-term trajectory; (3) nuclear energy saw rapid growth in the 1970s that was abruptly slowed by a sequence of major accidents and regulatory responses — not a *failed technology* (reactors continue to operate and new ones are being built), but a case in which the realised diffusion path deviated sharply from what data available at an early forecast origin would have predicted. Typical forecast performance across the full 120-technology ensemble is reported in Section 4.16 (and summarised in main paper Figure 3). The modeling choice underlying our main solar and wind forecasts (main paper Figure 4) is discussed in Section 4.19.

Backtests of useful energy generation with the standard B-R model use the universal prior with the model-direct adaptive widening calibrated in Section 4.16. This is the same method as our main-text forecasts. Backtests of market shares and of the growing-market model use no post-hoc widening.

4.17.1 Historical energy demand growth

To better understand the growth dynamics of energy technologies, we first examine historical trends in primary energy demand and useful energy generation for individual technologies. We calculate the growth in primary energy as log-differences, defined in Eq. 4.47. For convenience, we repeat the formula with new notation here,

$$g(t) = \frac{\log E(t) - \log E(t-h)}{h},$$

where h is the number of years between two consecutive data points, and $E(t)$ is the primary energy generated at time t.

Steady growth in primary energy consumption. Figure 4.36 shows the historical primary energy consumption in terawatt-hours (TWh) and the corresponding year-on-year increase in percentages, using data from [300], showing that primary energy has increased by approximately 2% per year for the past 100 years. Early growth hovers around 1%

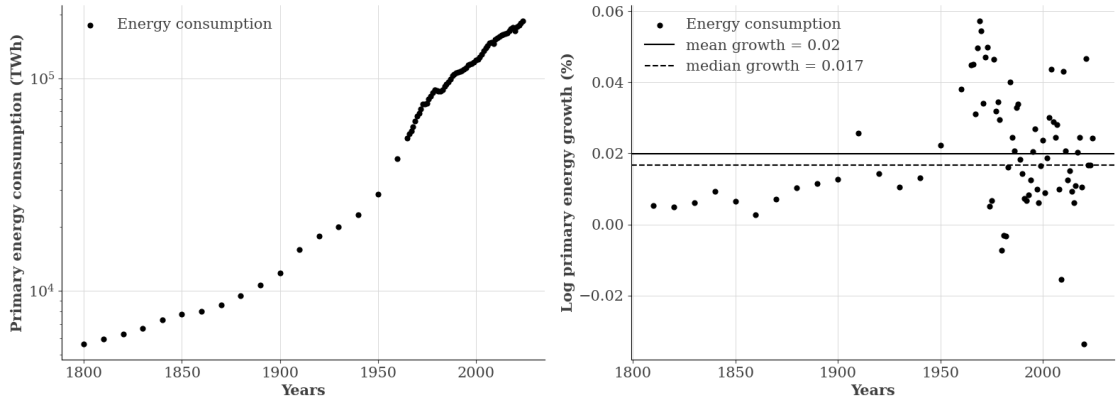


Figure 4.36: Historical primary energy consumption in TWh and the year-on-year growth rate. Primary energy has increased by approximately 2% per year for the past 100 years.

per year. From 1920 onwards, the average growth rate hovers around 2% per year, with some notable deviations. The period from 1960 to 1970 saw sustained growth above 2%, sometimes reaching 5%. The financial crisis in 2008 caused a temporary dip, and the COVID-19 pandemic in 2020 caused a significant drop — the only two years in the past

70 years where energy consumption decreased year-on-year. Despite these deviations, the long-term 2% trend has remained remarkably stable.

Growth in useful energy generation for individual technologies. Figure 4.37 shows the historical useful energy generation for selected energy technologies (in shares and TWh).

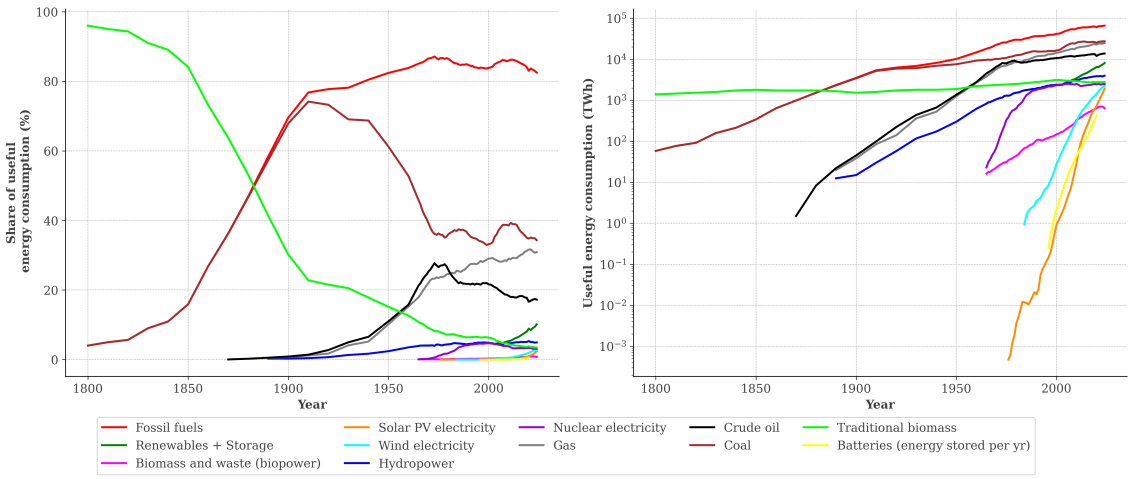


Figure 4.37: Historical useful energy generation for selected technologies in **(a)** shares and **(b)** TWh (in log scale). Useful energy generation for most fossil-fuel technologies (red) has increased in TWh over the past 20-30 years, while their share has remained approximately constant or even decreased. Examples shown are oil (black) and gas (gray).

Fossil fuels (sum of gas, oil, and coal, in red) have hovered around 80% of useful energy generation for the past 30 years, while their generation in TWh has steadily increased, evidenced by the constant line in panel (a) versus the red curve in panel (b). Oil (black) decreased in share while increasing in TWh of production. Gas (gray) increased in both.

Nuclear (purple) and traditional biomass (lime) are the only technologies shown that decreased in both share and total energy generation over the past decades. Nuclear fell slightly in share (from a plateau of $\approx$ 4.5% to $\approx$ 3%) and marginally in TWh [300]. Traditional biomass decreased significantly in share and marginally in TWh [300, 375].

4.17.2 Comparing backtest performance of three approaches

Here we present the backtest results using three different approaches. (1) Forecasting in shares using the shares B-R model. (2) Forecasting in TWh using the fixed-asymptote B-R model. (3) Forecasting in TWh using the growing-market B-R model. We start by introducing the growing market model.

4.17.2.1 Modeling market growth explicitly

To improve forecast performance in growing markets, we extend the B-R model by replacing the fixed asymptote L_i with a time-varying carrying capacity $L_i \exp(gt)$, where g is an exogenous annual growth rate. This is equivalent to assuming the market ceiling itself grows exponentially, a good approximation when aggregate energy demand grows at a roughly constant rate. Concretely, we model technology diffusion using the adjusted B-R curve

$$Y_{i,t} = \frac{L_i \exp(gt)}{\left(1 + \exp\left(-\beta k_i (t - t_{0,i}) + \beta \sigma_i (\rho \epsilon_{i,t-1} + \epsilon_{i,t})\right)\right)^{1/\beta}} \tag{4.63}$$

Estimating g from technology-specific time-series is difficult, especially for dominant technologies at early stages of diffusion where g and k are hard to separate. We set $g = 0.02$ based on Figure 4.36.

Limitations of the growing market assumption. In practice, as discussed in Section 4.11, technologies may decline rather than keep growing at the market rate after reaching maturity, so the 2% p.a. assumption does not hold uniformly. Oil useful energy consumption grew from approximately 8,500 TWh in 1983 to 13,500 TWh in 2022 (∼75% over 40 years, or ∼1.3% p.a.). Nuclear and traditional biomass actually *decreased*. Only hydropower and gas match the ∼2% p.a. assumption.

Energy substitution models capture these competitive dynamics [37, 259, 260, 376]: oil substituted wood during the early 20th century, and gas has substituted coal in heating

and electricity generation [377]. Despite these limitations, the 2% p.a. growing-asymptote assumption yields tractable Bayesian inference and reasonable backtest performance for technologies whose growth closely tracks aggregate energy demand.

The growing market S-curve model We use model (4.63) with $\beta = \frac{2}{3}$, $\rho = 0.8$, and $g = 0.02$ as universal parameters. To estimate the remaining parameters, we adjust the transformation function that linearizes the S-curve to account for market growth:

$$f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i}) = \frac{1}{\beta} \log \left(\frac{(L_i \exp(gt))^\beta}{Y_{i,t}^\beta} - 1 \right) + k_i(t - t_{0,i}).$$

The (t, t') element of the Jacobian matrix is

$$\left(\frac{\partial f(Y_{i,t})}{\partial Y_{i,t}} \right)_{t,t'} = \frac{\partial f(Y_{i,t})}{\partial Y_{i,t'}} = \delta_{t,t'} \frac{-(L_i \exp(gt))^\beta}{Y_{i,t}((L_i \exp(gt))^\beta - Y_{i,t}^\beta)} \tag{4.64}$$

and the log-determinant is

$$\log \det \left(\frac{\partial f(Y_{i,t})}{\partial Y_{i,t}} \right) = \sum_{t=1}^{T} \left[\beta g t + \beta \log(L_i) - \log \left(Y_{i,t}((L_i \exp(gt))^\beta - Y_{i,t}^\beta) \right) \right].$$

We then perform Bayesian inference as before using this adjusted transformation function and Jacobian.

Priors and assumptions. For models without asymptotic growth we require $L_i > \max_t(Y_{i,t})$ for $t \in [0, T]$, where T is the last time-point in the training data. As in Section 4.7, we place priors on scaled parameters $\tilde{L}_i = L_i / \max_t(Y_{i,t})$ and $\tilde{t}_{0,i} = t_{0,i} - t_1$, where t_1 is the first time-point in the training data. Figures 4.38b and 4.38e use the fixed-asymptote B-R model with the same universal prior as in the main analysis (Section 4.7).

In the growing-market model (4.63), the asymptote increases over time, so we require $L_i \exp(gt) > Y_{i,t}$ rather than $L_i > \max_t(Y_{i,t})$. In particular, $\tilde{L}_i$ can be less than 1 because

market growth compensates. We therefore adjust the priors for $\tilde{L}_i$ and $\tilde{t}_{0,i}$, while keeping k_i and σ_i unchanged. Furthermore, we set a smaller lower bound on $\tilde{t}_{0,i}$ to ensure that the inflection point is not too late, which would lead to implausible forecasts. We adjust the scales of $\tilde{L}_i$ and $\tilde{t}_{0,i}$ (and the shape of $\tilde{t}_{0,i}$) so that the prior predictive distribution remains reasonable. The exogenous growth assumption changes the relationship between these parameters and the data, particularly for L.

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2, 0, 5) & \text{(lower bound: } \tilde{L}_i \geq 0) \\ \log k_i \sim \log \Gamma(1, 0, 1) & \text{(unchanged)} \\ \log \tilde{t}_{0,i} \sim \log \Gamma(2, \max(5,\, T_{\max} - 6/\hat{k}_{\text{ssr},i}),\, 15) & \text{(lower bound: tech-specific)} \\ \log \sigma_i \sim \log \Gamma(1, 0, 1) & \text{(unchanged)} \end{cases} \tag{4.65}$$

The prior for $\tilde{L}_i$ now has a lower bound of 0 rather than 1.

For shares, L_i is bounded between 0 and 1. Hence, we keep all priors unchanged except for L_i, for which we use a uniform distribution between 0 and 1 as a non-informative prior. That is, the prior for shares is

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \mathcal{U}(0, 1), \\ t_{0,i} \sim \Gamma\big(3,\, \max(5,\, T_{\max} - 3/\hat{k}_{\text{ssr},i}),\, 30\big), \\ k_i \sim \Gamma(1,\, 0,\, 1), \\ \sigma_i \sim \Gamma(1,\, 0,\, 1). \end{cases} \tag{4.66}$$

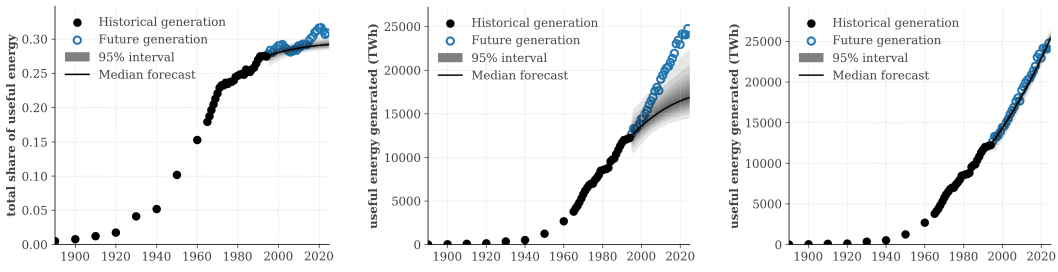
Prior adjustment for near-saturated shares. For the shares B-R model applied to gas and hydropower, both technologies are already near or at saturation in shares at the $T - 30$ training–test split. In this regime, the data-informed lower bound $\ell_{t_{0,i}} = \max(5,\, T_{\max} - 3/\hat{k}_{\text{ssr},i})$ can inadvertently push the prior support for $t_{0,i}$ to lie entirely *after* the last observed data point, resulting in degenerate posteriors. We therefore set a smaller

offset ($2/\hat{k}_{\text{ssr},i}$ instead of $3/\hat{k}_{\text{ssr},i}$) for these two cases, ensuring the prior places the inflection point within the observed data range.

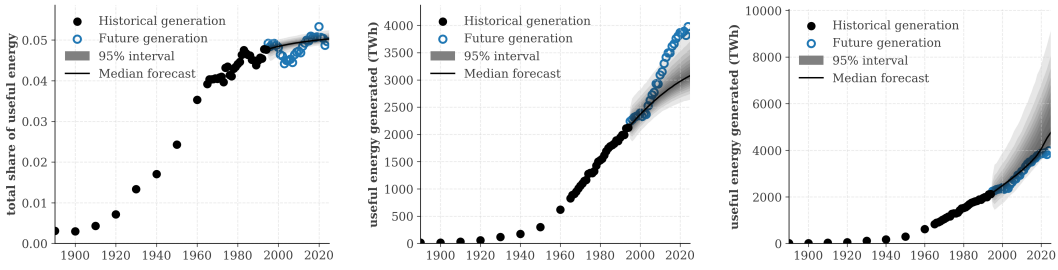
4.17.2.2 Comparing the three approaches.

We show results for two technologies where we expect the growing market model to work well, hydropower and gas. We then discuss the intrinsically more difficult example of oil.

Technologies that grow asymptotically with the market: Hydropower and gas. Figure 4.38 compares backtest results for gas and hydropower using each of the three methods for a 30-year forecast horizon. In the test period, both technologies' useful energy shares change only modestly (gas increases slightly; hydropower is flat), so the shares model is a reasonable approximation. Results would naturally vary under different training–test splits or prior configurations. Figure 4.38a shows that the shares forecast produces



(a) Gas backtest (shares) **(b)** Gas backtest (TWh, widened) **(c)** Gas backtest (TWh, growing S-curve)



(d) Hydropower backtest (shares) **(e)** Hydropower backtest (TWh, widened) **(f)** Hydropower backtest (TWh, growing S-curve)

Figure 4.38: Example backtest comparison for three forecasting methods on gas and hydropower useful energy generation shares and absolute generation. **(a-c)** Backtest on gas shares, useful energy with no growth assumptions, and useful energy with growth assumptions. **(d-f)** Backtest on hydropower shares, useful energy with no growth assumptions, and useful energy with growth assumptions.

reasonable results for this training–test split. The posterior places both technologies near saturation in shares (the posterior median of $\tilde{L}_i$ is close to the last observed value), which is consistent with the data. The forecasts appear slightly underdispersed: realized data points occasionally fall outside the 95% credible interval, likely because the B-R model's noise variance shrinks to zero as diffusion approaches the asymptote, understating residual uncertainty near maturity.

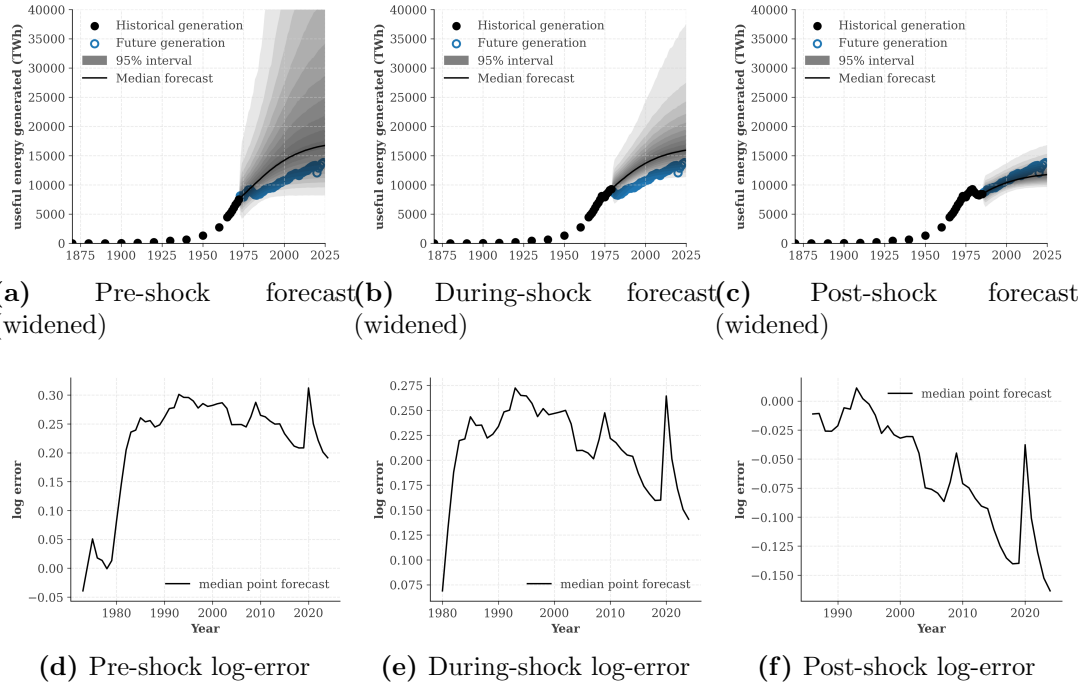
Figures 4.38b and 4.38e show that the fixed-asymptote B-R model applied to TWh is overconfident even with adaptive widening. Realized generation rises above the prediction interval because gas and hydropower keep growing with the market. Their useful energy shares stay flat while their absolute generation grows with overall demand, so generation exceeds the model's fixed ceiling. This is a structural failure of the model.

Figures 4.38c and 4.38f show that the growing market S-curve model captures future data points well, with the median forecast close to realized values. Both technologies grew at approximately 2% p.a. in this period, so the model is expected to perform well. Note that adjusting the priors for L_i and $t_{0,i}$ to account for the growing market size can lead to better performance.

These results inform our approach for solar and wind. For mature technologies with flat shares (gas, hydropower), the fixed-asymptote model fails in TWh because the growing market dominates. For solar and wind, however, we are in the opposite regime: both technologies are at early diffusion stages where total energy generation is still far below any plausible asymptote. In this regime, the fixed-asymptote model remains valid because the S-curve dynamics (not market growth) drive the forecast, and we use it for our main solar and wind projections in Section 4.19.

Predicting through an oil production shock. Figure 4.39 shows the distributional forecast and log-error for oil useful energy generation, using the fixed-asymptote B-R model,

when training before (1870–1972), during (1870–1979), and after the oil production shock (1870–1985). The figure illustrates how training-period timing relative to structural



(a) Pre-shock forecast (widened) **(b)** During-shock forecast (widened) **(c)** Post-shock forecast (widened)

(d) Pre-shock log-error **(e)** During-shock log-error **(f)** Post-shock log-error

Figure 4.39: Backtest for oil useful energy generation using the fixed-asymptote B-R model, with distributional forecasts and log-errors for pre-shock, during-shock, and post-shock training sets. **(a)** Pre-shock forecast captures future deviations well. **(b)** During-shock forecast. Most data points fall within the lower part of the prediction interval, with a few below it. **(c)** Post-shock forecast captures realized data points, which lie in the upper part of the distribution. **(d–f)** Corresponding log-errors. The median forecast for all three scenarios is reasonable, with log-errors generally within ±25%.

breaks affects forecast accuracy. Training pre-shock (1870–1972) or post-shock (1870–1985) yields reasonable performance: the pre-shock forecast captures future deviations within 95% prediction intervals, while the post-shock forecast achieves log-errors below 15% (in value) with all points inside prediction intervals. Yet post-shock forecasts systematically underestimate future data points. All points lie in the upper part of the distribution, possibly because the energy market is growing.

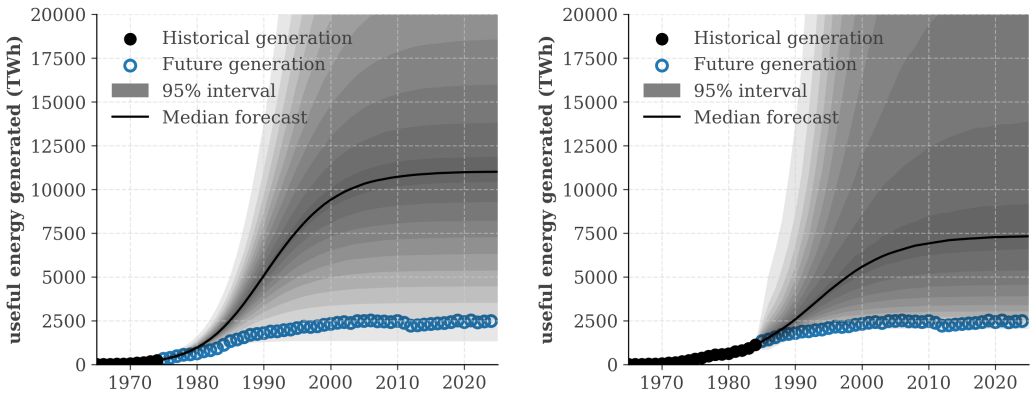
Training through the oil production shock (1870–1979) produces systematic overestimation in the median, with log-errors increasing to ∼20% as the model fails to capture subsequent drops. The general pattern is robust across prior configurations. Even in this

adversarial case the realized data remain within the prediction interval. The structural break is a low-probability event under the model, not an impossible one. A purely data-driven model has no mechanism to anticipate such production shocks, and the adaptive widening assigns them non-negligible probability.

4.17.3 Exogenous disruptions: nuclear energy

Nuclear energy is an instructive case of a technology whose diffusion trajectory was repeatedly disrupted by exogenous safety events. Unlike the technologies in our main cross-section dataset, nuclear has experienced multiple abrupt deceleration episodes and is therefore a stress-test for any smooth S-curve model.

Backtest results. Figure 4.40 shows backtests using the fixed-asymptote B-R model with forecast origins in 1974 (before the major growth decline) and 1984 (after the Three Mile Island accident), with the same universal prior used for solar backtests in Section 4.18. Current nuclear useful energy generation is approximately 2,500 TWh. From the 1974



(a) 50-year forecast horizon, origin 1974 (widened)

(b) 40-year forecast horizon, origin 1984 (widened)

Figure 4.40: Backtest for nuclear energy useful energy generation using the fixed-asymptote B-R model. **(a)** 50-year horizon from 1974 (before the major accidents): the model overestimates by ∼4×, but realized data still lie within the tails of the prediction interval. **(b)** 40-year horizon from 1984 (after Three Mile Island; pre-Chernobyl): the post-TMI deceleration is already visible in the training data, and the model captures the trend reasonably well.

forecast origin (panel a), the model predicts a 2024 median of ∼11,000 TWh — over 4× the realized value — yet realized data still lie within the lower tails of the 95% prediction interval. This is expected: a data-driven model has no mechanism to anticipate the sequence of accidents and regulatory changes that followed, but the wide prior on L ensures the structural break remains within the distribution's support.

By 1984, the data already show a clear deceleration following Three Mile Island, and the model is able to reasonably estimate the trend going forward (though it cannot anticipate Chernobyl in 1986 or Fukushima in 2011).

Historical context. Nuclear power was promoted as a peaceful use of atomic energy after WWII [378], and in the 1970s was projected to become a dominant energy source [379]. Today it accounts for approximately 3% of useful energy consumption. Figure 4.41 shows the historical useful energy share and log-growth. Three major accidents (Three

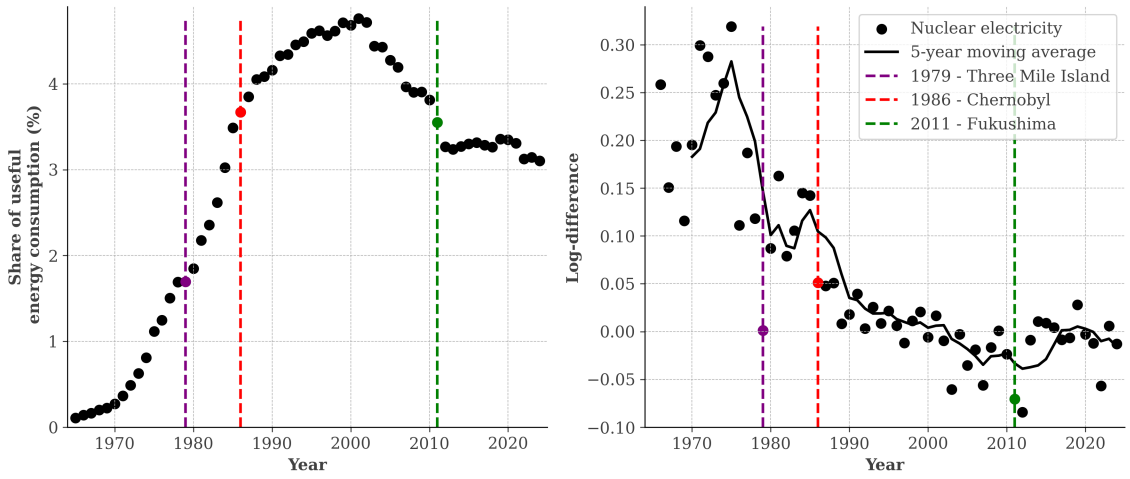


Figure 4.41: Useful energy consumption share and log-growth of nuclear electricity. Growth was approximately 30% p.a. during the rapid expansion phase. After Three Mile Island (1979), growth decreased; after Chernobyl (1986), log-growth fell from ∼15% to ∼5%; after Fukushima (2011), log-growth turned negative before recovering to ≈ 2% p.a.

Mile Island (1979) [380], Chernobyl (1986), and Fukushima (2011)) each produced sharp growth declines visible in Figure 4.41. No new US reactor has been ordered since 1978 [381], and public support shifted predominantly negative after these events [382, 383]. Nuclear electricity reached an approximate plateau around 1995 at ≈ 4.5% of useful energy

consumption; since 2011, log-growth has hovered near −2%. These repeated disruptions explain the forecast failures above: each accident reset the diffusion trajectory in a way that a smooth S-curve cannot capture.

The nuclear case illustrates a general limitation. The S-curve model cannot always capture sudden structural breaks in adoption driven by exogenous events. Such disruptions are rare across the technology ensemble, although this is partly due to selection bias.

4.18 Backtesting solar and wind time-series

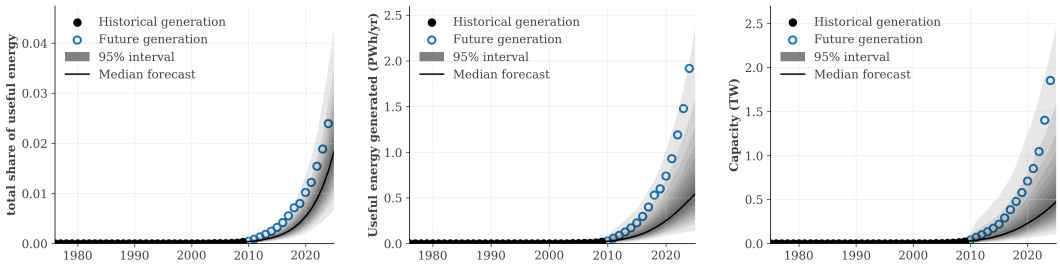
To confirm the robustness of our forecast results on solar PV and wind (main paper Figure 4), we backtest on historical data for a 15-year forecast horizon. This horizon corresponds to energy-system planning time-scales [198, 250] and matches the forecast horizons used in Section 4.19.

Before interpreting the backtest results, two features of solar PV and wind are important context. Both technologies have followed near-exponential growth through 2024, with no sustained deviation from trend (Section 4.19, Figure 4.48). When training from a 2010 forecast origin, nearly all historical data therefore lies in the exponential regime. This provides limited information about eventual saturation and makes the prior more influential than in the true out-of-sample exercise. The 2010 origin is also particularly early for solar: solar PV represented only 0.04% of global useful energy in 2010, compared to 2.4% in 2024 (the forecast origin of the main-text predictions). The backtest thus constitutes a conservative stress test, and the full out-of-sample forecasts from 2024 use 14 additional years of data with a considerably more informative posterior for L.

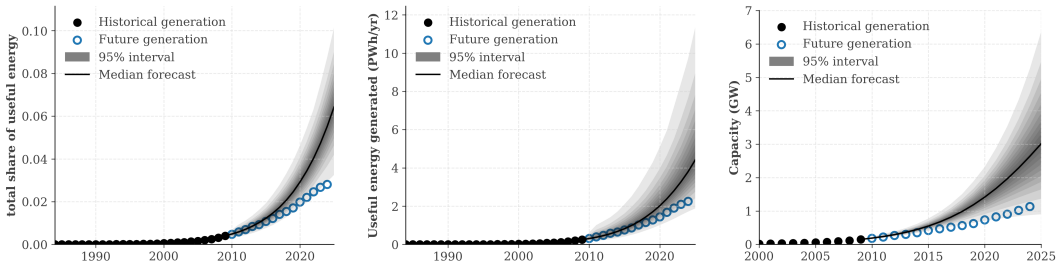
Backtests of useful energy generation and capacity use the universal prior with the model-direct adaptive widening calibrated in Section 4.16. This is the same method as our main-text forecasts. Backtests of market shares and of the asymptotic-growth model use

no post-hoc widening.

Figure 4.42 shows the median and 95% prediction interval for solar and wind useful energy generation, shares, and capacity. We marginally underpredict solar in the median and overpredict wind, except for wind useful energy generation. Almost all realizations fall within the 95% prediction interval. For solar useful energy generation, all realizations fall within the interval, but they lie systematically in its upper part, and the 2024 point is only just inside. This reflects the median underprediction of solar growth. The remaining exceptions are the 2024 wind useful-energy-share point below the interval and wind capacity below it in 2017 to 2019.



(a) Solar useful energy share **(b)** Solar useful energy generation (widened) **(c)** Solar capacity (widened)

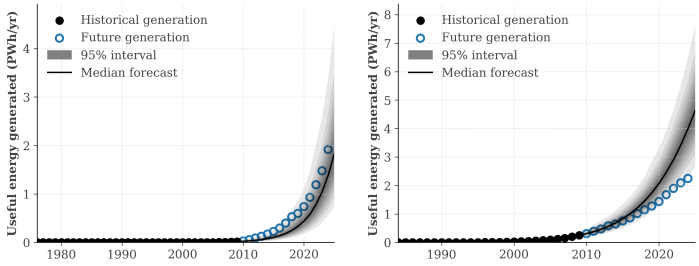


(d) Wind useful energy share **(e)** Wind useful energy generation (widened) **(f)** Wind capacity (widened)

Figure 4.42: Backtest results for solar and wind useful energy generation, shares, and capacity. Black dots are historical data. Blue circles are future realizations relative to the forecast origin. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. Panels labelled (widened) use the adaptive widening of Section 4.16. **(a)**–**(c)** show backtests for solar PV, and **(d)**–**(f)** show backtests for wind.

Backtest using the S-curve with asymptotic growth. Figure 4.43 shows the same backtest for solar PV and wind useful energy for a 15-year horizon, but using the

S-curve model with asymptotic market growth. If technologies are far from saturation, the forecast should be similar under both models. The results are indeed almost the same. When diffusion is far from saturation, $Y(t)/[L\exp(gt)] \ll 1$, so the S-curve trajectory is governed mostly by the exponential-growth phase of the B-R model. In this regime, the additional $\exp(gt)$ factor plays no large role in the near-term forecasted trajectory. The two model variants diverge materially only as the technology approaches its effective market ceiling.

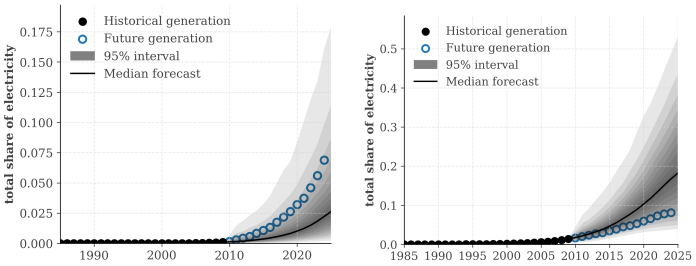


(a) Solar useful energy generation

(b) Wind useful energy generation

Figure 4.43: Backtest results for solar and wind useful energy generation using the asymptotic growth model. Black dots are historical data. Blue circles are future realizations. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. **(a)** shows the backtest for solar PV useful energy, and **(b)** shows the backtest for wind useful energy.

Additional solar and wind time-series Solar PV and wind time-series are also commonly reported as shares of total electricity generation. Figure 4.44 presents the backtest for electricity shares using data from OWID [210], providing an independent robustness check that the model generalises to this alternative diffusion measure. Prediction intervals are wider here because the electricity-shares series begin in 1985 rather than 1975, shortening the training window and reducing the information available about early diffusion dynamics. In-sample noise is also higher relative to the useful energy series, further contributing to the wider prediction intervals.



(a) Solar electricity share **(b)** Wind electricity share

Figure 4.44: Backtest results for solar PV and wind electricity shares. Black dots are historical data. Blue circles are future realizations. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. **(a)** shows the backtest for solar PV electricity share, and **(b)** shows the backtest for wind electricity share.

4.19 Forecasting solar and wind useful energy

We present distributional forecasts for solar PV and wind useful energy generation, shares, and capacity using the Bertalanffy-Richards curve with technology-invariant parameters $\rho = 0.80$, $\beta = 2/3$, $\epsilon_{i,T} = 0$. The historical data sources are described in Section 4.11, with records through 2024. Farmer and Lafond [30] provide the method for the exponential-growth forecasts. We compare our forecasts to IEA NZE [296], IPCC AR6 [234], and Way et al. [29].[29]

Throughout this section we keep two conceptually distinct issues separate. First, *installed capacity* (in TW) measures the maximum potential output of a technology, whereas the *energy actually delivered* depends additionally on the capacity factor and downstream conversion losses. Second, we distinguish *secondary energy* (raw electricity generated), *final energy* (electricity delivered to end-users, after transmission and distribution losses), and *useful energy* (the energy content of services provided, after end-use losses). We prefer useful energy because it best represents the energy services delivered by solar and wind and enables consistent comparison via shares. Where the IEA reports milestones in capacity (TW), we also forecast capacity directly. Section 4.19.2 below details the capacity-to-useful-energy conversion and shows that fitting capacity and converting yields

[29] See [66, 69, 113, 384, 385] for alternative approaches.

similar forecast distributions to fitting useful energy directly.

4.19.1 IEA net-zero milestones for solar capacity

The IEA publishes Net-Zero Emissions (NZE) 2050 milestones for solar PV installed capacity [296, 297]: deployment targets at 2030, 2035, 2040, and 2050 consistent with limiting warming to 1.5°C. We address two questions. First, given an assumed asymptote L, when does the B-R model predict each milestone is reached, and how does that change with L? Second, what does our distributional forecast for solar capacity imply for whether the milestones will be met?

4.19.1.1 Conditional milestone-year analysis

L is hard to identify from early data, and experts may disagree on its plausible range. We complement the distributional forecast with a conditional exercise. We treat L as a known input, fit $(\hat{k}, \hat{t}_0)$ by OLS on the GFP-linearised model (Eq. (4.16)), and compute the year at which diffusion reaches each IEA milestone as a function of L.

To estimate the year t^* such that diffusion $Y(t^*)$ equals milestone M_{t^*}, we invert the B-R model (Eq. (4.16)) to solve for t^*,

$$t^*(L, M_{t^*}) = \hat{t}_0 - \frac{1}{\beta\hat{k}} \log\left(\left(\frac{L}{M_{t^*}}\right)^{\beta} - 1\right) + \frac{\eta_{t^*}}{\beta\hat{k}}, \tag{4.67}$$

where $\eta_{t^*} = \hat{\sigma}(\epsilon_{t^*} + \rho\epsilon_{t^*-1})$. The noise term is additive in the time domain, so $\pm 1\sigma$ error bands are

$$t^*(L, M_{t^*}) = \hat{t}_0 - \frac{1}{\beta\hat{k}} \log\left(\left(\frac{L}{M_{t^*}}\right)^{\beta} - 1\right) \pm \frac{\sigma_\eta}{\beta\hat{k}}, \tag{4.68}$$

where $\sigma_\eta = \hat{\sigma}\sqrt{1+\rho^2}$. These OLS bands reflect only the stochastic noise η_{t^*} and do not account for posterior uncertainty in $\hat{k}$, $\hat{t}_0$, or $\hat{\sigma}$. They therefore understate the total uncertainty. We also perform a full Bayesian fit with the same universal priors as the

unconditional forecasts, treating L as fixed. Figure 4.45 compares the OLS, the Bayesian, and the Bayesian widened 68% bands.

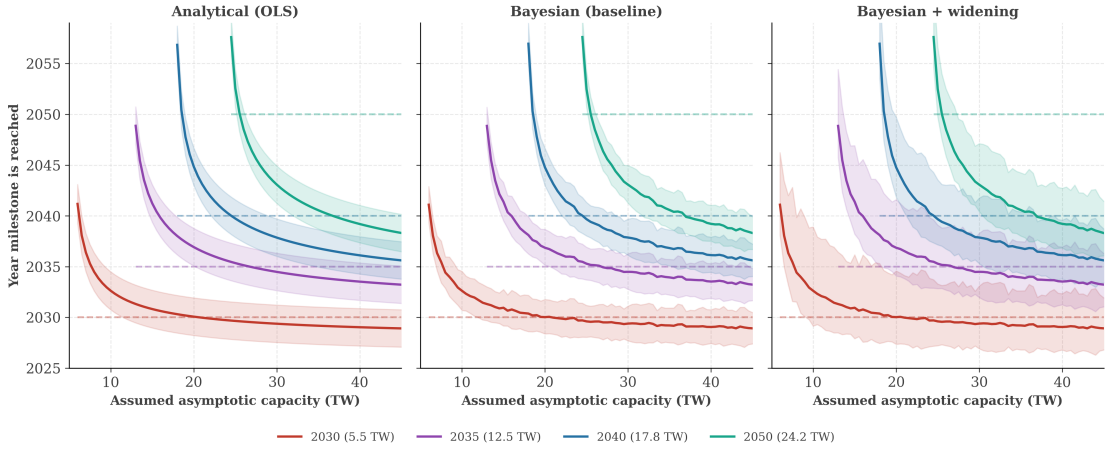


Figure 4.45: Year at which each IEA NZE milestone is reached as a function of the assumed asymptote L. **(a):** OLS analytical median with $\pm 1\sigma$ bands as per main paper Figure 4(b). **(b)** Bayesian posterior keeping L fixed, with posterior uncertainty in $\hat{k}$, $\hat{t}_0$, and $\hat{\sigma}$. **(c)** The same Bayesian posterior after the adaptive widening transform of Section 4.7.5. Dashed horizontal lines mark IEA target years. The widening is calibrated on the unconditional 120-technology forecasts, so panel (c) is an approximation in this fixed-L case.

The two methods agree closely on the median. The OLS and Bayesian median milestone years differ by less than 0.1 year at every L for the IEA solar capacity milestones. The widening transform preserves the median by construction, so panel (c) shares the median of panel (b). The year each milestone is reached is therefore robust at the median. It does not depend on the method or on the widening, and only the width of the bands changes. This is what matters for the application to solar PV below, which reports median years (Table 4.21).

Validation on the 120-technology dataset. We test the conditional method on the 120 technologies of Section 4.11. For each technology we take the asymptote L from the full-sample OLS fit on the GFP-linearised B-R model with free β (the same fit used for prior calibration elsewhere in the pipeline). Diffusion is defined as $d = Y/L$, and milestone M corresponds to $Y \geq M \cdot L$. For each series we then fit $(\hat{k}, \hat{t}_0)$ by OLS on the

GFP-linearised model with fixed $\beta = 2/3$ using observations with $d < d_{\text{fit}}$. We predict the year diffusion reaches each milestone fraction M and compare to the realised year.

Table 4.20 reports the median prediction error (predicted minus realised milestone year) at two values of d_{fit} that bracket solar PV's 2024 situation. The lower diffusion ($d_{\text{fit}} = 2.5\%$) is close to solar PV's posterior-median current diffusion ($\approx 2.2\%$). The higher diffusion ($d_{\text{fit}} = 10\%$) matches the diffusion implied by the IEA NZE 2050 asymptote $L = 26$ TW. At the lower diffusion only 56 of the 120 historical analogues have enough early data to fit the method. Median errors grow from +4.7 years at $M = 0.25$ to +12.1 years at $M = 0.90$. At the higher diffusion the sample is larger ($N = 86$) and errors are smaller, from +1.1 years at $M = 0.25$ to +6.0 years at $M = 0.90$.

Table 4.20: Median prediction error (predicted minus realised milestone year, in years) on the 120-technology dataset. We take each technologyś asymptote L from the full-sample free-β OLS fit on the GFP-linearised B-R model. Diffusion is $d = Y/L$. We fit $(\hat{k}, \hat{t}_0)$ by OLS with fixed $\beta = 2/3$ on observations with $d < d_{\text{fit}}$, predict the year diffusion reaches each milestone fraction M, and compare to the realised year. N is the number of technologies with a successful fit and a realised milestone. The two rows bracket solar PVś situation in 2024: $d_{\text{fit}} = 2.5\%$ approximates our posterior-median current diffusion ($\approx 2.2\%$), while $d_{\text{fit}} = 10\%$ reflects the diffusion implied by the IEA NZE 2050 asymptote $L = 26$ TW ($2.2/26 \approx 8.5\%$).

d_{fit}	N	M=0.25	M=0.50	M=0.75	M=0.90
2.5%	56	+4.7	+7.9	+12.0	+12.1
10%	86	+1.1	+1.6	+3.3	+6.0

How much can we trust the method? The signed median error is positive in every cell of Table 4.20. That is, the conditional method predicts milestones *later* than technologies realised them. The milestone-year predictions of Figure 4.45 are therefore conservative. Under the IEA NZE asymptote assumption ($L = 26$ TW, $d_{\text{fit}} = 10\%$), each IEA capacity target divided by L gives its milestone fraction M: 0.21 for 2030, 0.48 for 2035, 0.68 for 2040, and 0.93 for 2050. The corresponding median biases are approximately +1, +2, +3, and +6 years. Under our posterior on L, where solar's current diffusion is about 2.2%, the same milestones lie at smaller M but require a much longer extrapolation,

and the corresponding median biases are approximately +5, +8, +12, and +12 years.

The median is robust, but are the prediction bands the right width? The widening transform was calibrated on the unconditional forecasts, so applying it to these fixed-L bands is an approximation. We test it by backtesting coverage on the 120 technologies, one milestone at a time. Take the 50% milestone, the year a technology first reaches half of its asymptote. For each technology the fixed-L Bayesian fit gives a predictive distribution over that year, and we know the year it was actually reached. The 68% coverage is the fraction of technologies whose actual year falls inside the central 68% of their own predictive distribution. A well-calibrated band has coverage equal to its nominal level. We compute this for each milestone (25%, 50%, 75%, 90%) and for several values of d_{fit}, the diffusion up to which we fit each technology. Without widening the bands are too narrow, just as in the unconditional forecasts. Widening moves coverage closer to nominal. Fitting up to 10% diffusion, for example, raises the 68% coverage of the 50% milestone from 33% to 55%, and its 90% coverage from 45% to 76%. Coverage is better when we fit up to a larger diffusion, where more data is available.

Application to solar PV. Table 4.21 reports the year reached for each milestone at three representative asymptotes, with Bayesian widened 68% bounds. At $L = 26$ TW the model meets the 2030 milestone (5.5 TW) and the 2035 milestone (12.5 TW) in their target years, and reaches the 2040 and 2050 milestones roughly one year early. At $L = 40$ TW the model reaches the final 2050 milestone (24.2 TW) in 2039, approximately a decade ahead of target.

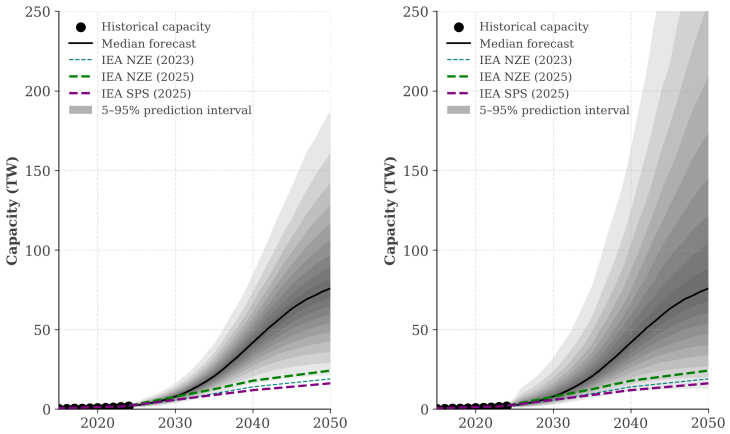
4.19.1.2 Distributional forecast for solar PV capacity

The conditional analysis above assumes L is known. Figure 4.46 shows the distributional forecast for solar PV capacity before and after the widening transform of Section 4.7.5. The 24.2 TW by 2050 IEA NZE target lies below the 15th percentile of the widened forecast,

Table 4.21: Estimated year of reaching IEA net-zero-by-2050 milestones for solar PV capacity, conditional on asymptote L. The median year t^* is computed analytically from the Bertalanffy-Richards curve with $\beta = \frac{2}{3}$. The lo/hi columns are the 68% band after applying the adaptive widening transform of Section 4.7.5 (Eq. 4.36), with w^* looked up at $d_{\text{eff}} = Y_{2024}/L$ from the same anchors as the unconditional capacity forecasts. A fixed-L Bayesian fit with the universal priors of Section 4.7 yields a posterior median for t^* within 0.1 year of the OLS value for every L tabulated. A dash indicates the milestone exceeds the assumed asymptote.

L (TW)	IEA milestone	Median year	68% lo	68% hi	Years early/late
10	2030 (5.5 TW)	2033	2029	2039	3 late
10	2035 (12.5 TW)	–	–	–	–
10	2040 (17.8 TW)	–	–	–	–
10	2050 (24.2 TW)	–	–	–	–
26	2030 (5.5 TW)	2030	2027	2034	on time
26	2035 (12.5 TW)	2035	2032	2039	on time
26	2040 (17.8 TW)	2039	2036	2043	1 early
26	2050 (24.2 TW)	2049	2046	2053	1 early
40	2030 (5.5 TW)	2029	2026	2033	1 early
40	2035 (12.5 TW)	2034	2031	2037	1 early
40	2040 (17.8 TW)	2036	2034	2040	4 early
40	2050 (24.2 TW)	2039	2037	2043	11 early

indicating that even under pessimistic diffusion scenarios solar PV is likely to meet all IEA net-zero capacity milestones. The unconditional median reaches approximately 75 TW in 2050, well above the highest IEA target.



(a) Solar PV capacity forecast (unwidened).

(b) Solar PV capacity forecast (widened).

Figure 4.46: (a) Distributional B-R forecast for solar PV installed capacity under the universal prior, unwidened. **(b)** Same forecast after the adaptive widening transform of Section 4.7.5. Dashed lines show the IEA NZE and Stated Policies scenarios from the 2023 [296] and 2025 [297] editions. The most demanding IEA milestone (24.2 TW by 2050) lies below the 15th percentile of the widened forecast.

4.19.2 From capacity to useful energy

Solar useful energy forecasts can be built from two alternative data sources: (i) **useful energy (direct):** historical useful energy actually delivered (in PWh/yr) from Way et al. [29]; or (ii) **installed capacity (converted):** historical installed capacity (in TW) from IRENA [298], converted to useful energy via

$$E_{\text{useful}} = C \times 8760 \times \text{CF} \times 0.87 \times 0.9, \tag{4.69}$$

where C is installed capacity in TW and 8760 is the number of hours per year. The capacity factor CF (typically 0.10–0.25 for solar PV globally [386]) converts installed capacity to secondary energy (electricity generated). The factor 0.87 is the median secondary-to-final energy conversion ratio computed from 1,167 vetted IPCC AR6 scenarios [234], accounting for transmission and distribution losses estimated at 4–15% globally [387]. The factor 0.9 converts from final to useful energy, following Way et al. [29] and accounting for end-use losses [388].

Using CF = 0.17 (the historical median global value [386]), 76 TW (approximately the median solar capacity forecast in 2050) yields ≈ 88 PWh of useful energy per year via Eq. (4.69), broadly consistent with the median predicted useful energy of ≈ 80 PWh in main text Figure 4 in 2050. The conversion is sensitive to the assumed capacity factor. A slightly lower value within the historical range (CF ≈ 0.15) gives a capacity-route median of ≈ 78 PWh, almost identical to the direct estimate. Even at CF = 0.17, the two routes produce very similar distributional forecasts, as quantified below (Fig. 4.47). The unwidened 5^{th}–95^{th} percentile capacity forecast at 2050 is ≈ 21–187 TW, translating to ≈ 25–219 PWh of useful energy and consistent with the corresponding unwidened useful-energy direct interval. After applying the adaptive widening transform of Section 4.7.5, these intervals widen to ≈ 13–430 TW (≈ 15–500 PWh) for the capacity route and ≈ 13–470 PWh for the useful-energy direct route. The widening is applied identically to both routes, so the similarity between them is preserved (panel c of Fig. 4.47 below). The widened lower bound is conservative since the asymmetric transform widens the upper tail more strongly than the lower (Section 4.7.5).

To confirm that the choice of data source does not materially affect the forecast, we compare the two distributional forecasts at each year using the two-sample KS statistic (Eq. (4.62)) with $F_1 = \hat{F}_{\text{useful}}$ and $F_2 = \hat{F}_{\text{cap}}$, the empirical CDFs of the useful-energy and capacity-converted forecast samples respectively, with x in PWh/yr, and the overlap coefficient

$$\mathrm{O} = \int \min\big(f_{\text{useful}}(x),\ f_{\text{cap}}(x)\big)\, dx, \tag{4.70}$$

where $D = 0$ (resp. O = 1) indicates identical distributions.

Figure 4.47 presents the comparison in three panels. On a log scale (panel a), the two historical series show a slight divergence up to around 2000, after which they align closely through 2024, confirming that the constant-CF conversion is a reasonable approximation

over the observed period. The early divergence may reflect a capacity factor that differed from the fixed value of 0.17 during the early phase of solar PV deployment. The widened forecast distributions at 2030, 2040, and 2050 are broadly overlapping (panel b), with comparable medians and interquartile ranges across the forecast horizon, though the useful-energy (direectly predicted) median is marginally lower in 2050 uder a 0.17 capacity factor. The KS D statistic remains below 0.10 and the overlap above 0.9 throughout (panel c). The same conclusion holds for the unwidened forecasts. We conclude that forecasting capacity and converting to useful energy, or forecasting useful energy directly, generates similar distributional forecasts for solar PV.

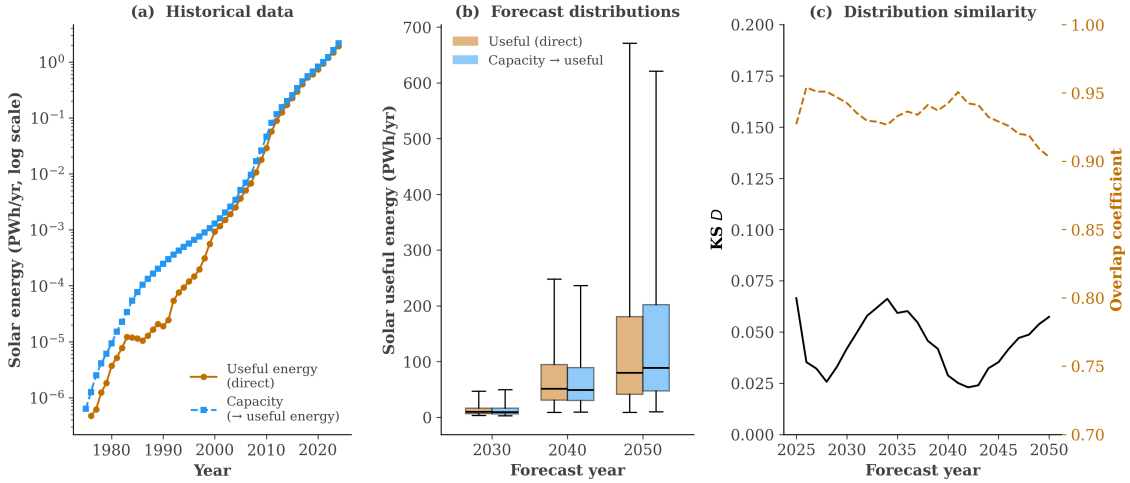


Figure 4.47: **(a)** Historical useful energy (orange) and capacity converted to useful energy via Eq. (4.69) (blue), on a log scale. The two series show a slight divergence up to around 2000, after which they align closely. **(b)** Side-by-side boxplots of the forecast distributions at 2030, 2040, and 2050, both widened via the adaptive sqrt-lower transform of Section 4.7.5 (box = IQR, whiskers = 1.5× IQR). **(c)** Kolmogorov–Smirnov D statistic (black, left axis) and overlap coefficient (orange dashed, right axis) from 2025 to 2050.

4.19.3 Forecasts for useful energy

4.19.3.1 Setup

We now turn to distributional forecasts for useful energy generation and useful energy shares.

Converting between the energy-accounting conventions defined above (secondary, final, useful) requires two steps. Sadovskaia et al. [387] estimate global transmission and distribution losses at 4–15% (secondary to final energy). Way et al. [29] assume end-use losses of 10% for solar and wind electricity (final to useful energy). We apply the two factors sequentially.

We compute the median secondary-to-final energy ratio across vetted IPCC AR6 scenarios [234] for 2030, 2040, and 2050, obtaining a conversion factor of 0.87. We then multiply by 0.9 to convert from final to useful energy, following Way et al. [29] and accounting for end-use losses [388]. To compare with IEA electricity generation scenarios, we multiply the reported final energy values by 0.9 to obtain useful energy estimates [296, 297]. The derivation of these conversion factors and a validation that this approach yields similar forecast distributions to fitting useful energy directly are provided in Section 4.19.2.

Before comparing our forecasts to scenarios, we examine how our model can predict departure from exponential behavior even though, for early solar PV and wind data, observed growth is close to exponential.

4.19.3.2 Analyzing the growth of solar and wind

This subsection examines how our B-R forecasts incorporate information about when solar and wind will decelerate from their current near-exponential growth. We establish formally that the near-exponential growth observed so far is exactly what the B-R model predicts for early-stage technologies. We then identify the point at which the median B-R forecast diverges from a pure exponential extrapolation. This is the earliest horizon at which the S-curve dynamics become detectable in the data.

So far, solar PV and wind have exhibited growth patterns very close to exponential [29–31]. The B-R model predicts this. For a technology in its early growth phase ($t \ll t_0$)

we have $\exp(-\beta k(t - t_0)) \gg 1$, so the B-R model reduces to

$$y_t \equiv \log Y_t \approx (\log L - kt_0) + kt - \sigma\big(\epsilon_t + \rho\epsilon_{t-1}\big), \qquad \epsilon_t \sim \mathcal{N}(0,1), \tag{4.71}$$

i.e. a deterministic linear trend at rate k with MA(1) noise. Our ability to predict future deployment hinges upon deviations from this exponential trend. Following Farmer and Lafond [30], we analyze when the median B-R forecast departs significantly from such an exponential null model.

The deterministic trend implied by the model is $y_1 + \hat{\mu}\,(t - t_1)$, where $y_1 = \log Y(t_1)$ is the log of the first observation and $\hat{\mu} = \overline{\Delta y}$ is the sample mean of the log-differences. The OLS residuals from this trend are $\hat{r}_i = y(t_i) - \hat{y}(t_i)$, with residual standard error

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \hat{r}_i^2}{n-2}}, \tag{4.72}$$

where the denominator $n - 2$ reflects one degree of freedom lost to the estimated slope $\hat{\mu}$ and one to the constraint that the trend passes through (t_1, y_1), which forces $\hat{r}_1 = 0$. The in-sample prediction interval is

$$\hat{y}(t) \;\pm\; 2\,\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(t - \bar{t})^2}{S_{xx}}}, \tag{4.73}$$

where $\bar{t}$ is the sample mean of the time index and $S_{xx} = \sum_i (t_i - \bar{t})^2$. The factor of 2 approximates the exact critical value $t_{n-2,\,0.025}$; for our sample sizes ($n \geq 42$) the approximation error is less than 1%. This is the standard OLS prediction interval for a new observation at time t within the sample period [389]. We define the normalized deviation from exponential growth as

$$d(t) = \frac{\hat{r}_t}{\hat{\sigma}}. \tag{4.74}$$

A technology deviates significantly from exponential growth at time t when $|d(t)|$ exceeds

the bound implied by Eq. 4.73.

For out-of-sample forecasting of a random walk with drift, we follow Farmer and Lafond [30] and use a τ-step-ahead forecast variance that accounts for the uncertainty in the estimated drift,

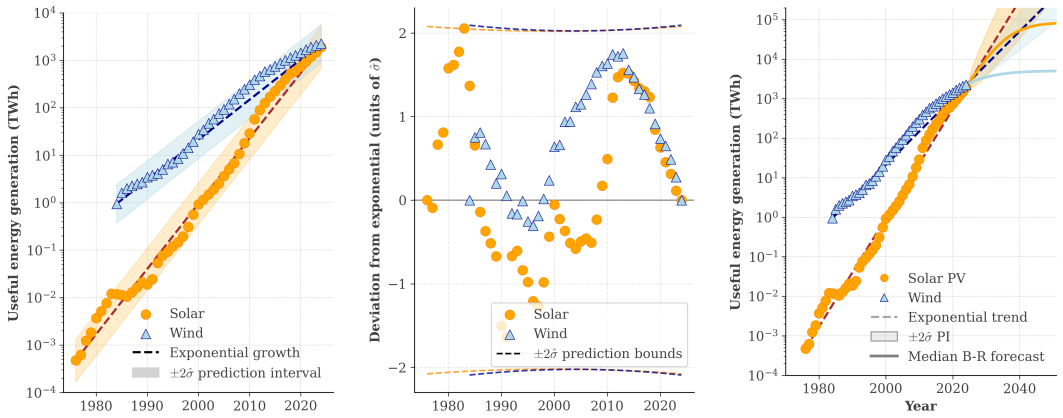
$$V(\tau) = \frac{\hat{s}^2}{1+\hat{\phi}^2}\left[-2\hat{\phi} + \left(1 + \frac{2(m-1)}{m}\hat{\phi} + \hat{\phi}^2\right)\left(\tau + \frac{\tau^2}{m}\right)\right], \tag{4.75}$$

where $\hat{s}^2 = \frac{1}{m-1}\sum_{t=2}^{T}(\Delta y_t - \hat{\mu})^2$ is the sample variance of the log-differences and $m = T-1$ is the number of log-differences. The τ^2/m term is due to the fact that the drift is estimated. Then we use

$$\hat{y}(T+\tau) = y_T + \hat{\mu}\,\tau \;\pm\; 2\sqrt{V(\tau)} \tag{4.76}$$

as a 95% prediction interval for the τ-step-ahead forecast.

For solar PV, $\hat{\mu} = 0.32$ (37% annual growth, doubling every 2.2 years), $\hat{\phi} = 0.41$, $\hat{s}^2 = 0.035$, and $\hat{\sigma} = 0.50$. For wind, $\hat{\mu} = 0.19$ (21% annual growth, doubling every 3.6 years), $\hat{\phi} = 0.32$, $\hat{s}^2 = 0.009$, and $\hat{\sigma} = 0.45$. Figure 4.48 presents the results.



(a) Useful energy generation **(b)** Deviations from exponential growth **(c)** Exponential extrapolation with prediction intervals

Figure 4.48: (a) Historical useful energy generation on a logarithmic scale, with exponential trend (dashed) and $\pm 2\hat{\sigma}$ prediction bands (shaded). **(b)** Normalized deviations $d(t)$. The wave-like pattern is consistent with the positively correlated errors implied by the MA(1) process. **(c)** Exponential extrapolation with prediction intervals (shaded), exponential trend line (dashed), and median B-R forecasts (solid).

Figure 4.48a shows that both technologies have followed approximate exponential growth over the past several decades. Figure 4.48b shows that historical deviations exhibit a wave-like pattern consistent with the positively correlated errors of the MA(1) process, but no sustained systematic departure from the trend within the historical sample. This is exactly what the B-R model predicts for technologies still in their early exponential phase: finite-L S-curves look indistinguishable from pure exponential growth until diffusion rises well above a few percent of the asymptote. The finite estimate of L is therefore not identified from observed saturation in the data. It is instead informed by our empirical Bayes prior on L, which encodes evidence from the cross-technology distribution of asymptotes (Section 4.7). Figure 4.48c shows when the median B-R forecast is expected to begin departing from exponential growth.

4.19.3.3 Comparing forecasts to scenarios

Figure 4.49 shows forecasts for shares and generation of useful energy from solar PV and wind, compared to the IEA NZE scenario [296, 297] and the Way et al. fast-transition trajectory [29]. The IEA NZE scenario for 1.5°C warming [296] and the Way et al. fast-transition trajectory [29] both lie below our median solar forecast. For wind, IEA NZE and Way et al. project higher deployment than our unconditional median forecast. Both sit on the upper end of the 90% prediction interval (5th–95th percentile). Our median forecast projects a combined solar and wind share of approximately 70% and total generation of approximately 85 PWh by 2050, compared to the IEA NZE projection of 50% and 45 PWh [296]. Our forecasts indicate that solar PV is likely to be the dominant source of renewable energy by 2050, while wind plays a much smaller role than in the IEA NZE scenario, albeit with a fat tail that includes higher wind deployment. These widened bounds are probably conservative for solar. In backtesting, fewer than 5% of technologies fall below their 5th-percentile forecast at early diffusion, so the lower bound is cautious (the Q-Q

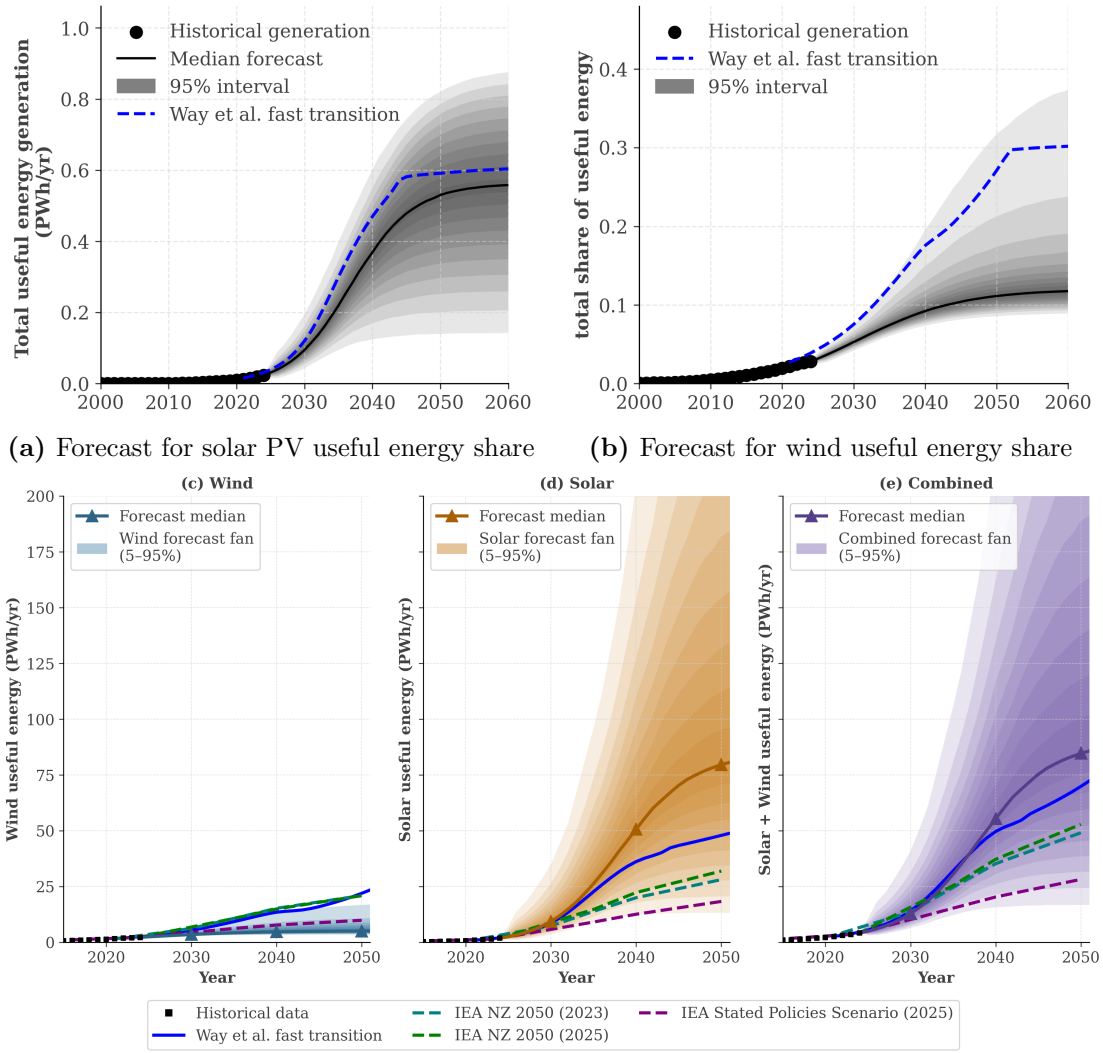


Figure 4.49: Same Bertalanffy-Richards distributional forecast as main-text Figure 4(d-f) for **(a)** solar useful-energy share and **(b)** wind useful-energy share. **(c) - (e)** are identical forecasts to main-text Figure 4(d-f), here compared to the IEA NZE scenario [296, 297] (green dashed; final electricity multiplied by 0.9 to obtain useful energy following Way et al. [29]) and the Way et al. fast-transition trajectory [29], rather than the AR6 ensemble.

curve in main paper Figure 3 sits below the diagonal at low percentiles). Because the widening is calibrated across technologies, we are cognizant that it may also over-widen the upper tail for an early-stage technology like solar.

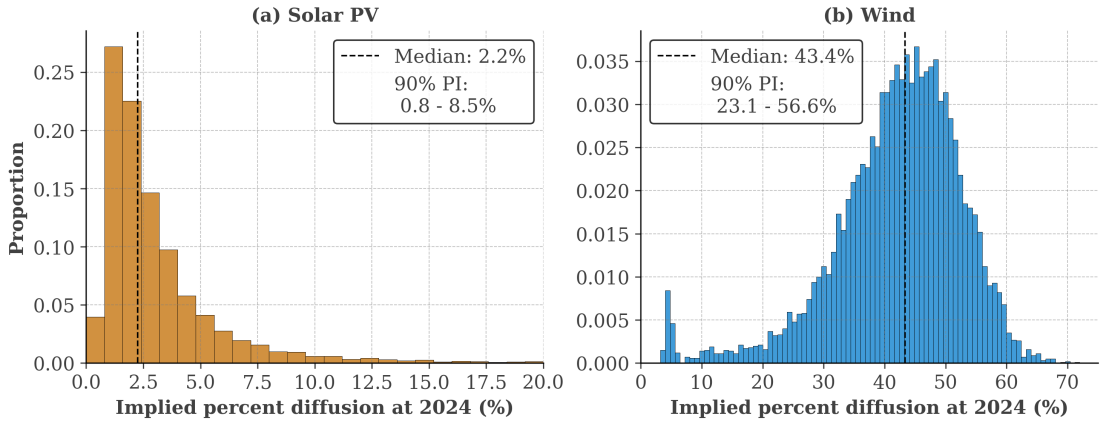
By 2050, our median forecast implies that solar will have an over 5-fold higher useful energy share than wind (over 50% versus ≈ 10%). The implied ratio in absolute useful energy generation is substantially larger. The median solar forecast reaches approximately 80 PWh/yr versus approximately 5 PWh/yr for wind, a roughly 16-fold difference. The

share and generation ratios diverge because total useful energy demand is not constant. When both technologies are measured relative to the same growing total, the gap in forecast seems to be smaller. This motivates the exogenous growth model presented below, which explicitly accounts for demand growth and yields shares and absolute generation forecasts that are more consistent with each other.

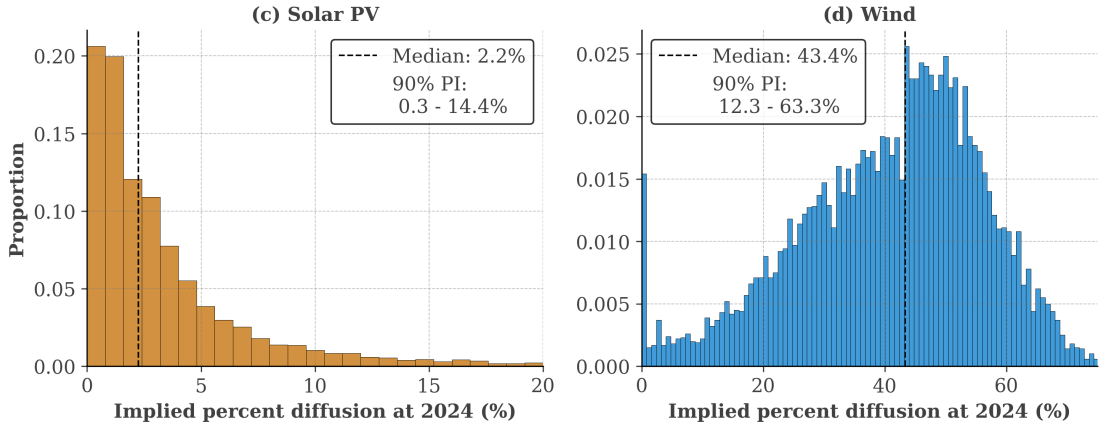
Inferred percentage of adoption. The posterior distribution for the asymptote implies a distribution over the current diffusion level. For a technology still growing, $\max_t Y_{i,t} = Y_{i,T}$, so the normalized asymptote is $\tilde{L}_i = L_i / Y_{i,T}$. The implied current adoption percentage is therefore

$$D_{i,T} = \frac{Y_{i,T}}{L_i} \times 100 = \frac{100}{\tilde{L}_i} \ \%. \tag{4.77}$$

Because $\tilde{L}_i$ has a posterior distribution, so does $D_{i,T}$. We apply the asymmetric widening transform of Section 4.7.5 to the posterior on L before computing the implied diffusion. The widening transform preserves the median by construction and stretches the tails. The widened posterior is the one used in the main paper because it is calibrated to match the empirical PIT distribution observed in backtesting (Section 4.16). Figure 4.50 reports the implied diffusion distribution. The top row (a, b) shows the un-widened and the bottom row (c, d) shows the widened implied diffusion of the 2024 generation. The median implied diffusion is identical in both rows by construction. The 5–95% interval is wider in the bottom row. The median implied diffusion is approximately 2.2% for solar PV and 43.4% for wind. The solar posterior places solar firmly in early diffusion. The wind posterior places wind at a more advanced stage but with greater uncertainty about how advanced. Re-fitting the wind-time-series from forecast origins one to five years earlier yields a posterior on L that narrows toward the same low value, indicating the data is becoming increasingly informative about the asymptote.



(a) Un-widened (raw posterior)



(b) Widened (used in main paper)

Figure 4.50: Posterior distribution of the current diffusion level $D_{i,T} = 100/\tilde{L}_i$ for solar PV and wind useful energy generation. **Top row (un-widened):** raw posterior. **Bottom row (widened, used in main paper):** after applying the asymmetric widening transform of Section 4.7.5. Vertical dashed lines mark the median implied diffusion (approximately 2.2% for solar PV and 43.4% for wind).

Exogenous demand growth. Figure 4.51 shows forecasts as in main-text Figure 4, but assuming 2% annual exogenous demand growth, using the methodology of Section 4.17. Solar reaches approximately 60 PWh/yr by 2050, compared to approximately 80 PWh/yr under the B-R model (without exogenous growth). This may seem counter-intuitive, but reflects a change in the model's parameters and a corresponding change in the L prior: to accommodate the exponential growth term, we use a lower L prior (since some growth is now absorbed by the exogenous term). This produces a lower median forecast in Solar's case. That is, the result is sensitive to the prior choice. Nonetheless, the qualitative implications of the forecast remain similar: the median solar forecast lies well above the

IEA NZE scenario, and the AR6 scenarios lie well below the median. The main difference is that, under the exogenous growth model, the median forecast up to 2040 lies slightly below the Way et al. [29] fast transition. Under the B-R model (without exogenous growth), the median lies marginally above it. For wind, the median forecast lies below the IEA NZE scenario under both B-R variants, but the exogenous growth model indicates higher upper bounds of the distributional forecast for wind. In both cases, the forecasts indicate solar is still in the early diffusion stage, while wind is in a more advanced stage of diffusion. We extend these forecasts to 2060 rather than 2050, because the exogenous demand growth keeps the trajectories rising as the technologies approach saturation, and the extra decade shows how they separate from the scenarios. The widening transform is not calibrated for this model, so we show the un-widened forecast. Applying it would only widen the bounds and would not change the conclusion. Under the exogenous model the median wind forecast tracks the median AR6 scenario, whereas in the main paper it lies below.

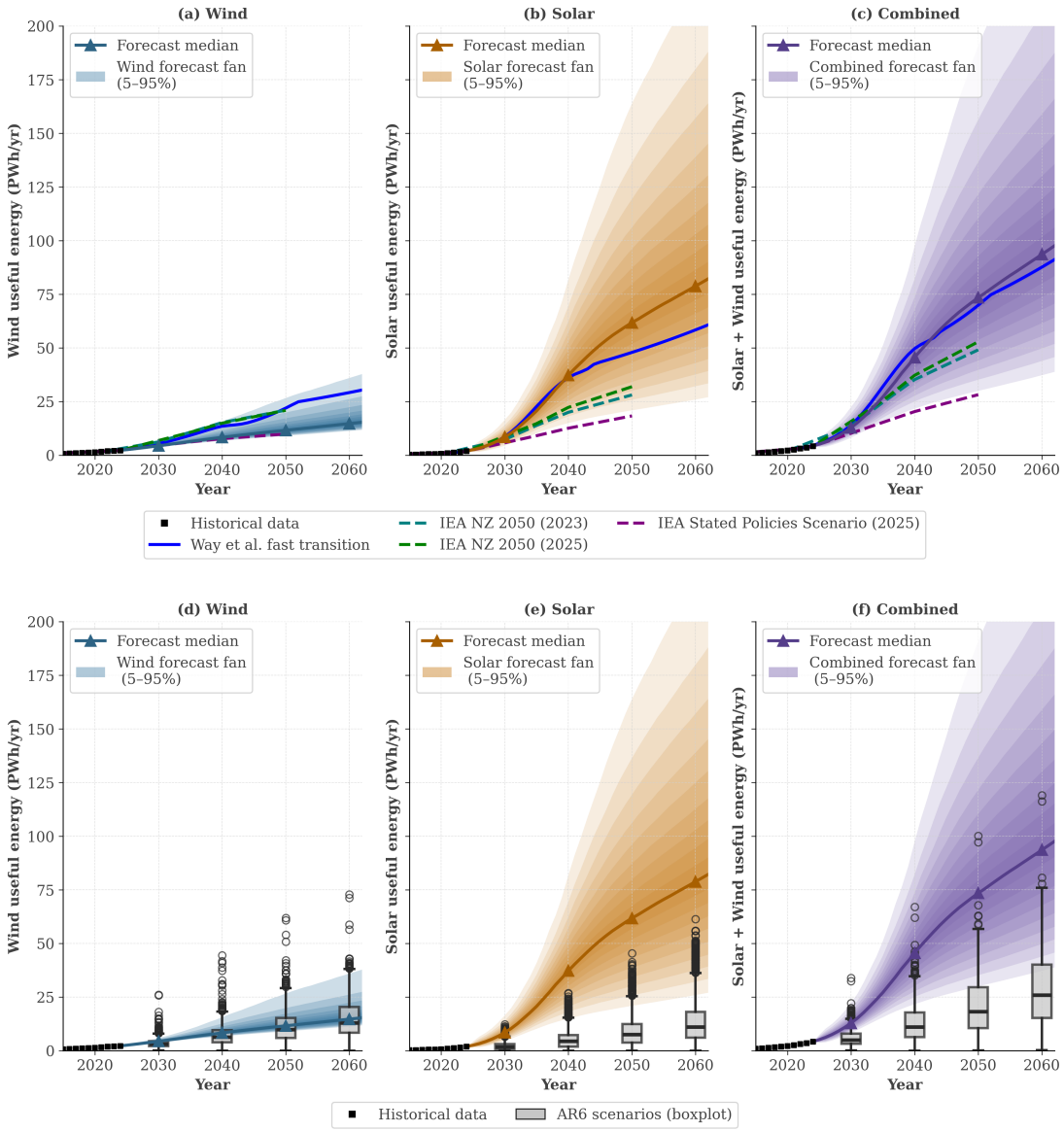


Figure 4.51: Un-widened forecasts assuming 2% annual exogenous demand growth, for solar PV, wind, and the two combined, shown to 2060. **Top row:** our forecast against the Way et al. [29] fast-transition trajectory and the IEA NZE scenarios [296, 297]. **Bottom row:** the same forecast against the IPCC AR6 scenarios [234], in the layout of main paper Figure 4(d-f), with wind, solar, and the two combined. The shaded band is the 5–95% prediction interval.

Comparison to literature capacity projections. Several studies project future solar PV capacity: 21.7 TW by 2050 and 84.5 TW by 2100 [390], and 133 TW by 2100 [391]. Assuming a capacity factor of 0.17 and applying the useful-energy conversion factor of $0.9 \times 0.87 \approx 0.78$ as per above, these capacity projections correspond to useful energy levels of approximately 25 PWh, 98 PWh, and 154 PWh, respectively, all falling within

the prediction interval for the forecast we have shown for both solar PV useful energy and capacity.

Achieving these higher capacity levels would require continued growth in energy and/or electricity demand potentially beyond the baseline. Surges in electricity demand have occurred, driven by AI and data centres among other factors. Electricity demand grew 4.4% in 2024 and 3.3% in 2025 [392]. Advanced technologies such as space-based solar power [393, 394] or perovskite tandem solar cells [390] could deliver improvements to capacity factors and/or the useful energy conversion factor, which would allow for higher useful energy generation for a given capacity. However, these technologies are currently in early stages of development, and it is uncertain whether they will be deployed at scale by 2050.

4.19.4 A pessimistic prior

Assume the position of a skeptic who believes that solar PV is likely not to grow much further. For example, they may believe that solar PV is already in a phase of decelerating growth. In such a case one may opt for more pessimistic priors for the asymptote L and the inflection point parameter t_0. Here we present forecasts for solar PV and wind useful energy assuming the prior

$$\text{backtesting prior:} \quad \pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2,\, 1,\, 5), \\ t_{0,i} \sim \Gamma\big(2,\, \max(5,\, T_{\max} - 3/\hat{k}_{\text{ssr},i}),\, 10\big), \\ k_i \sim \Gamma(1,\, 0,\, 1), \\ \sigma_i \sim \Gamma(1,\, 0,\, 1). \end{cases} \tag{4.78}$$

In this prior, the mode of the asymptote L is 6 times the current level (approximately 11.5 PWh/yr for solar). For solar, where $\hat{k}_{\text{ssr}} \approx 0.3$ and $T_{\max} = 50$, the mode of the inflection point $t_0 = \max(5, 50 - 3/0.3) + 10 \approx 50$ (equivalent to t_0 being at the last

observation).

Figure 4.52 shows the widened forecasts under this pessimistic prior. The widening transform of Section 4.7.5 was calibrated under the universal prior of Eq. (4.35) and is not re-calibrated here, hence should be treated as a conservative approximation. The median solar forecast in 2050 has decreased to approximately 63 PWh/yr, from approximately 80 PWh/yr under the prior of Eq. (4.35). Even under a pessimistic prior, the data is informative enough to converge to a posterior that still predicts substantial growth in solar PV useful energy generation. That is, the median posterior asymptote is 66 PWh/yr, compared to the prior median of 18 PWh/yr (and prior mode of $11 - 12$ PWh/yr). The median wind forecast in 2050 has remained at approximately 5 PWh/yr compared to the prior used in the main paper, and its posterior is now very narrow. The width of the distributional forecasts has decreased, reflecting the more concentrated prior. Even under this pessimistic prior the conclusions are qualitatively the same as in the main paper. Solar still grows far beyond its current level, while wind stays small.

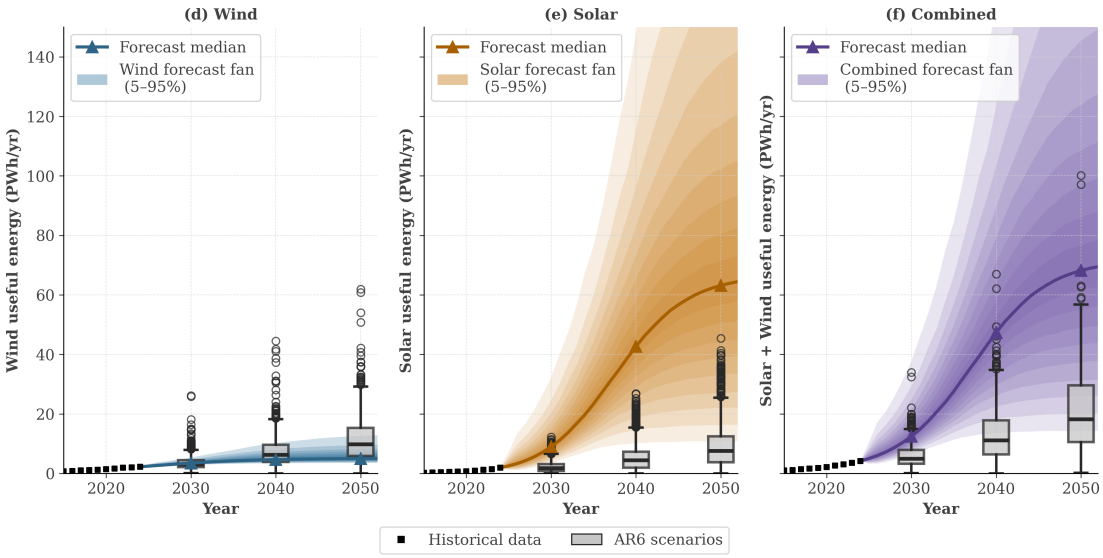


Figure 4.52: Bertalanffy-Richards forecasts with 90% prediction intervals in 5% steps (shaded), under a pessimistic prior for the asymptote L and the inflection point t_0. Green dashed lines show the IEA NZE scenario [296, 297], converted to useful energy by multiplying by 0.9 following Way et al. [29]. The median solar forecast in 2050 is approximately 63 PWh/yr. The median wind forecast in 2050 has remained at approximately 5 PWh/yr.

Figure 4.53 compares prior predictive distributions to the posterior predictive distri-

bution for solar PV and wind useful energy generation. The data has shifted the solar PV distribution towards much higher values, and the posterior median is substantially higher than the prior median in 2050. The data has shifted the wind distribution to much lower values. This illustrates how the data can overcome a pessimistic or optimistic prior, yielding forecasts of substantial growth (or limited growth, when the optimistic prior is incorrect). Despite this, it also indicates that the prior still has an influence on the forecast, as the median forecast has decreased compared to the prior of Eq. (4.35).

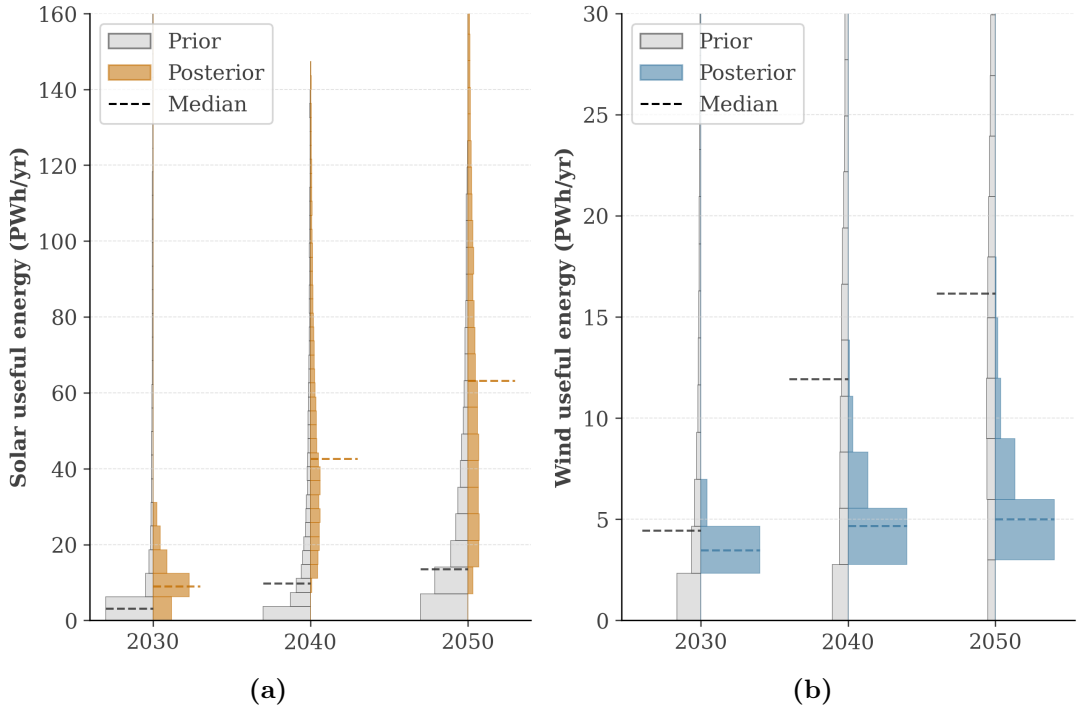


Figure 4.53: The figure compares the prior predictive distribution (grey) to the posterior predictive distribution under a pessimistic prior for the asymptote L and the inflection point t_0. **(a)** Solar PV (orange). **(b)** Wind (blue) useful energy generation. The data has shifted the solar PV distribution towards much higher values, and the posterior median is substantially higher than the prior median in 2050. The data has shifted the wind distribution to much lower values.

Across these three variations, the universal prior, exogenous demand growth, and a pessimistic prior, our qualitative conclusion is unchanged.

Chapter 5

Empirical solar and wind forecasts expose a solar-growth shortfall in AR6 mitigation scenarios

This chapter is taken from: Wagenvoort, B. & Ives, M. (2026). Empirical solar and wind forecasts expose a solar-growth shortfall in AR6 mitigation scenarios [395].

Integrated assessment model (IAM) scenarios are widely used to inform mitigation strategy, yet they are not probabilistic and may misrepresent technology diffusion. Here we combine distributional, empirically grounded forecasts of solar photovoltaics and wind electricity shares with the Intergovernmental Panel on Climate Change (IPCC) sixth assessment report (AR6) vetted scenario database to identify pathways consistent with observed deployment dynamics. We find that most AR6 scenarios fall below the forecast range for solar growth, indicating that rapid solar-led transitions are under-represented. Probable scenarios (those consistent with the solar+wind forecast) exhibit faster electrification and fossil-fuel phase-out, earlier net-zero (median c. 2068; with c. 10% before 2050), greater renewables-based hydrogen production, and similar levels of near-term carbon capture deployment but lower long-term reliance. Within this probable subset, end-century warming remains below 2°C with high probability ($> 80\%$), conditional on continued deployment trends and mitigation of non-energy emissions. Findings are robust across electricity and useful energy shares and levels. These results will hopefully motivate the generation of scenario projections with fewer ad hoc constraints on renewable growth and a better sampling of possible high-renewable transition pathways.

Keywords: Energy transitions; Solar PV; Wind energy; IPCC AR6; Climate scenarios; Net-zero; Technology forecasting

5.1 Introduction

Over the past half-century, the scientific community has published warming projections from anthropogenic greenhouse gas emissions and decarbonisation pathways to limit warming [396], most prominently through the Intergovernmental Panel on Climate Change (IPCC) [397] and the International Energy Agency (IEA) [398]. Both organisations regularly publish decarbonisation scenarios that strongly influence governments, financial institutions, industry, investors, and the public [24].

Published projections from the IPCC and IEA have, however, systematically underestimated renewable energy deployment [28, 224, 399] and overestimated future technology costs [30, 400]. In particular, most historical projections have underestimated solar PV deployment [24, 28, 29, 400], drawing criticism that such scenarios portray low-warming pathways as unnecessarily difficult and costly [24, 29, 401], prompting calls for better-parameterised energy-economy models [28, 402–406].

The Integrated assessment modelling (IAM) community has worked to address these criticisms [27], notably through the ADVANCE Inter-model Comparison project [407]. The sixth assessment report (AR6) incorporates more renewable capacity than AR5, with many scenarios now adopting declining renewable cost curves [30, 163, 164] and non-linear renewable growth [29, 261], yielding more rapid, low-warming transition scenarios. As cost declines in clean technologies become increasingly apparent, the range of projected future warming is narrowing [408]. While AR5 portrayed low-warming scenarios as nearly impossible, AR6 concludes they remain challenging but achievable [27], and the highest warming scenarios are now considered very unlikely [409].

Recent research has indicated that solar PV is destined to be a major, affordable energy source in most countries [410], and likely to meet the majority of our future electricity demand [29, 411]. Distributional forecasts grounded in historical data suggest that solar

PV will likely become the primary future provider of *energy*, and that many AR6 scenarios produce substantially lower solar useful energy estimates than these empirically grounded forecasts [284]. Such empirically grounded forecasts do not align well with most fast, low-warming pathways in AR6, which rely on substantial levels of carbon capture and removal [412], a potentially costly alternative. Accordingly, published estimates of the probability of limiting warming below 2°C remain quite low (∼5%) [177, 413]. These findings raise critical questions: which AR6 scenarios are consistent with historically validated deployment data? What do probable scenarios reveal about fossil fuel phase-out, renewable deployment, and warming outcomes? Do alternative socio-economic assumptions produce probable pathways? How do these scenarios compare to the full AR6 ensemble in terms of net-zero timing and cumulative emissions?

We present a filtering framework that complements the AR6 scenario database. By comparing each AR6 scenario to empirically grounded, probabilistic forecasts for solar PV and wind electricity generation shares, we identify the subset consistent with historical deployment data. The forecasts are derived from observed deployment patterns and are independent of IAM outputs. Hence, any agreement or disagreement between the two is therefore informative rather than circular. The framework is general and can accommodate any distributional forecast and scenario ensemble. Here we apply it using recent, back-tested S-curve forecasts [284] and all 1189 vetted AR6 scenarios [27]. We characterize how scenarios consistent with forecasts differ in energy portfolios, fossil fuel phase-out timelines, and warming outcomes.

5.2 Methods

5.2.1 Approach

We filter the AR6 scenario ensemble to a smaller subset consistent with the historical growth trajectories of the combined solar PV and wind electricity generation share, retaining the structural diversity of scenarios from different integrated assessment models. We refer to scenarios passing this filter as *probable* scenarios, and those that do not as *improbable*.

We apply the framework using back-tested S-curve forecasts for solar PV and wind [284] and the 1189 vetted AR6 scenarios. We compare each scenario to the forecast distributions and determine consistency at each time step (2030, 2040, . . . , 2100). The probable scenarios provide a more empirically grounded and thus policy-relevant basis for analysing the energy transition and assessing warming outcomes.

In determining whether a scenario is probable, we make no assumptions about other technologies (nuclear, hydro, geothermal, fossil fuels with CCS, or CDR). Our analysis focuses on solar PV and wind contributions, assuming that the combined solar PV and wind electricity generation share provides a sufficient indicator of net-zero scenario achievability. The forecasts are unconditional and based solely on historical generation data; our likelihood estimates are therefore less reliable for scenarios that rely heavily on technologies other than solar and wind.

5.2.2 Calculating electricity generation shares from AR6 scenarios

We calculate electricity generation shares directly from technology-specific secondary electricity generation (SE) data in the AR6 database [27]. The AR6 database reports technology-specific generation at the secondary energy level but lacks technology-specific

final energy disaggregation. Since transmission and distribution (T&D) losses apply uniformly across all electricity technologies, shares calculated at either level are identical. See Appendix 5.7 for a derivation. For results using useful energy shares, see Appendix 5.12.1 for the calculation method and Appendix 5.13.5 for results.

5.2.3 Probable scenario identification

In our main analysis, we apply a binary likelihood criterion at each time step: a scenario is either consistent with the forecast (*probable*) or not (*improbable*). For brevity in equations, we retain the notation $\mathcal{L}$ (likelihood). Likelihood is determined by whether the scenario's combined solar and wind electricity generation share falls within the probability mass of the forecast distribution. We then identify scenarios consistent across all time steps up to a given year (e.g., 2050) and analyse this subset for insights into probable energy transition trajectories. Alternative formulations (product likelihood, yearly likelihood), including ranking and weighting of scenarios, and sensitivity analyses appear in Appendix 5.9.

We calculate scenario likelihoods based on the sum of solar and wind electricity shares. Since the sum of two S-curves with different growth rates is not itself an S-curve (see Appendix 5.10), we forecast each technology independently and sum the resulting distributions. This approach allows technology trade-offs where over-performance in one technology can compensate for underperformance in another, reflecting realistic transition dynamics.

5.2.3.1 Likelihood measure

For scenario s at time t, we define the combined solar and wind electricity generation share as $Y_t^{(s)} = Y_{\text{solar},t}^{(s)} + Y_{\text{wind},t}^{(s)}$. The distributional forecast for the combined share is

$$y_t \sim \mathbb{P}(Y_{\text{solar},t} + Y_{\text{wind},t} \mid \text{historical data}) \tag{5.1}$$

A scenario s is probable at time t when

$$\mathcal{L}(Y_t^{(s)}) = \mathbb{I}\left(Y_t^{(s)} \in [y_t^{2.5}, y_t^{97.5})\right) \tag{5.2}$$

where $y_t^{2.5}$ and $y_t^{97.5}$ are the 2.5th and 97.5th percentiles of the forecast distribution at time t, and $\mathbb{I}$ is the indicator function. That is a scenario is probable if it falls within the 95% prediction interval at time t (giving it value 1), and improbable if it lies outside that region (giving it value 0). Equation 5.2 provides a single-year compatibility score. To assess consistency across the full scenario trajectory (trajectory likelihood), we multiply scores across time steps using

$$\mathcal{L}(Y^{(s)} \mid T) = \frac{\prod_{t\in\{2030,2040,\ldots,T\}} \mathcal{L}(Y_t^{(s)})}{\sum_s \prod_{t\in\{2030,2040,\ldots,T\}} \mathcal{L}(Y_t^{(s)})} \tag{5.3}$$

Equation 5.3 constitutes our primary analysis: it identifies scenarios that maintain consistency with the forecast across all time steps up to T. We classify a scenario as *probable* when $\mathcal{L}(Y^{(s)} \mid T) = 1$ (i.e., the scenario falls within the 95% prediction interval of the forecast for all time steps up to T), and *improbable* when $\mathcal{L}(Y^{(s)} \mid T) = 0$ (i.e., the scenario falls outside the 95% prediction interval for at least one time step). Appendix 5.9 and 5.13 present other methods and show results consistent with those presented in the main text.

5.2.3.2 Interpretation and alternative formulations

The trajectory likelihood in the main text is not a direct probability distribution over the AR6 scenario set, but rather a binary classification based on alignment with the forecast. Plausible trajectories not represented in AR6 fall outside the scope of this analysis. In Appendix 5.9 we present alternative methods, one of which constitutes a proper conditional probability over the AR6 scenario set under two conditions: (i) the forecast distributions are well-calibrated, as demonstrated for the forecasts used in this paper by back-testing in

Wagenvoort et al. [284]; and (ii) forecast errors are temporally independent across decades, an assumption we adopt here.

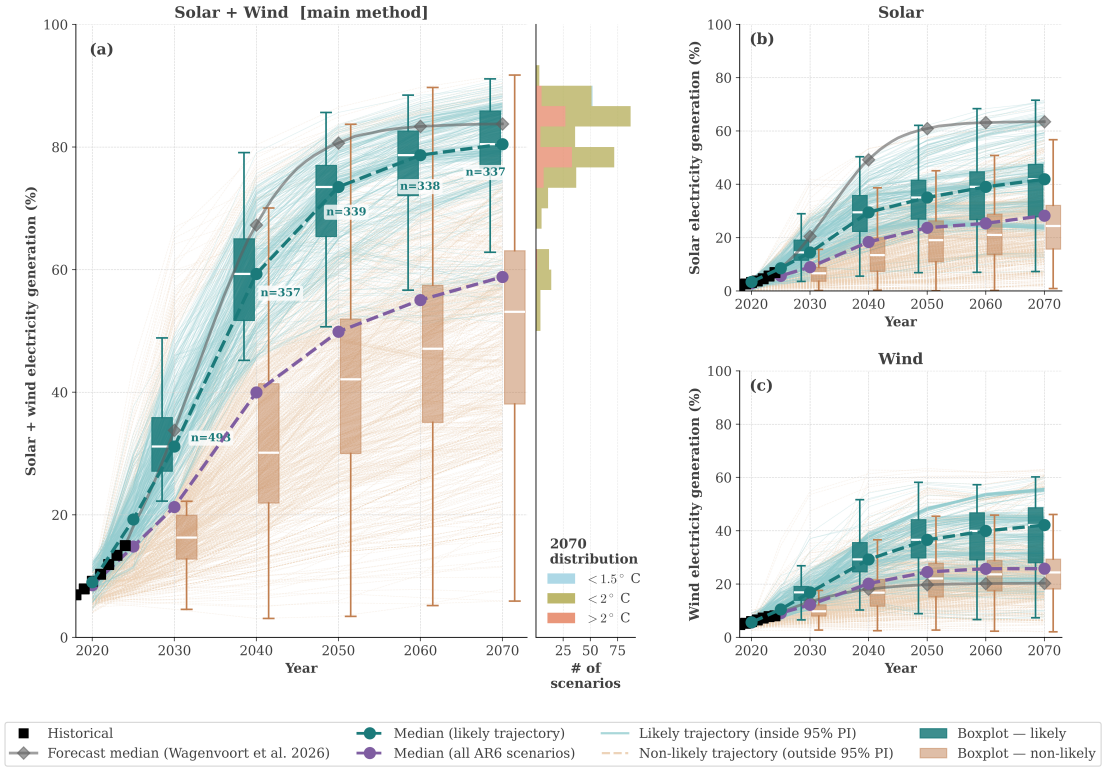
5.3 Results

We present three main results. First, we present the probable vs improbable AR6 scenarios, and how their solar+wind trajectories differ. We also compare the forecast to other scenarios to showcase the usability of our tool beyond the AR6 scenarios. Second, we examine the energy portfolio evolution and warming scenarios of probable vs. improbable AR6 scenarios. Third, we investigate the shared socioeconomic pathways (SSPs), and their relative representation in the probable vs full AR6 scenarios.

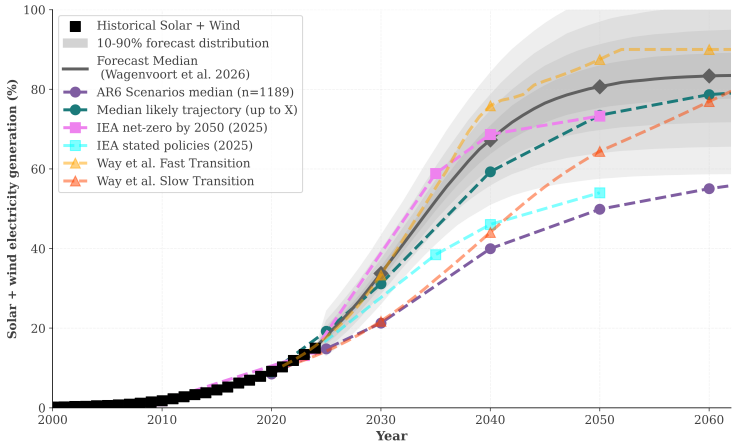
5.3.1 Probable vs. improbable renewable growth pathways

We compare the filtered set of probable and improbable AR6 scenarios based on wind and solar electricity generation share (Figure 5.1). We also compare the distributional forecast to other scenarios. The main text uses electricity generation shares as the primary measure of solar and wind penetration. To test robustness, the Appendix presents results using three complementary measures: electricity generation levels (PWh), which account for differences in total electricity demand growth across scenarios, since high shares could reflect low demand rather than high deployment (Appendix 5.13.4); useful energy shares, which extend beyond the electricity sector to all end-use energy demand and account for end-use efficiency losses (Appendix 5.13.5); and useful energy levels, (Appendix 5.13.6). We also present results for alternative likelihood formulations including product and trajectory likelihoods (Appendix 5.9), likelihood-weighted results, and a larger set of AR6 scenarios (Appendix 5.13). The core conclusions hold across all measures and formulations, strengthening confidence that the distinction between probable and improbable scenarios

is not an artefact of the choice of metric.



(a) Probable vs. improbable AR6 scenarios



(b) Comparison to other scenarios

Figure 5.1: Comparison of solar PV, wind, and solar PV + wind electricity generation for probable (teal) and improbable (tan) AR6 scenarios. Historical data (black squares) shown for 2020–2024. Forecast median (grey diamonds), median of all AR6 scenarios (purple), median of probable scenarios (teal). **(a–c)** Progression of probable and improbable AR6 scenarios over time. Histogram indicates solar+wind electricity share in 2070 coloured by surface warming. **(a)** Solar PV + wind (main method), and the disaggregation into **(b)** Solar PV, **(c)** wind electricity shares. The number of probable scenarios decreases over time to 337 in 2070. **(d)** Median probable scenario trajectories (teal) compared to IEA scenarios (pink, cyan) and Way et al. transitions (yellow, red).

Our method filters the AR6 scenarios (Figure 5.1, median in purple) into a probable

(teal) and an improbable (tan) set. The two sets do not overlap at the first year of assessment (2030). Beyond this point, some scenarios that initially grow slowly may accelerate and fall within the 95% prediction interval at a later stage; these are still considered improbable. Similarly, scenarios that prescribe too-rapid deployment and then plateau early are also excluded. Probable scenarios tend to generate between 75 and 85% of solar + wind electricity by 2070 (median probable AR6 scenarios, teal). Most of those limit warming to below 2°C (green bars in histogram), some exceed 2°C (red bars), while only 2 limit to below 1.5°C (blue bars). The yearly likelihood methodology produces complete distinct sets at every year (see Appendix 5.13). The number of scenarios decreases from 498 in 2030 to 337 by 2070. The discussion in Section 5.5 addresses the limitations of statistical power for a lower number of scenarios.

Consistent with prior evidence of solar underestimation in IAM scenarios [28, 224], most probable AR6 scenarios exhibit slower solar growth than the forecast, lacking the approximately exponential growth characteristic of S-curves [36, 91]. Moreover, the median of 1189 AR6 scenarios (purple) falls below the 10th percentile. For combined solar plus wind electricity generation, most scenarios fall below the forecast median between 2040 and 2050: Probable AR6 scenarios and the forecast align until 2030, after which their growth rates diverge–Probable AR6 scenarios level off while the forecast continues to grow.

For these forecasts, combined solar and wind share can reach 100% in a small fraction of the distribution beyond 2050 (Figure 5.1b). This result comes from summing independent forecast distributions rather than modelling them jointly; we discuss this choice and alternatives in Appendix 5.10. Appendix 5.13 compares forecasts against scenarios weighted by their combined solar and wind likelihood to validate our consistency measures.

The IEA net-zero by 2050 scenario [297] (pink) exceeds the median probable AR6 scenario trajectory (teal), closely tracking the forecast median for the next decade before levelling off (Figure 5.1b). The IEA stated policies scenario [297] (cyan) falls well below the

median of probable scenario trajectories and even below the 10th percentile throughout, indicating that current policy ambition alone is insufficient to match historically grounded deployment rates.

The Way et al.[29] fast transition (yellow) and slow transition (red) straddle the median probable scenario trajectories until 2050, after which the slow transition intersects it. Both straddle the forecast median, with the fast transition tracking it reasonably well. These scenarios are conditional forecasts projecting cost declines via Wright's Law given an assumed deployment trajectory. That probable AR6 scenario trajectories lie between the fast and slow transitions confirms their consistency with plausible cost-based projections of the energy system.

5.3.2 Energy portfolio evolution

We now examine how the set of probable scenarios translates into the global energy portfolio mix, warming and net-zero. Figure 5.2 compares probable and improbable AR6 scenarios.[1] Table 5.5 in Appendix 5.13.2 provides statistical comparisons between probable, improbable, and all AR6 scenarios for each variable and time point. See Appendix 5.13 for additional variables, including global GDP and population.

Probable scenarios measured in shares limit what we can say about the total growth in solar + wind. High shares may be due to limited energy demand growth, as opposed to high deployment. To address this, we also analyse electricity generation levels in Appendix 5.13.4 and useful energy shares and levels in Appendices 5.13.5 and 5.13.6; results are consistent across all measures. Nonetheless, probable scenarios measured in shares project a significant increase in total electricity generation from 2030 onward: generation nearly triples to a median of 90 PWh by 2060, compared to 31 PWh in 2024 [210]. Fossil fuel electricity shares fall to near zero in the median over the same period, while solar and wind rapidly

[1] Alternative formulations (product likelihood 5.9.3, trajectory likelihood 5.9.1) and detailed results in Appendix 5.13.

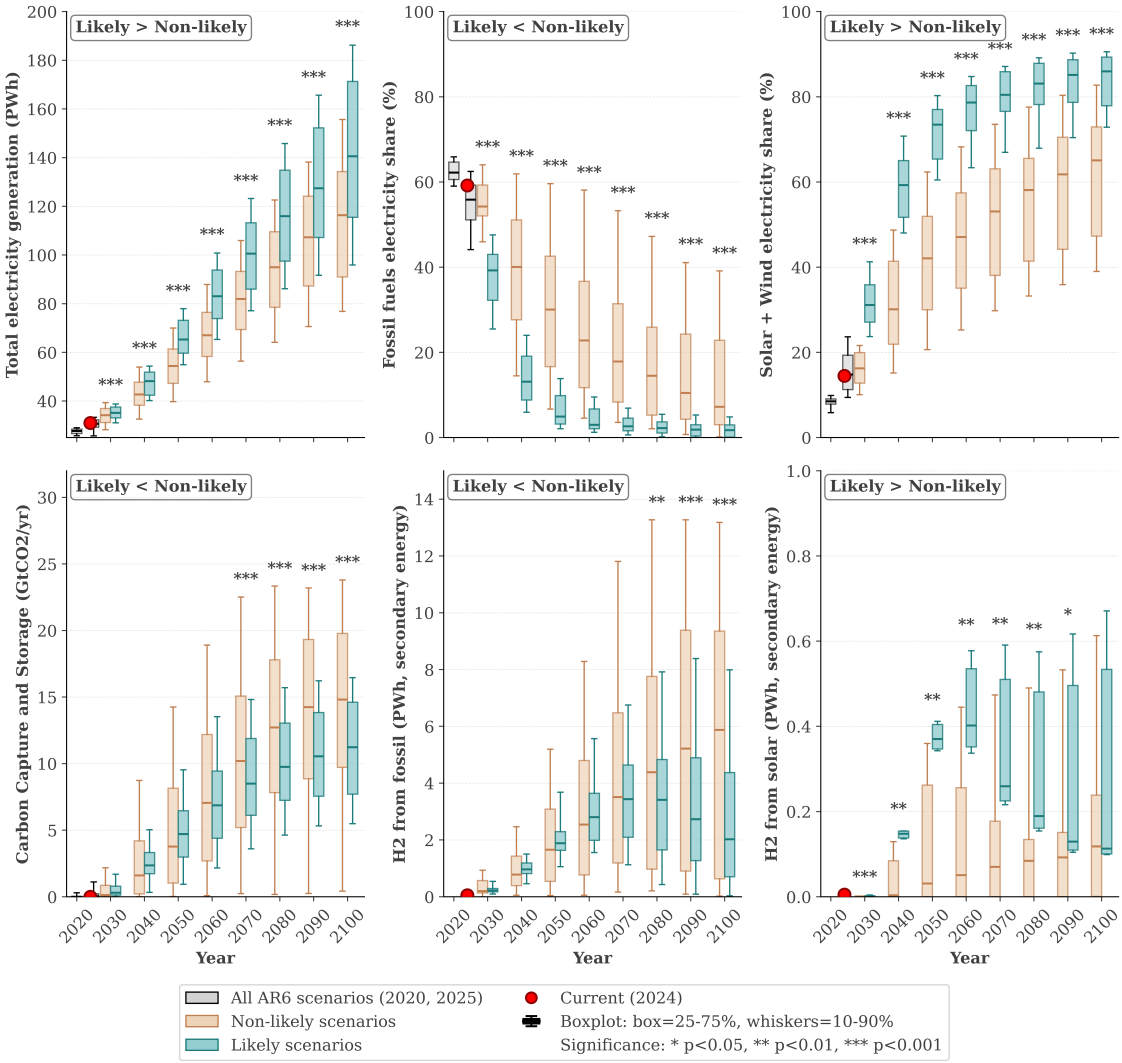


Figure 5.2: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan). Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. Grey box-plots represent AR6 scenarios in 2020 and 2025, with the red dot marking the 2024 observed levels. **(a)** Total electricity generation. **(b)** Fossil fuel electricity generation. **(c)** Solar and wind electricity generation share. **(e)** Yearly carbon sequestration from CCS. **(d)** Blue hydrogen (from fossil fuel incl. CCS) production (secondary energy). **(f)** Hydrogen production from solar (secondary energy).

become the dominant sources, reaching a median share of ∼75% by 2050.

Both probable and improbable scenarios project similar CCS growth up to 2060, after which probable scenarios deploy less CCS while improbable scenarios rely more heavily on it to achieve net-zero. CCS therefore plays an important role across both subsets. Results are similar for alternative formulations (product and trajectory) in Appendix 5.13, though the difference in CCS deployment between subsets is more pronounced for the

weighted trajectory likelihood. For hydrogen production, both subsets show comparable blue hydrogen production up to 2070, but probable scenarios show significantly higher renewables-based hydrogen production. Data limitations produce a systematic decrease in hydrogen production from both fossil fuels and solar PV in probable scenarios after 2070. These energy portfolio patterns are consistent when assessed using electricity generation levels (Appendix 5.13.4) and useful energy shares and levels (Appendices 5.13.5 and 5.13.6), although the timing at which CCS deployment diverges between subsets varies slightly across measures.

5.3.3 Net-zero timing and emissions budgets

We compare the distribution of net-zero year, emissions and surface warming of probable and improbable scenarios (Figure 5.3).

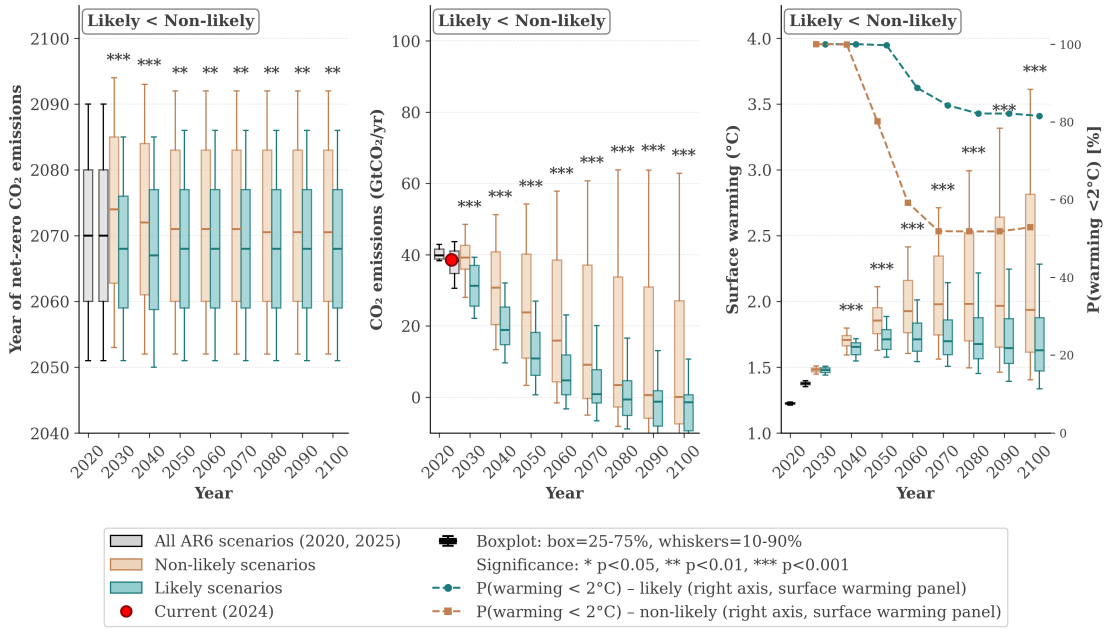


Figure 5.3: Boxplot comparing climate outcomes for probable (teal) and improbable (tan) AR6 scenarios. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. Grey box-plots represent AR6 scenarios in 2020 and 2025, with the red dot marking the 2024 observed levels. **(a)** Distribution of net-zero year for probable (teal) and improbable (tan) AR6 scenarios. **(b)** Distribution of cumulative CO_2 emissions to net-zero for probable and improbable AR6 scenarios. **(c)** Surface warming and probability of limiting warming to below 2°C for probable and improbable AR6 scenarios.

10% of probable scenarios reach net-zero before 2050, with the median net-zero year occurring in 2068. Other metrics yield slightly earlier median net-zero years (2050 for product likelihood shown in appendix 5.13) with a higher proportion of scenarios reaching net-zero before 2050 (18% for non-binary trajectory likelihood, shown in appendix 5.13). These results suggest that while net-zero by 2050 is possible, it is not the central outcome implied by the probable AR6 scenarios.

Probable scenarios reduce CO_2 emissions immediately, placing the majority below 2°C warming budgets (> 80%). Some probable scenarios still reach nearly 2.5°C; understanding the characteristics of these scenarios could help identify potential risks and mitigation opportunities. Moreover, nearly no scenarios limit warming to below 1.5°C, consistent with observations that global warming may have exceeded 1.5°C in 2024 [414]. Some overshoot and return to 1.5°C with help of CCS. Hence, while probable scenarios indicate limited warming, especially extreme warming, they also indicate that additional efforts are required to limit warming well below 2°C, and achieve net-zero by 2050. These climate outcomes are robust across all complementary measures: probable scenarios limit warming to below 2°C with > 80% probability whether assessed using electricity generation shares, levels, or useful energy (Appendices 5.13.4, 5.13.5, and 5.13.6), while median net-zero year shifts only marginally across measures.

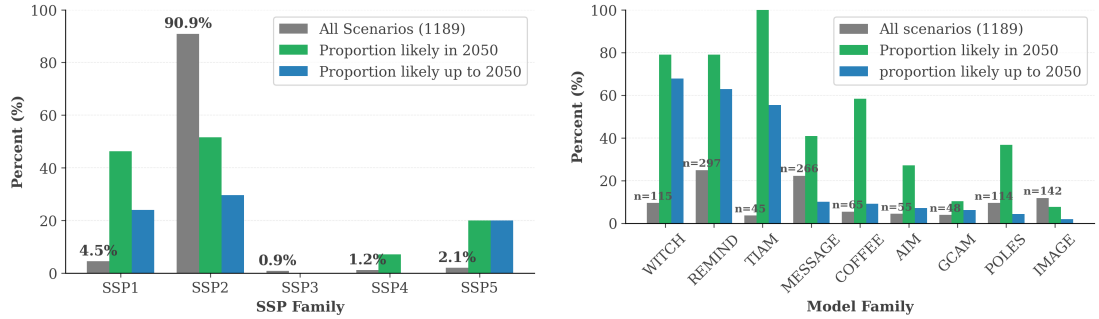
5.3.4 From Shared Socioeconomic Pathways to policy-relevant mitigation pathways

Since AR5, many climate mitigation scenarios have applied one of the five Shared Socioeconomic Pathways (SSPs), developed to support integrated climate research across the IPCC Working Groups [415]. The SSPs represent different socioeconomic futures (e.g. GDP growth, population growth, consumer attitudes towards renewables) and are designed to

capture the uncertainty of our complex global socio-economic system, describing worlds with varying challenges to mitigation and adaptation [415]: SSP1 (sustainability), SSP2 (middle of the road), SSP3 (regional rivalry), SSP4 (inequality), and SSP5 (rapid growth). Most AR6 scenarios adopt SSP2 as their underlying set of socioeconomic assumptions [27]. Figure 5.4 reveals a strong compositional bias: 91% of the 1189 vetted scenarios are SSP2 scenarios, whereas the probable subset is more balanced across SSP1, SSP2, and SSP5. This finding carries an important implication: not only SSP2 (middle-of-the-road) pathways are consistent with historical deployment data; high economic and renewable growth scenarios (SSP5) are equally consistent. Meanwhile, SSP3 (regional rivalry) scenarios are barely represented in the probable subset—not because such futures are impossible, but because historical solar and wind growth patterns are inconsistent with a regionalised rivalry scenario. This uneven representation reflects a structural limitation of the AR6 ensemble rather than a conclusion about SSP3's impossibility. Future scenario development should actively explore a wider range of socio-economic assumptions, particularly SSP1 and SSP5, to test the boundaries of uncertainty and avoid inadvertently embedding SSP2 as the de facto baseline.

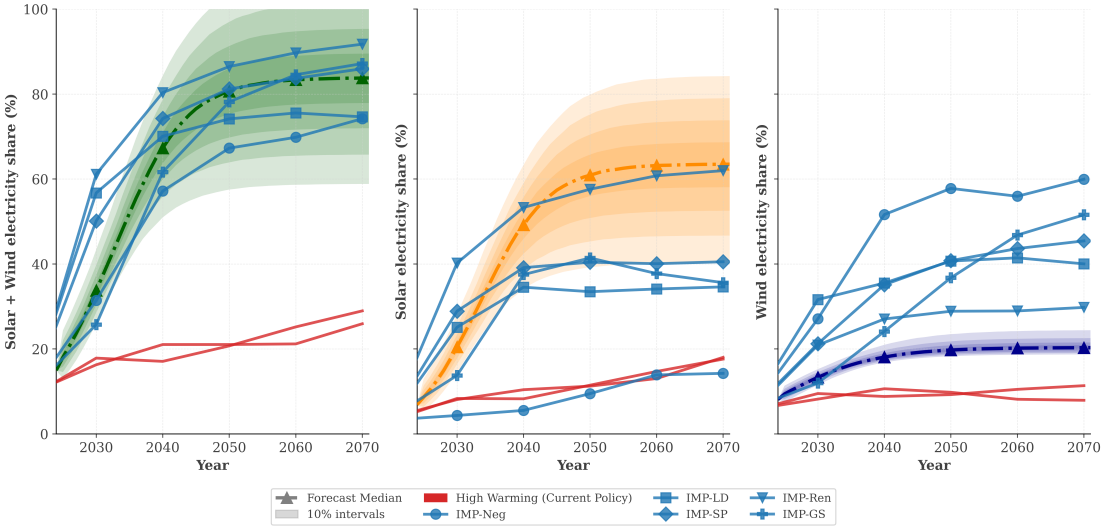
Given the limitations of SSP application, the AR6 report focuses on a small number of illustrative pathways (IPs) that highlight key mitigation-strategy themes [27]. The AR6 WGIII report features seven IP scenarios: two high-warming scenarios under current policies, CurPol [416] and ModAct [417], and five illustrative mitigation pathways—IMP-Ren [418], IMP-Neg [416], IMP-LD [68], IMP-GS [419], and IMP-SP [420]. These IPs are purely illustrative; no assessment of their likelihood is implied [421]. They highlight the implications of different societal choices for future emissions and sectoral transformations, drawn from scenarios that passed a second vetting process for plausible near-term trends such as CCS levels in 2030 [422]. We compare these IPs to the probabilistic forecast to assess alignment and to tease out the implications of the societal choices they embody.

Figure 5.4c compares seven mitigation pathways to the probabilistic forecast for solar, wind, and combined solar plus wind electricity generation shares. Figure 5.4c shows that



(a) SSP scenario likelihoods

(b) Models consistent with historical growth patterns



(c)

Figure 5.4: **(a)** Share of Shared Socioeconomic Pathway (SSP) scenarios in AR6. Grey: proportion of each SSP. Light green: percentage within the SSP group probable in 2050. Blue: percentage probable across the full trajectory to 2050 (consistent in 2030, 2040, and 2050). **(b)** Bar chart of IAM models, ordered by the proportion of their scenarios that are probable.**2. row** Comparison of 7 illustrative mitigation pathways (5 fast transition in blue and 2 current policy in red)to the probabilistic forecast for solar + wind, solar, and wind electricity generation shares. Shaded regions show 80% prediction interval (10% steps). Red lines indicate high-warming scenarios; blue lines indicate illustrative mitigation pathways.

solar shares for each IP fall well below the forecast median from 2030 onward, except IMP-Ren, which tracks the median reasonably well. The current policy (red) scenarios fall below the 5th percentile. Wind shares exceed the forecast median for all mitigation pathways except current policy scenarios, which again fall below the 5th percentile. Combined solar

plus wind shows that illustrative mitigation pathways (blue) tend to lie within the 80% prediction interval, while current policy scenarios fall below the 10th percentile throughout.

Three illustrative pathways—IMP-Ren, IMP-LD, and IMP-SP—show faster near-term growth than the forecast median but eventually track the forecast reasonably well. Mitigation pathways (blue) tend to project higher wind growth and lower solar growth than the forecast median. The two current policy scenarios (red) fall well below the 10th percentile for both solar and wind, indicating that current policy ambition does not match historically grounded deployment rates.

IMP-Ren is the only IMP that tracks the forecast median for solar, which assumes rapid solar deployment as the primary transition driver [418]. As of 2040, for combined solar plus wind, all 5 IMPs (blue) fall within the 80% prediction interval. This does not mean they all deploy similar levels of solar and wind generation. Appendix 5.13.4 compares the absolute electricity generation levels of these pathways to forecasts of solar and wind generation in PWh, showing that IMP-LD (low demand IMP) deploys significantly less solar and wind generation than the forecast median, while IMP-Ren (renewables-focused IMP) and IMP-GS (general sustainability IMP) deploy significantly more. This highlights that similar shares can arise from different combinations of deployment and demand growth.

Most IMPs deploy limited CCS compared to the wider AR6 ensemble, except IMP-NEG, which focuses on negative emissions technologies. IMP-GS deploys a moderate amount of CCS (approximately 5 $GtCO_2$/yr by 2070). IMP-Ren deploys less than 3 $GtCO_2$/yr at net-zero, aligning well with the probabilistic forecast. IMP-LD and IMP-SP deploy negligible CCS yet remain probable pathways.

WITCH and REMIND are by far the two models that generate scenario trajectories most consistent with the forecasts (Figure 5.4b). Approximately 60–70% of scenarios from these models are probable with historical growth patterns of solar and wind. TIAM is also reasonably represented, while other models mostly generate improbable scenarios. This

raises the question of what characteristics of these models lead to more (or fewer) probable scenarios.

5.4 IAM constraints on energy transitions

Creutzig et al. [28] identify a number of reasons for why IAMs have underestimated the potential of solar energy in their decarbonization scenarios. Most IAMs promote climate action through the use of carbon pricing as a proxy for all possible climate policies, which fails to capture the technology-specific policies and personal preferences of customers that have promoted solar over other technologies. The IAMs were also overly optimistic about the costs and potentials of alternative low-carbon technologies, such as carbon capture and storage and nuclear power.

They also point to the fact that solar has experienced a higher learning rate than expected in the models. The majority of the IAMs do apply endogenous technological learning (ETL) to capture such learning rates for a number of energy generation technologies (AIM and GCAM are exceptions). ETL is a modeling technique designed to replicate the empirical relationship between the cost of technologies and the cumulative investment in that technology, whereby the costs of a technology decline with cumulative experience - i.e. experience curves or learning rates [30]. A model run with ETL is very likely to select technologies with high learning rates, and will invest heavily in these technologies quite early on (depending on how myopic the models are), unless they are prevented from doing so by constraints [423].

We would contend that insufficiently high learning rates for solar do not fully capture what restricts the endogenous growth of solar PV in the IAM modeled scenarios. We also need to examine the many constraints imposed on the deployment and cost declines placed on solar. Constraints we have identified for solar and wind come in many forms, including

'floor costs' on technologies, build-rate constraints, deployment constraints, and integration constraints. Table 5.1 provides an overview of how each of the IAMs generally utilise such constraints and whether we see evidence of their use in the AR6 scenarios.

Floor costs are potentially the most common, although it is difficult to tell with any accuracy how much they are used, as documentation of their nature and explanation of their application are not as clear as would seem prudent in such high-impact modelling efforts. Way et al. [29] presented graphically some of the documented floor costs applied in IAM modeled scenarios in AR6 and previous assessment reports, showing how they have been routinely revised as actual costs have been driven below the original floor costs - suggesting such constraints have had questionable justification, and have been applied too conservatively in the past. Looking for potentially undocumented floor costs in the AR6 database (Table 5.1), we found that many models still hold solar PV capex costs well above their current global average levels of $500/kW in 2023 [424]. Furthermore, the problem appears to go beyond the use of floor costs, with even the most recent studies using quite obsolete or very conservative cost values for solar and wind [400].

Build constraints in the IPCC IAMs are also not well documented, but evidence suggests some scenarios do limit how quickly modular technologies like solar can scale globally. For example, the CurPol, NDCplus, Bridget, 2Deg2020, and 2Deg2030 scenarios, in the AR6 scenarios produced using each of the major IAMs (AIM, IMAGE, MESSAGEix, POLES, REMIND, TIAM, WITCH and COFFEE), as well as the IMP-GS illustrative mitigation pathway, utilize a stated build limit whereby "the increase of the share of renewables in total electricity generation per year (starting in 2020, until 2050 and up to 50 %, maximum) cannot exceed a 1.4%-point increase per year" [419]. This constraint is problematic in hindsight given the growth of renewable's share in total electricity generation has exceeded 1.5% since 2022 [425] and the median forecast exceeds this constraint in virtually every year from 2025 to 2040, reaching their self-imposed 50% maximum as quickly as 2034.

Deployment constraints include economic constraints or macroeconomic feedbacks, constraints based on regional resource availability or the technical potential on solar and wind, and adjustment costs that limit how fast technologies can ramp up or down. These constraints can be quite subtle and difficult to detect. For example, Jenkins et al. [426] found that some MESSAGE-GLOBIOM 1.0 scenarios deployed what appear to be non-optimal amounts of carbon sequestration technologies given the carbon price regimes they applied, despite the model applying cost-optimization. This suggested the IAM algorithm was constraining clean-technology deployment more than build-rate restrictions could explain (e.g. the model did not begin to build CCS early enough for what was required in later decades, suggesting a lack of perfect foresight in its optimization was the constraint).

Integration costs can also limit deployment of variable renewables in some models. Such constraints are an attempt to mimic the problems electricity networks might have deploying transmission, ancillary or storage infrastructure to balance the intermittency of renewables. Fortunately, integration costs have been the subject of an inter-model comparison project [407] which has resulted in updates to infrastructure investment and curtailment equations and parameters in many IAMs. For instance, AR5 versions of REMIND used simplified methods to approximate VRE integration limits. Recent improvements documented in Ueckerdt et al. [427] led to much greater uptake of renewables in REMIND scenarios in AR6 (relative to AR5). The REMIND model is now responsible for the majority of high VRE deployment scenarios.

As mentioned previously, the AIM and GCAM model do not use ETL but instead uses exogenous technology cost trajectories and "share-weighted" logit decision formulations for energy technology market share. As shown in Table 5.1 AIM and GCAM are still susceptible to some of the same constraints as the ETL models, especially the use of outdated costs. They also have an additional avenue for constraint in that the logit choice approach means that even if one technology is cheapest AIM and GCAM will still allocate

some shares to legacy alternatives. This approach is designed to mimic varying regional and consumer preferences, which are relevant, but could result in unrealistic constraints on highly disruptive technologies like solar PV, which has dominated new global energy investments in recent years.

Finally, many IAMs can apply relatively high discount rates ($3-5\%$) which make high capex, low opex investments like wind and solar less attractive in near-term years [428]. If such delays in deployment lead to an overshoot of the scenario climate goal, future carbon removal technologies like Bioenergy with Carbon Capture and Storage (BECCS) are more likely to be used to meet climate targets [429]. Compounding this problem is that many such scenarios allow deployment levels of BECCS that exceed independent estimates of the global BECCS potential as derived from literature reviews [430] and expert elicitation [431].

5.5 Discussion

Three key insights emerge from this analysis. Firstly, probable scenarios — those aligned with observed solar PV and wind deployment trends — feature rapid decarbonisation and high electrification. Secondly, solar PV and wind play a dominant role in enabling low-warming futures within the probable subset; CCS retains a role but to a lesser extent than in the full AR6 ensemble. Finally, provided non-energy-related emissions are also sufficiently mitigated, warming will very likely remain below 2°C.

In terms of representing the growth potential of renewables, AR6 is a clear improvement on AR5, but we believe the scenarios in AR6 still inadequately represent technological change and innovation dynamics [24, 28, 29]. Many employ endogenous technological learning but impose exogenous ad hoc constraints on solar PV growth, such as floor costs, deployment caps, and build limits that do not appear well justified. This leads to

IAM	Floor costs or overly conservative costs on VRE	Build-Rate Constraints	Deployment Constraints	Integration / Flexibility Handling
AIM/CGE	Most AR6 AIM scenario appear to have a USD2010 $1,287 floor cost on solar and $1,838 on onshore wind in 2050	Logit-based market share allocation [432], limits to capital stock turnover, build limits on new technologies	Economic/resource constraints only	Assumes smooth integration each period, exogenous storage investment based on VRE-shares
GCAM	GCAM does not employ ETL but in most AR6 GCAM scenario the cost of solar capex does not go below USD2010 $1,421, or below $1,878 for onshore wind, or below $1,611 for offshore wind after 2050	Logit-based market share allocation, capital turnover limits	Technology lifetimes and S-curve retirements; regional resource availability and policy levers	No diurnal or seasonal resolution in power, so instead imposes an 'integration cost' to VRE generation as penetration rises
IMAGE	No AR6 IMAGE scenario has USD2010 solar capex costs lower than $2,500, or onshore wind lower than $6,690, or offshore wind lower than $20,000	Maximum market share growth parameters, turnover constraints on power plants	Resource/technical potentials; integration cost penalties; policies bound expansion	integration costs based on electricity capacity and modelled with exogenous storage investment based on VRE-shares
MESSAGEix (MESSAGE-GLOBIOM)	No technology costs for MESSAGE scenarios provided in AR6 database	Capacity addition constraints per timestep, dynamic constraints on diffusion [433][2],	Resource/technical potentials; policy constraints (caps, phase-outs)	Transmission/storage in optimization, storage driven by capacity, flexibility equation and VRE-curtailment [407]
POLES	No evidence of floor costs for POLES scenarios in AR6 database but the floor cost of USD2010 $570 mentioned in [434] are never reached	Investment rate and planning limits, technology penetration ceilings	Annual capacity planning and load-curve simulation; renewable potentials	Detailed operational model via load segments and planning rules, only within-day storage based on representative days [407]
REMIND	No floor costs found for REMIND in AR6 scenarios but they are an option mentioned in the REMIND documentation [435]	Limits on annual capacity additions	Endogenous investments with adjustment costs on capex to penalise rapid ramp-up or ramp-down of technologies, capital market imperfections such as limits on regional growth in debts [436]	Hourly dispatch and price-feedbacks via DIETER coupling; aggregated grid integration costs based on VRE share [407]
TIAM	For TIAM scenarios in AR6 the solar capex costs appear to hit temporary limits of USD2010 $220 from 2060 to 2090, $460 for offshore wind and $639 for onshore wind from 2060-2090	explicit maximum annual capacity growth constraints	investment flow constraints	Grid caps, VRE penetration and storage rules, transmission expansion tiers
WITCH	Most WITCH AR6 scenarios appear to have a USD2010 $277 floor cost on capex of solar,$639 on onshore wind, and $407 on offshore wind by 2060	Capital stock turnover limits	Technology-specific flexibility coefficients enforce feasible VRE mixes	Aggregated grid integration cost markups based on total capacity and VRE share [407]

Table 5.1: Comparison of build-rate and deployment constraints, grid integration handling and floor costs employed in selected AR6 IAMs.

systematic exclusion of rapid-deployment pathways that historical data support. Such constraints have not been as readily applied to CCS technologies in AR6, which probably explains why the improbable scenarios deploy CCS at a significantly greater scale as of 2060, than the probable subset. Ironically, it may be that the lack of evidence for learning on CCS [437] that explains this asymmetry: low learning rates mean CCS does not dominate cost-optimising models in the way solar PV does and therefore does not need such constraints.

We recognize that many of the constraints placed on the deployment of solar in the

IAMs reflect phenomena that can be very real. However, the modelling community's attempts to model in detail the many highly complex real world phenomena that drive and constrain the deployment of energy technologies appear to have resulted in systemic biases. Such constraints need to be applied more transparently, and with recognition of their uncertainty, and options that might mitigate such challenges need to also be considered [438]. Comparatively simple models, such as that developed for Way et al. [29], have effectively demonstrated the perils of overusing ad hoc constraints that do not eventuate [410]. Such simpler models, particularly those based on long-term empirical trends [30], should be utilized more diligently to test the more complex IAMs for such systemic biases [406].

The AR6 scenarios are not produced as part of a coordinated controlled experiment, and so do not represent a random statistical sample; nor do they fully explore the space of possibilities. The AR6 database was assembled for a number of reasons including providing a comparable, transparent, harmonized, and common collection of modelled scenarios. They are generated by a wide community running independent experiments, and the database owners aim to be as inclusive as possible. While such inclusiveness is laudable, it has allowed a strong bias to develop in the ensemble. Our results suggest far more scenarios incorporating higher solar PV growth rates are needed to better represent historically supported pathways. The set of AR6 scenarios used for the WGIII report only vet for some historical inconsistencies [422] and too many of the AR6 scenario set omit rapid-solar pathways due primarily to ad hoc growth constraints on solar, and many exhibit nearly linear growth trajectories despite historical evidence that technological change follows S-curve dynamics [36, 91, 292, 439–441].

Concurrently, scenario developers should strive to have AR7 better represent the socio-economic uncertainty embodied in the full range of the Shared Socio-economic Pathways (SSPs), and the IPCC authors are aware of the non-statistical nature of the scenario

ensemble [442]. However, some coordination effort seems prudent given the problems we have unearthed in our analysis. Pirani et al. [443] recommend some form of scientific and technical oversight and an expanded coordination effort for AR7, including a community vetting process and automated checks for quality control. This may not be sufficient unless there is a clear purpose and guiding principles for vetting the collection to ensure a spread of scenarios that better represent our current knowledge and uncertainty. Some 91% of the AR6 scenarios apply the same SSP2 middle-of-the-road set of exogenously defined projections of population, GDP, and other major energy-demand drivers. The IMPs only apply two of the SSPs, with IMP-GS applying SSP2, and IMP-LD, IMP-Ren and IMP-SP applying SSP1, (IMP-Neg is not associated with an SSP). SSP1 has almost half the energy demand of SSP5 in 2100 [421]. This clearly means the AR6 scenarios, and the IMPs, do not adequately test the boundaries of socio-economic uncertainty, which in turn provides less utility for decision-makers seeking to test the resilience of their strategies to such uncertainty, and reduces the utility of forecasting frameworks such as the one presented in this work, to explore what such forecasts might mean for the future.

The probable subset spans SSP1, SSP2, and SSP5, indicating that high economic growth (SSP5) and sustainability-focused (SSP1) futures are both consistent with observed solar and wind growth trajectories. This is a meaningful finding: scenario modellers should not treat SSP2 as a de facto empirical baseline. A richer exploration of SSP1 and SSP5 assumptions in AR7 would improve the utility of the ensemble for decision-makers and allow this kind of tool to more fully characterise socio-economic uncertainty. A better representation of the full range of SSPs would improve our understanding of the implications of socio-economic uncertainty for future energy transitions and warming outcomes. For example, the Way et al. [29] fast transition (shown in Figure 5.1b) applies energy demand growth similar to SSP5 but still achieves full decarbonisation in line with the Paris goals, demonstrating that decarbonisation does not require curtailing economic growth and could

succeed even under demand rebound if solar costs continue their long-term decline.

Our approach has inevitable limitations given the complexity of what is being modelled. Most importantly, we assume that the share of solar PV and wind in future electricity generation is sufficient to characterise the likelihood of all transition scenarios (see Methods 5.2 for technical details). While solar and wind are the fastest-growing low-carbon electricity sources with growth that is correlated to decarbonization rates [28], we do not explicitly model electrification of end-uses, GDP growth, or population changes. We rely instead on aligning our forecasts with scenarios incorporating these factors, assuming that solar and wind growth likelihood is consistent with the associated plausible trajectories in the AR6 database. Despite the AR6 vetting process, most AR6 solar projections fall below the 2.5th percentile of the forecast distribution (see Figure 5.1), indicating that the majority of scenarios fail to capture solar's growth potential. This makes our estimates of net-zero year and cumulative emissions *conservative*. For instance, the median CCS deployment in the probable subset could be lower than our results suggest if the AR6 ensemble included more scenarios with rapid solar and wind growth, rather than relying on CCS to achieve net-zero.

We present a range of likelihood formulations with different mathematical foundations and show they yield qualitatively similar results. The choice of method may depend on the specific research question and desired granularity. For example, if the goal is to assess a scenario with high solar PV growth but low wind growth, a product likelihood requiring both technologies to independently fall within the prediction interval may be more appropriate. The framework can accommodate alternative forecasts methods, whether from S-curve models, expert elicitation, or machine learning, can be applied to other technologies or other scenario ensembles (e.g. AR7, IEA) to explore counter-factuals such as what would happen if solar and wind deployment halted.

In some cases it may be better to analyse useful energy shares as opposed to electricity

generation shares, as the former provides a more holistic overview of the energy system, including hard-to-electrify sectors such as industry and transport. Appendix 5.12 provides a method for calculating useful energy shares from AR6 scenarios, and Appendix 5.13.5 provides results using useful energy shares.

Shares can convey the relative importance of different technologies in the energy mix, but do not say how much energy demand grows by. That is, a scenario can have a high renewable share simply due to an assumed decrease in energy demand. Levels can provide insights into the absolute scale of deployment. In some cases (such as SSP1) the growth in energy demand is low so that while shares are high, levels are not as high as in other scenarios (such as SSP5) where energy demand grows rapidly. If the goal is to understand investment risks for a high solar PV and/or wind growth scenario, it may be better to measure likelihood in levels rather than shares, as the former may be more directly related to investment decisions. We present our method applied to useful energy levels in Appendix 5.13.6.

The main analysis employs a binary consistency criterion (Eq. 5.2 and 5.3) to determine whether a scenario is consistent with the forecast up to a given year. This approach identifies probable scenarios across a given trajectory, but does not provide a direct probability distribution over the AR6 scenario set. Alternative formulations, such as a product likelihood requiring both solar and wind to be individually consistent, yield similar qualitative results with fewer scenarios classified as probable (see Appendix 5.9). Additionally, the trajectory consistency criterion assumes temporal independence across decades and evaluates only fixed AR6 trajectories, not hybrid paths. The main analysis treats all probable scenarios equally rather than weighting by likelihood values (Appendix 5.9.2), though alternative formulations yield qualitatively similar results (Appendix 5.9).

Finally, our results are conditional, not unconditional forecasts. They rest on three key conditions: (i) solar PV and wind deployment continues along historically established

diffusion trajectories and is not derailed by, for example, trade barriers, critical mineral shortages, or sustained policy reversals; (ii) non-energy-related emissions (land use, industry, agriculture) are also sufficiently mitigated; and (iii) the underlying forecasts remain well-calibrated, as demonstrated by back-testing across multiple technologies [284]. If any of these conditions is violated, the warming and net-zero outcomes we report would shift accordingly. Our results therefore represent empirically grounded probability estimates rather than absolute predictions of future energy transitions.

5.5.1 Future research

The probabilistic framework presented here is a first attempt to estimate transition trajectory likelihoods from historical data and scenario analysis. It relies on the deployment of solar and wind to determine likelihood of scenarios that include a broad range of technologies. The forecasts applied here [284] have limitations and could be improved (e.g. through the application of Bayes Model Averaging to better represent S-curve model structure uncertainty). Because the framework is forecast-agnostic, improved or alternative forecasts — whether from updated S-curve models, machine-learning approaches, or expert elicitation — can be substituted directly. We therefore encourage future research to expand and improve on this approach, including adding other promising technologies to the likelihood determination, such as energy storage, and energy efficiency efforts, or even including trends that show predictive strength such as grid infrastructure build out rates, fossil fuel phase-out rates, or electrification rates. For example, energy networks based on molecules, such as geothermal power [444], and thermal energy networks for heating, could prove highly cost effective and gain momentum even in a highly electrified future. If such technological solutions are not included in the calculation of scenario likelihood employed here, the framework will produce likelihoods that are biased against scenarios that employ them at any significant scale.

Expanding from scenario-based to fully probabilistic analysis could also provide new insights for policymakers and investors on optimal strategies to achieve net-zero emissions, limit warming, and promote sustainable economic growth. Moreover, calculating probabilities for pathways allowing for combinations of scenarios could yield a more nuanced and complete understanding of probable transition trajectories. This would require developing methods to combine scenarios while maintaining consistency with empirical data and forecasts, representing a promising avenue for future research.

5.5.2 Concluding remarks

Our approach complements IAMs. Scenario models identify which futures are technically feasible given policy interventions and behavioural change. Empirical forecasts assess which of those futures are historically plausible. IAMs can model policy discontinuities, structural breaks, and technology interactions that S-curve forecasts cannot capture, while forecasts provide empirical grounding and uncertainty quantification that IAMs currently lack. Together, the two identify futures that are both technically feasible and empirically supported.

Probable scenarios limit surface warming to below 2°C with $> 80\%$ probability and almost certainly to below 2.5°C — well below the 4°C that some AR6 scenarios suggest. Net-zero is likely to be achieved before 2068, with a 10% probability by 2050. While these findings are conditional on latest state-of-the-art methods for forecasts, they indicate that a fast transition from fossil fuels to renewable energy is both necessary and achievable within the remaining carbon budgets [445–447]. Further, it demonstrates the need for additional policies to ensure that we reach net-zero by 2050, and limit warming to well below 2°C.

For the IAM modelling community, the analysis exposes systematic bias in the most recent scenario ensemble. Just as biased pessimism may discourage ambition, the finding that rapid deployment is empirically likely could become self-reinforcing. Forecasts can

shape the expectations that produce outcomes, and this performativity should be acknowledged; corrections to any systematically pessimistic anchor are encouraged. High-warming, low-solar pathways are over-represented in the ensemble, particularly for solar PV where median forecasts predict up to twice the deployment of median probable scenarios. In the probable subset, fast, low-warming energy transitions represent the central tendency. We offer this framework as a diagnostic that can improve scenario utility and provide a more grounded foundation for the urgent decisions that will shape the global energy system and the world's climate for decades to come.

For policymakers, our findings send a clear message: to decarbonise our energy systems the policies that have encouraged the deployment of wind and solar have been more successful than many realize. But the optimism of our results depends on such efforts continuing. Policymakers should continue to prioritise dismantling any barriers to rapid solar PV and wind deployment. Much of our heating, cooling, and transport will need to be electrified for a deeply decarbonised future, and an abundance of cheap solar and wind energy will likely be central to making this cost-effective, alongside the necessary grid infrastructure and energy storage investments. The fact that our forecasts, and those of others [29, 440], indicate that deep renewables penetration is not only possible but empirically grounded, should embolden policymakers to act accordingly.

Supplementary Methods

5.6 Overview of Likelihood Methods

This appendix describes all likelihood methods used in this study and their mathematical formulations. Table 5.2 provides a summary of the nine methods (M0–M8) examined. **M0**, the binary trajectory likelihood applied to the sum of solar and wind electricity generation shares, is the primary method presented in the main text (Section 5.2). Methods **M1**–**M3** are binary variants that differ in time scope (yearly vs. trajectory) and technology criterion (sum vs. product). Methods **M4**–**M8** use continuous probability-based likelihood scores rather than binary classification, enabling ranking and weighting of scenarios by their consistency with empirical forecasts. Electricity shares do not account for changes in total electricity generation, which may differ across scenarios. To address this, we also apply **M0** to electricity generation levels (Section 5.11). Electricity generation only describes a portion of the energy system, so we apply **M0** to useful energy shares and levels (Section 5.12). All energy measures yield qualitatively similar results, indicating that the main findings are robust to the choice of energy metric and likelihood method.

The following sections detail: the equivalence of secondary and final electricity shares (Section 5.7); derived probability distributions used for downstream analysis (Section 5.8); the mathematical formulation of each alternative likelihood method M1–M8 (Section 5.9); a proof justifying independent S-curve forecasting (Section 5.10); and the methodology for calculating useful energy shares from AR6 data (Section 5.12).

Table 5.2: Overview of likelihood methods. **M0** is the primary method (main text). *Time scope*: consistency required at each individual year (yearly) or across the full trajectory. *Technology criterion*: solar and wind assessed jointly via their sum (permitting trade-offs) or independently via their product (requiring both to be individually consistent). *Weighting*: binary methods classify scenarios as probable/improbable; continuous methods assign probability-based scores that can be used unweighted (likelihood > 0) or weighted (by the likelihood value). The *Probable (2050)* column reports the number of vetted AR6 scenarios (i.e. out of 1189) classified as probable for trajectories up/in to 2050. All methods use electricity generation shares; **M0** is additionally applied to useful energy shares and useful energy levels as robustness checks.

ID	Formulation	Time scope	Technology	Weighting	Probable (2050)
Binary likelihood (indicator function, Eq. 5.2)					
M0	Trajectory *(main text)*	Full trajectory	Sum	Binary	339
M1	Yearly	Single year	Sum	Binary	592
M2	Trajectory product	Full trajectory	Product	Binary	30
M3	Yearly product	Single year	Product	Binary	65
Continuous likelihood (probability-based, Eq. 5.7)					
M4	Trajectory	Full trajectory	Sum	Unweighted	265
M5	Trajectory	Full trajectory	Sum	Weighted	265
M6	Yearly	Single year	Sum	Unweighted	569
M7	Yearly	Single year	Sum	Weighted	569
M8	Trajectory product	Full trajectory	Product	Unweighted	24

5.7 Secondary electricity and final electricity share equivalence

For technology i in scenario s at time t, the electricity generation share is:

$$\text{Share}^{\text{elec}}_{i,s,t} = \frac{E^{\text{secondary}}_{i,s,t}}{\sum_{j \in \text{elec}} E^{\text{secondary}}_{j,s,t}} \times 100 \tag{5.4}$$

where $E^{\text{secondary}}_{i,s,t}$ is Secondary Energy—Electricity—i and the denominator sums over all electricity technologies (solar, wind, hydro, nuclear, geothermal, biomass, coal, gas, oil).

This equivalence holds because T&D efficiency $\eta^{\text{T\&D}}$ cancels in the ratio:

$$\frac{E^{\text{secondary}}_{i,s,t} \times \eta^{\text{T\&D}}_{s}}{\sum_j E^{\text{secondary}}_{j,s,t} \times \eta^{\text{T\&D}}_{s}} = \frac{E^{\text{secondary}}_{i,s,t}}{\sum_j E^{\text{secondary}}_{j,s,t}} \tag{5.5}$$

5.8 Derived probability distributions

We derive probability distributions for scenario characteristics (net-zero timing, warming level, fossil fuel shares) using indicator functions applied to scenario likelihoods from Eq. 5.7.

5.8.1 Net-zero probability

The probability of achieving net-zero before time T_0 is

$$\mathcal{P}\left(t < T_0 \mid T_0 < \infty\right) = \sum_s \mathcal{L}(Y_t^{(s)}) \cdot \mathbb{I}\left(T_0^{(s)} < T_0\right) \tag{5.6}$$

where $T_0^{(s)}$ is net-zero year for scenario s and t is the analysis year.

5.8.2 Probabilities for all other metrics

The same approach is applied to all other metric or scenario characteristics. That is, for any metric X_s (warming level, fossil fuel share, electricity generation): replace the indicator with $\mathbb{I}(X_s < x)$.

5.9 Alternative Likelihood Formulations

The main text uses a binary trajectory likelihood formulation (**M0**) based on the sum of solar and wind shares, treating all scenarios that fall within the 95% prediction interval of the forecast distribution as equally consistent with empirical data. Binary variants of this measure include yearly (instead of trajectory) and product (instead of sum) formulations (**M1**–**M3**; see Table 5.2). This section describes three non-binary alternatives used in **M4**–**M8**: (1) trajectory likelihood with likelihood measured as the probability mass between two adjacent scenarios, (2) likelihood weighting, which weights scenarios by their likelihood

values rather than treating all probable scenarios equally; (3) product likelihood, requiring both technologies to be individually consistent.

5.9.1 Trajectory likelihood with scenario-adjacent probability mass (M4, M5)

Likelihood of scenario s at time t is the probability mass between adjacent scenarios:

$$\mathcal{L}(Y_t^{(s)}) = \mathbb{P}\left(Y_t^{(s)} < y_t \leq Y_t^{(s')}\right) \tag{5.7}$$

where s' is the scenario immediately above s in combined share, excluding scenarios at the distribution extremes (0th and 100th percentiles). To assess consistency across the full pathway, we multiply scores across time steps using

$$\mathcal{L}(Y^{(s)} \mid T) = \frac{\prod_{t\in\{2030,2040,\ldots,T\}} \mathcal{L}(Y_t^{(s)})}{\sum_s \prod_{t\in\{2030,2040,\ldots,T\}} \mathcal{L}(Y_t^{(s)})} \tag{5.8}$$

Equation 5.8 constitutes our primary analysis: it identifies scenarios that maintain consistency with the forecast across all time steps up to T. We classify a scenario as *probable* when $\mathcal{L}(Y^{(s)} \mid T) > 0$.

5.9.1.1 Interpretation and alternative formulations

The trajectory likelihood can be considered to approximate a conditional probability over the AR6 scenario set under two conditions: (i) the forecast distributions are well-calibrated, as demonstrated by back-testing in Wagenvoort et al. [284]; and (ii) the historical observations contain sufficient information about interdependence between solar and wind Weighting these results yields proper probabilities conditional on AR6 scenarios, and the technologies considered (in the case presented here, solar and wind).

The ordering of scenarios in the unweighted trajectory likelihood may assign zero likelihood to one scenario and non-zero likelihood to another even if both fall at the same forecast percentile. This happens when scenarios ascribe virtually identical amounts of solar and wind. Nevertheless, it admits a direct translation to a true conditional probability measure, since the likelihood of a scenario is directly related to its forecast percentile. To generate a best approximation of the probability distribution, one needs a sufficient number of scenarios that are spread out across forecast distribution. Furthermore, for future research, including more technologies and metrics, would further generate a better approximation.

The binary trajectory likelihood, by contrast, is a coarser measure that classifies scenarios as probable or inconsistent based on a threshold, without distinguishing among scenarios within the consistent category.

5.9.2 Likelihood weighting (M5, M7)

Likelihood weighting applies to both trajectory (**M5**) and yearly (**M7**) formulations. For a scenario characteristic X_s (e.g., cumulative emissions, warming level) at time t:

$$\mathcal{P}(X < x) = \sum_s \frac{\mathcal{L}(Y_t^{(s)})}{\sum_{s'} \mathcal{L}(Y_t^{(s')})} \cdot \mathbb{I}(X_s < x) \tag{5.9}$$

This reweights the distribution toward scenarios that better match the forecast. Likelihood-weighted results are presented in Appendix 5.13.

5.9.3 Product likelihood (M2, M3, M8)

Product likelihood requires both solar and wind energy shares to be individually consistent with their respective forecasts. Binary product versions are **M2** (trajectory) and **M3** (yearly); the non-binary version is **M8**. For technology i (solar or wind) with historical

generation $Y_{i,0:T}$ and scenario s at time t as $Y_{i,t}^{(s)}$, the distributional forecast is:

$$y_{i,t} \sim \mathbb{P}(Y_{i,t} \mid Y_{i,0:T}) \tag{5.10}$$

The cumulative distribution function for scenario s technology i at time t:

$$\mathcal{F}_{s,i,t} = \mathbb{P}\left(Y_{i,t}^{(s)} \leq y_{i,t}\right) \tag{5.11}$$

Probability scenario exceeds forecast:

$$\mathcal{Q}_{s,i,t} = 1 - \mathcal{F}_{s,i,t} \tag{5.12}$$

Likelihood of scenario s for technology i at time t (excluding extremes):

$$\mathcal{L}\left(Y_{s,i,t}\right) = \mathcal{Q}_{s,i,t} - \mathcal{Q}_{s',i,t} \tag{5.13}$$

where s' is the scenario immediately above s: $Y_{s,t} < Y_{s',t}$ with no s'' where $Y_{s,t} < Y_{s'',t} < Y_{s',t}$.

Normalized joint likelihood across both technologies for scenario s at time t:

$$\mathcal{L}\left(Y_{s,t}\right) = \frac{\prod_i \mathcal{L}\left(Y_{s,i,t}\right)}{\sum_s \prod_i \mathcal{L}\left(Y_{s,i,t}\right)} \tag{5.14}$$

Product likelihood imposes a more stringent filtering criterion than sum likelihood. Under sum likelihood, scenario s is deemed probable if its combined solar and wind share $Y_{\text{solar},t}^{(s)} + Y_{\text{wind},t}^{(s)}$ falls within the forecast distribution for total renewable share. Under product likelihood, both $Y_{\text{solar},t}^{(s)}$ and $Y_{\text{wind},t}^{(s)}$ must individually fall within their respective

univariate forecast distributions. Formally, this requires:

$$\mathcal{L}_{\text{product}}\left(Y_{s,t}\right) > 0 \iff \mathcal{L}\left(Y_{s,\text{solar},t}\right) > 0 \wedge \mathcal{L}\left(Y_{s,\text{wind},t}\right) > 0 \tag{5.15}$$

This criterion substantially reduces the set of probable scenarios: only 24 scenarios satisfy product likelihood for pathways to 2050, compared to 265 under sum likelihood. The stringency arises from an empirical mismatch between forecast and scenario characteristics. The distributional forecast $\mathbb{P}(Y_{\text{solar},t} \mid Y_{\text{solar},0:T})$, conditioned on recent historical data exhibiting exponential solar growth, projects substantially higher solar deployment rates than the typical AR6 scenario. Conversely, AR6 scenarios systematically allocate greater relative deployment to wind, with median wind-to-solar ratios exceeding those implied by the forecast. Consequently, product likelihood excludes scenarios with $Y^{(s)}_{\text{wind},t}/Y^{(s)}_{\text{solar},t} > \left(Y_{\text{wind},t}/Y_{\text{solar},t}\right)_{\text{forecast,median}}$ even when $Y^{(s)}_{\text{solar},t} + Y^{(s)}_{\text{wind},t}$ aligns with the combined forecast distribution. This creates a selection effect wherein scenarios satisfying product likelihood exhibit $\mathrm{d}Y^{(s)}_{\text{solar},t}/\mathrm{d}t$ more closely matching empirical trends, typically resulting in faster overall renewable deployment and lower cumulative emissions to net-zero.

Product likelihood is theoretically justified when solar and wind deployment are governed by independent constraints (e.g., distinct resource availability, separate policy mechanisms, non-overlapping supply chains) such that underperformance in one technology cannot be compensated by the other. However, this independence assumption may be violated in energy systems exhibiting portfolio flexibility, technological substitutability, and integrated renewable energy planning. The choice between sum and product likelihood thus reflects assumptions about technological interdependence and system-level flexibility rather than purely statistical considerations.

5.9.4 Comparison and limitations

Product likelihood's stringency may be excessive. Energy systems exhibit portfolio flexibility where underperformance in one technology can be offset by over performance in another. Requiring both solar and wind to be individually consistent eliminates scenarios that are feasible but rely more heavily on one technology. For example, a scenario with exceptional wind growth but moderate solar growth may be eliminated despite the scenario having a combined deployment of the two technologies that is within historical growth patterns.

5.10 Analysing the sum of S-curves

We calculate the likelihood of AR6 scenarios by summing distributional forecasts for solar and wind shares, then comparing this sum to the combined solar and wind shares in AR6 scenarios. This approach assumes that historical data contain sufficient information to capture correlations, substitutions, and competitive dynamics between these two technologies. Wagenvoort et al. [284] use the Bertalanffy-Richards (B-R) curve

$$Q_i(t) = L_i / \left(1 + e^{-k_i \beta (t - t_{0i})}\right)^{1/\beta} \tag{5.16}$$

with $\beta = 2/3$. We use the base-case B-R forecasting approach from [284], specifically the shifted-origin variant without post-hoc widening of prediction intervals. Below we show that the sum of two such curves with different growth rates cannot itself be a B-R curve, justifying our approach of forecasting each technology independently and summing distributions rather than fitting a single combined curve.

5.10.1 The sum of two s-curves need not be a s-curve

Consider the B-R curve of the form $f(t) = L/\left(1 + e^{-k\beta(t-t_0)}\right)^{1/\beta}$ for $\beta > 0$ (i.e. inflection point exists). This is the solution to the ordinary differential equation

$$\frac{df}{dt} = kf\left(1 - \left(\frac{f}{L}\right)^{\beta}\right). \tag{5.17}$$

Let f_1 and f_2 be any two B-R curves with common $\beta > 0$[3] such that

$$\frac{df_i}{dt} = k_i f_i\left(1 - \left(\frac{f_i}{L_i}\right)^{\beta}\right), \qquad i = 1, 2. \tag{5.18}$$

Define their sum $S(t) = f_1(t) + f_2(t)$. Differentiating,

$$\frac{dS}{dt} = k_1 f_1 + k_2 f_2 - k_1 \frac{f_1^{1+\beta}}{L_1^{\beta}} - k_2 \frac{f_2^{1+\beta}}{L_2^{\beta}}. \tag{5.19}$$

Suppose $S(t)$ were itself a B-R curve with parameters (k, L) and the same β. Then S must satisfy

$$\frac{dS}{dt} = kS\left(1 - \left(\frac{S}{L}\right)^{\beta}\right). \tag{5.20}$$

Expanding the right-hand side,

$$\frac{dS}{dt} = k(f_1 + f_2) - \frac{k}{L^{\beta}}(f_1 + f_2)^{1+\beta}. \tag{5.21}$$

For (5.19) and (5.21) to be equal for all t, we require:

1. **Linear terms:** $k_1 f_1 + k_2 f_2 = k(f_1 + f_2)$ which implies $k_1 = k_2 = k$

2. **Nonlinear terms:** The expansion $(f_1 + f_2)^{1+\beta}$ contains mixed terms like $(1+\beta) f_1^{\beta} f_2$ and $(1 + \beta) f_1 f_2^{\beta}$ that do not appear in (5.19). These cross-terms vanish only if

[3] This result extends to cases where $\beta_1 \neq \beta_2$, and to sums of more than two curves by induction.

$f_1(t)/f_2(t)$ is constant for all t, which requires $t_{0,1} = t_{0,2}$. That is, it requires the two technology s-curves to have the same inflection point.

Therefore, $S(t)$ is a B-R curve if and only if $k_1 = k_2$ and $t_{0,1} = t_{0,2}$. Note that the asymptotes L_1 and L_2 need not be equal. When $k_1 = k_2 = k$ and $t_{0,1} = t_{0,2} = t_0$, the sum has the form $S(t) = (L_1 + L_2)/\left(1 + e^{-k\beta(t-t_0)}\right)^{1/\beta}$.

Since historical data show $k_{\text{solar}} \neq k_{\text{wind}}$ [29, 210], the sum of solar and wind s-curves cannot be represented as a single s-curve, justifying our approach of forecasting the two technologies independently and summing the distributions. In our use case that means sums can exceed 100%. We treat these as though solar PV and wind provide 100% of the electricity generation.

5.11 Calculating electricity generation levels

The main text analyses electricity generation *shares*, which normalise for demand growth but cannot distinguish whether high shares arise from rapid deployment or from low total demand. To address this, we also compare AR6 scenarios to forecasts of absolute electricity generation at the final energy level. This section describes how we calculate technology-specific electricity generation in PWh (1 PWh = 10^{15} Wh) from AR6 data.

Methodology. For each electricity technology i in scenario s at time t, we convert secondary energy to final electricity generation:

$$E^{\text{final}}_{i,s,t} = E^{\text{secondary}}_{i,s,t} \times \eta^{\text{T\&D}}_{s,t} \tag{5.22}$$

where $E^{\text{secondary}}_{i,s,t}$ is the technology-specific generation reported as Secondary Energy—Electricity—i (in EJ/yr), and $\eta^{\text{T\&D}}_{s,t}$ is the scenario- and year-specific transmission and distribution efficiency.

T&D efficiency determination. The T&D efficiency is calculated from the ratio of aggregate final to secondary electricity:

$$\eta_{s,t}^{\text{T\&D}} = \frac{\text{FE—Electricity}_{s,t}}{\text{SE—Electricity}_{s,t}} \tag{5.23}$$

When this ratio is unavailable or implausible (> 0.95, indicating a data quality issue), we use a default value of $\eta^{\text{T\&D}} = 0.87$, representing the median global electricity T&D efficiency. This is consistent with the T&D efficiency used in Stage 1 of the useful energy calculation (Section 5.12.1).

Unit conversion. AR6 data are reported in EJ/yr. We convert to PWh using 1 EJ = 1/3.6 PWh.

Forecast comparison and likelihood. The empirical consistency filter (Eq. 5.2) applies identically: we replace electricity generation shares with absolute generation levels (in PWh) and compare to distributional forecasts of solar and wind electricity generation.

5.12 Calculating useful energy shares

5.12.1 Calculating useful energy shares from AR6 scenarios

The main text analyses electricity generation shares calculated at the secondary energy (SE) level (Section 5.2). Electricity only describes a portion of total energy use. As a robustness check we repeat the analysis using *useful energy* shares, which account for other sectors and fuel types, and account for transmission and distribution (T&D) losses and end-use conversion efficiency. The AR6 database reports Secondary Energy (SE, generation at power plants) and Final Energy (FE, delivery to end-users after transmission/distribution losses) [27]. Neither quantifies useful energy (mechanical work, heat, light delivered after

end-use conversion). We convert AR6 data to useful energy using sector- and fuel-specific efficiencies from Cullen and Allwood [448], De Stercke [449], and Way et al.[29].

We measure technology contributions as shares of total useful energy rather than absolute values. Absolute generation values can increase indefinitely with demand growth, potentially leading to asymptotic behaviour where all technologies appear to succeed. Shares normalise for demand growth, providing a meaningful comparison across scenarios with different energy demand trajectories.

Methodology Calculation proceeds in two stages corresponding to distinct energy pathways: electricity (from Final Energy with technology shares from Secondary Energy) and direct fuel use (from Final Energy). A key methodological point: FE already includes transmission and distribution losses, avoiding double-counting.

Stage 1: Electricity conversion. For electricity, the AR6 database reports technology-specific generation as Secondary Energy (SE) at power plants and, in some cases, technology-specific or aggregate Final Energy (FE) delivered to end-users. We calculate useful energy for each technology directly, applying T&D losses and end-use efficiency. For each electricity technology i:

$$E_{i,s,t}^{\text{useful}} = \begin{cases} \text{FE}_{i,s,t} \times \eta_{\text{elec}}^{\text{end-use}} & \text{if tech-specific FE available} \\ E_{i,s,t}^{\text{secondary}} \times \eta_s^{\text{T\&D}} \times \eta_{\text{elec}}^{\text{end-use}} & \text{otherwise} \end{cases} \tag{5.24}$$

where $E_{i,s,t}^{\text{secondary}}$ is generation from technology i (Secondary Energy—Electricity—i), $\text{FE}_{i,s,t}$ is technology-specific final energy (Final Energy—Electricity—i, when available), and $\eta_{\text{elec}}^{\text{end-use}} = 0.90$ is the electrical end-use conversion efficiency.

T&D efficiency determination. When technology-specific FE is unavailable, we calculate a scenario-specific T&D efficiency $\eta_s^{\text{T\&D}}$ hierarchically:

1. **Sector-specific FE available**: If sector-disaggregated Final Energy is reported (FE|Sector|Electricity for Transportation, Residential and Commercial, and Industry), calculate T&D efficiency from the ratio of summed sector-specific FE to total SE:

$$\eta_s^{\text{T\&D}} = \frac{\sum_{\text{sector}} \text{FE|Sector|Electricity}_{s,t}}{\text{SE|Electricity}_{s,t}} \tag{5.25}$$

2. **Aggregate FE available**: If sector-specific data is unavailable but total electricity FE and SE are reported, use their ratio:

$$\eta_s^{\text{T\&D}} = \frac{\text{FE|Electricity}_{s,t}}{\text{SE|Electricity}_{s,t}} \tag{5.26}$$

3. **Default empirical value**: If no FE data is available, use the IPCC-vetted empirical average:

$$\eta_s^{\text{T\&D}} = 0.87 \tag{5.27}$$

   This represents the median global electricity T&D efficiency across diverse systems.

This direct calculation approach accounts for T&D losses exactly once. When technology-specific FE is available, T&D losses are already embedded. When using SE, we apply the scenario-specific or empirical T&D efficiency. All electricity technologies experience identical T&D losses and end-use efficiency, so the calculation method does not introduce systematic bias across technologies.

Stage 2: Fossil fuel conversion. For non-electric energy carriers (solids, liquids, gases) consumed in sector j:

$$E_{k,j,s,t}^{\text{useful}} = E_{k,j,s,t}^{\text{final}} \times \eta_{k,j}^{\text{conversion}} \tag{5.28}$$

where $E_{k,j,s,t}^{\text{final}}$ is final energy delivery of fuel type k to sector j (reported as Final

Energy—j—k in the AR6 database), and $\eta_{k,j}^{\text{conversion}}$ is the sector-and-fuel-specific conversion efficiency.

Conversion efficiency parameters. We adopt the sector- and fuel-specific conversion efficiencies from Way et al. [29], based on Cullen and Allwood [448] and De Stercke [449] (Table 5.3). Our key methodological improvement is recognising that Final Energy already incorporates transmission and distribution losses, eliminating the need for separate T&D calculations and preventing double-counting.

Table 5.3: Conversion efficiency parameters for calculating useful energy from AR6 scenario data. Values adopted from Way et al. [29], based on Cullen and Allwood [448] and De Stercke [449]. For electricity, end-use efficiency is applied directly to Final Energy (which already includes T&D losses).

Sector	Energy Carrier	Efficiency	End-use Device
Transportation			
	Liquids	0.20	Internal combustion engines
	Gases	0.20	Compressed natural gas vehicles
	Electricity	0.90	Electric motors and batteries
Residential and Commercial			
	Solids	0.65	Combustion stoves and boilers
	Liquids	0.65	Oil heating systems
	Gases	0.65	Natural gas heating and cooking
	Electricity	0.90	Electric heating, cooling, lighting, appliances
Industry			
	Solids	0.35	High-temperature process heat
	Liquids	0.35	Industrial furnaces and boilers
	Gases	0.35	Process heat and steam generation
	Electricity	0.90	Electric motors, electrolysis, electric furnaces

Useful energy shares for forecast comparison. After calculating useful energy from all technologies and fuel types, we compute each technology's share of total useful energy—the metric compared to forecasts:

$$\text{Share}_{i,s,t}^{\text{useful}} = \frac{E_{i,s,t}^{\text{useful}}}{\sum_{i'} E_{i',s,t}^{\text{useful}}} \times 100 \tag{5.29}$$

where the denominator sums useful energy across all electricity technologies (solar, wind, hydro, nuclear, geothermal, biomass, coal, gas, oil) and non-electric fuel categories (solids,

liquids, gases). The empirical consistency filter (Eq. 5.2) applies identically: we replace electricity generation shares with useful energy shares and compare to useful energy distributional forecasts. Because T&D efficiency cancels in electricity share ratios but not in useful energy shares (where the denominator includes non-electricity sectors at different efficiencies), the two metrics can yield different scenario subsets.

Worked Example: Useful Energy Calculation This example demonstrates the calculation sequence for AIM/CGE 2.0 SFCM_SSP2_Ref_1p5Degree in 2050.

Step 1: Calculate T&D efficiency. Check if technology-specific FE is available. For this scenario, only aggregate FE is reported:

$$\text{SE—Electricity} = 243.82 \text{ EJ/yr}$$

$$\text{FE—Electricity} = 197.53 \text{ EJ/yr}$$

$$\eta^{\text{T\&D}} = \frac{197.53}{243.82} = 0.8102 \text{ (19\% losses)}$$

Step 2: Calculate useful energy directly for each technology. Apply T&D efficiency and end-use efficiency to SE:

$$\text{SE—Electricity—Solar} = 53.85 \text{ EJ/yr}$$

$$E_{\text{solar}}^{\text{useful}} = 53.85 \times 0.8102 \times 0.90 = 39.27 \text{ EJ/yr}$$

$$\text{SE—Electricity—Wind} = 88.81 \text{ EJ/yr}$$

$$E_{\text{wind}}^{\text{useful}} = 88.81 \times 0.8102 \times 0.90 = 64.76 \text{ EJ/yr}$$

Step 3: Calculate useful energy shares. Including fossil fuels (industrial solids: 5.02 EJ/yr) and all other technologies, total useful energy is 236 EJ/yr. This yields the

final metric for likelihood analysis:

$$\text{Share}_{\text{solar}}^{\text{useful}} = \frac{39.27}{236} = 0.166 \ (16.6\%)$$
$$\text{Share}_{\text{wind}}^{\text{useful}} = \frac{64.76}{236} = 0.274 \ (27.4\%)$$

Step 4: Fossil fuel useful energy (unchanged). Industrial solids as exemplar:

$$\text{FE—Industry—Solids} = 14.33 \text{ EJ/yr}$$
$$E_{\text{solids}}^{\text{useful}} = 14.33 \times 0.35 = 5.02 \text{ EJ/yr}$$

Figure 5.5 illustrates the conversion sequence. For electricity, Final Energy (197.53 EJ/yr) already reflects T&D losses from Secondary Energy (243.82 EJ/yr). Applying the 90% end-use efficiency yields 177.78 EJ/yr useful electricity—an effective 72.9% conversion rate from SE. Solar provides 39.27 EJ/yr (16.6% of total useful energy) and wind provides 64.76 EJ/yr (27.4%). Industrial fossil fuels convert at their sector-specific 35% efficiency directly from Final Energy.

Supplementary Results

5.13 Supplementary Results

This appendix presents supplementary results complementing the main text. We first demonstrate that our findings are robust across different likelihood formulations (Appendix 5.13.1). We then present additional results using the main text method (Appendix 5.13.2), results using alternative continuous likelihood formulations M4–M8 (Appendix 5.13.3). As electricity generation shares do not account for changes in total electricity generation, which may differ across scenarios, we also present results using electricity generation levels

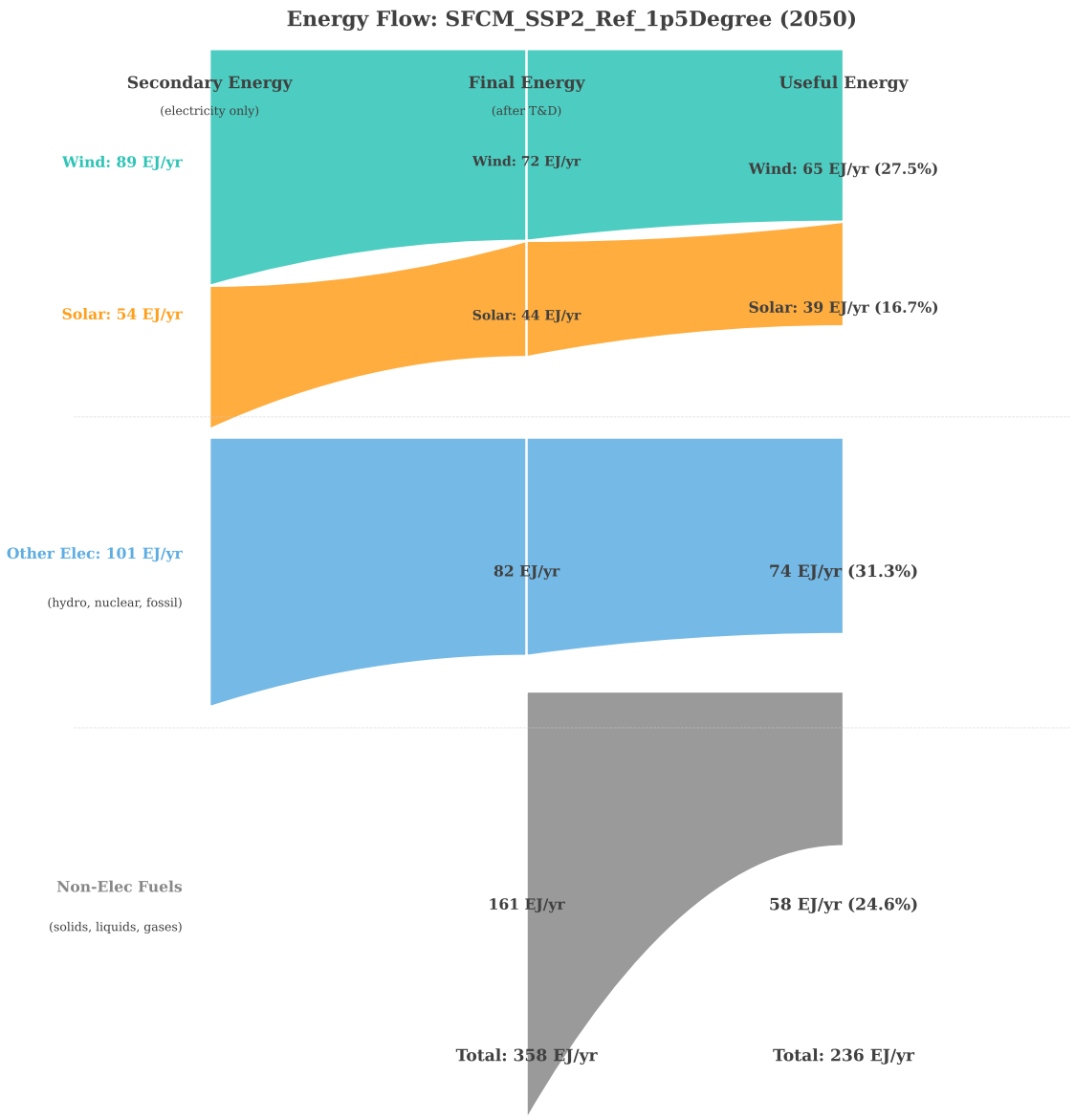


Figure 5.5: Sankey-style energy conversion for AIM/CGE 2.0 SFCM_SSP2_Ref_1p5Degree (2050). Band thicknesses are proportional to energy (EJ/yr). Solar and wind generation at power plants (Secondary Energy) experience T&D losses before delivery to end-users (Final Energy, 197.53 EJ/yr from 243.82 EJ/yr SE, 81% efficient). End-use conversion ($\eta^{\text{end-use}} = 0.90$) applied to FE yields 177.78 EJ/yr useful electricity—an effective 72.9% conversion rate from SE. Industrial solids convert final energy at 35% efficiency. Solar provides 16.6% of total useful energy (39.27 EJ/yr) and wind provides 27.4% (64.76 EJ/yr).

(Appendix 5.13.4). Electricity generation only describes a portion of the energy system, so we also present results using useful energy shares and levels (Appendices 5.13.5 and 5.13.6). See Table 5.2 for all method definitions.

5.13.1 Robustness across binary likelihood formulations (M0–M3)

We first demonstrate that results are robust to the choice of likelihood formulation. Table 5.4 compares results using the four binary likelihood methods (M0–M3, Table 5.2). The Cohen's d values in each column indicate the difference between probable and improbable subsets for each metric, and the Mann-Whitney U test results (stars) indicate whether differences are statistically significant. Results are broadly similar across formulations, confirming that the main conclusions are not sensitive to the specific binary likelihood definition.

Across the metrics, results most often show similar directional differences between probable and improbable scenarios regardless of formulation. When they do not, the differences are often not statistically significant, suggesting those metrics are less robust to the choice of formulation. Notably, **M2** (pathway product) shows higher median CCS deployment (GtCO2/yr) in probable scenarios for 2050, while all other formulations show significantly lower CCS deployment. Such differences may arise because yearly likelihood measures are more sensitive to year-to-year variations, while trajectory measures are more sensitive to overall consistency across the full trajectory.

Stringency of likelihood measures We measure stringency in terms of difference between probable and improbable scenarios for key metrics, as measured by Cohen's d. The main text binary trajectory likelihood measure (**M0**) has the second-highest stringency for both energy and socioeconomic metrics, while the trajectory product binary measure (**M2**) has the highest. Yearly measures (**M1** and **M3**) have lower stringency than trajectory

Table 5.4: Robustness of energy metric comparisons (probable vs. improbable) across four binary likelihood methods. M0 corresponds to the main paper. All methods use a binary split (likelihood > 0 vs. $= 0$) with no resampling or weighting. The consistent direction of ΔMedian across methods confirms findings are not sensitive to the specific binary likelihood definition.

		M0		**M1**		**M2**		**M3**	
Metric (Unit)	**Year**	ΔMed	d	ΔMed	d	ΔMed	d	ΔMed	d
Total electricity generation (PWh)	2030	1.0	0.21^{***}	1.0	0.21^{***}	3.6	$\mathbf{0.89}^{***}$	3.6	$\mathbf{0.89}^{***}$
	2050	10.9	$\mathbf{1.06}^{***}$	10.2	$\mathbf{0.94}^{***}$	16.4	$\mathbf{1.79}^{***}$	15.6	$\mathbf{1.40}^{***}$
	2100	24.2	$\mathbf{0.77}^{***}$	27.3	$\mathbf{1.06}^{***}$	48.4	$\mathbf{1.35}^{***}$	30.7	$\mathbf{0.81}^{***}$
Fossil fuels electricity share (%)	2030	-15.0	$\mathbf{-1.55}^{***}$	-15.0	$\mathbf{-1.55}^{***}$	-10.8	$\mathbf{-0.64}^{***}$	-10.8	$\mathbf{-0.64}^{***}$
	2050	-25.1	$\mathbf{-1.82}^{***}$	-31.0	$\mathbf{-2.47}^{***}$	-16.5	$\mathbf{-1.42}^{***}$	-16.9	$\mathbf{-1.19}^{***}$
	2100	-5.5	$\mathbf{-1.07}^{***}$	-26.7	$\mathbf{-2.07}^{***}$	-4.5	$\mathbf{-1.03}^{***}$	-2.6	$\mathbf{-0.78}^{***}$
Solar + Wind electricity share (%)	2030	14.9	$\mathbf{1.61}^{***}$	14.9	$\mathbf{1.61}^{***}$	11.8	$\mathbf{1.01}^{***}$	11.8	$\mathbf{1.01}^{***}$
	2050	31.4	$\mathbf{2.33}^{***}$	33.1	$\mathbf{3.16}^{***}$	25.7	$\mathbf{1.75}^{***}$	25.5	$\mathbf{1.52}^{***}$
	2100	20.9	$\mathbf{1.70}^{***}$	30.8	$\mathbf{3.47}^{***}$	21.4	$\mathbf{1.78}^{***}$	-3.3	0.42^{**}
Carbon Capture and Storage ($GtCO_2$/yr)	2030	0.2	-0.06^{ns}	0.2	-0.06^{ns}	-0.0	-0.12^{ns}	-0.0	-0.12^{ns}
	2050	0.9	-0.10^{ns}	1.0	-0.16^{ns}	2.9	0.42^{ns}	3.0	0.46^{ns}
	2100	-3.6	-0.46^{***}	-3.7	-0.25^{***}	0.5	0.08^{ns}	-2.7	-0.25^{***}
H_2 from fossil fuels (PWh)	2030	0.0	-0.12^{ns}	0.0	-0.12^{ns}	0.1	-0.22^{ns}	0.1	-0.22^{ns}
	2050	0.2	-0.02^{ns}	0.5	-0.04^{ns}	0.3	0.01^{ns}	0.1	0.06^{ns}
	2100	-3.9	$\mathbf{-0.57}^{***}$	-4.5	-0.37^{***}	-3.1	$\mathbf{-0.52}^{*}$	1.0	-0.05^{ns}
H_2 from solar (PWh)	2030	0.0	$\mathbf{0.68}^{***}$	0.0	$\mathbf{0.68}^{***}$	0.0	0.19^{***}	0.0	0.19^{***}
	2050	0.3	$\mathbf{1.37}^{**}$	0.3	$\mathbf{3.29}^{***}$	–	–	-0.0	$\mathbf{-1.22}^{ns}$
	2100	-0.0	0.43^{ns}	0.2	$\mathbf{1.17}^{***}$	–	–	-0.1	-0.37^{ns}

Binary methods (no weighting): **M0** (main paper): Trajectory, sum-based solar+wind likelihood — binary split (likelihood > 0 means scenario is within the 95% PI at every year up to t). **M1**: single-year binary split using sum-based likelihood (scenario within 95% PI in year t only). **M2**: Trajectory, product-based likelihood ($P(\text{solar}) \times P(\text{wind})$) — binary split. **M3**: single-year binary split using product-based likelihood.

ΔMed = median(probable) − median(improbable) in reported units. d = Cohen's d (**bold** if $|d| > 0.5$). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ^{ns} not significant (Mann–Whitney U, one-sided).

measures (**M0** and **M2**), plausibly because they permit more year-to-year variation while trajectory measures require consistency across the full trajectory.

5.13.2 Additional Results

Metrics comparing probable and improbable scenarios Table 5.5 accompanies Figure 5.2, comparing the probable scenarios with the improbable scenarios for 6 key energy transition metrics.

Mann-Whitney U tests show significant differences between probable and improbable scenarios for six metrics, with probable scenarios characterised by higher electrification, higher solar and wind electricity shares, and lower fossil fuel shares. Probable scenarios are

Table 5.5: Statistical comparison of key energy metrics between probable and improbable AR6 climate pathways

Metric (Unit)	Year	ΔMedian	MW stat	Cohen's d
Total electricity generation (PWh)	2030	1.0	194827***	0.21
	2050	10.9	224599***	**1.06**
	2100	24.2	199935***	**0.77**
Fossil fuels electricity share (%)	2030	-15.0	28894***	**-1.55**
	2050	-25.1	27557***	**-1.82**
	2100	-5.5	54476***	**-1.07**
Solar + Wind electricity share (%)	2030	14.9	322915***	**1.61**
	2050	31.4	271304***	**2.33**
	2100	20.9	255861***	**1.70**
Carbon Capture and Storage ($GtCO_2$/yr)	2030	0.2	152627***	-0.06
	2050	0.9	137883*	-0.10
	2100	-3.6	90756***	-0.46
H_2 from fossil fuels (PWh)	2030	0.0	41135	-0.12
	2050	0.2	64113***	-0.02
	2100	-3.9	42151***	**-0.57**
H_2 from solar (PWh)	2030	0.0	1231***	**0.68**
	2050	0.3	565**	**1.37**
	2100	-0.0	445	0.43

Note: ΔMedian = difference in median values (Probable minus Improbable). MW stat = Mann-Whitney U statistic with superscript significance (*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). Cohen's d = effect size (**bold** if $|d| > 0.5$, indicating medium to large effect).

therefore more aligned with rapid decarbonisation trajectories and earlier net-zero targets. CCS deployment does not differ significantly between the two subsets until 2050, suggesting that CCS is not a key near-term distinguishing factor based on the probabilistic forecast of solar and wind electricity generation shares. This may be because scenarios often limit solar deployment, while we determine likelihood in solar+wind, thus not capturing this. Hydrogen plays a more central role in probable scenarios, with significantly higher shares of hydrogen from solar PV and similar shares of hydrogen from fossil fuels—consistent with a transition pathway reliant on green rather than blue hydrogen.

Additional metrics for probable vs. improbable scenarios Complementing Figures 5.2 and 5.3, we show additional metrics comparing probable and improbable scenarios in Figure 5.6 and Table 5.6. Table 5.6 also compares probable scenarios to all

AR6 scenarios, including yearly GHG emissions.

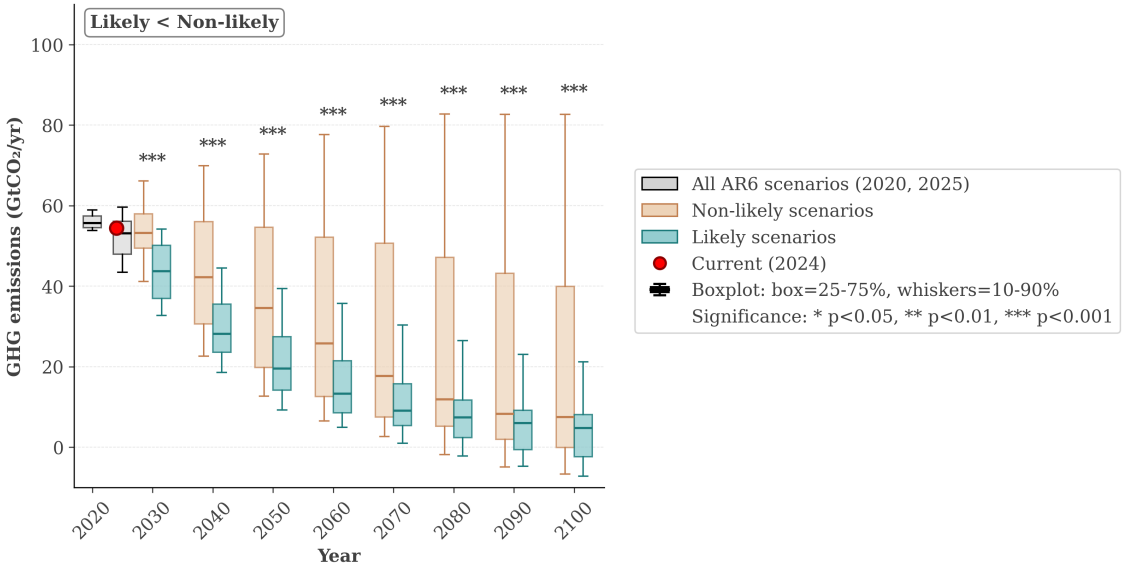


Figure 5.6: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan) for probable scenarios up to year T for GHG emissions. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test.

Mann-Whitney U tests show significant differences between probable and improbable scenarios for emission, warming, and net-zero metrics. Probable scenarios feature more ambitious decarbonisation trajectories and earlier net-zero targets. The difference between probable and all vetted AR6 scenarios is significant for most of the same metrics, as expected given that the probable subset is drawn from the full ensemble.

Table 5.7 shows additional statistical tests comparing probable and improbable scenarios for the 12 metrics shown in main paper Figure 5.2 and appendix Figure 5.6.

Table 5.6: Extended statistical comparison for all metrics: probable trajectories versus improbable and versus all AR6 scenarios

		Probable vs Improbable			Probable vs All AR6	
Metric (Unit)	**Year**	**ΔMedian**	**MW stat**	**Cohen's d**	**ΔMedian**	**t-stat**
Total electricity generation (PWh)	2030	1.0	194827***	0.21	0.4	2.48*
	2050	10.9	224599***	**1.06**	7.9	12.85***
	2100	24.2	199935***	**0.77**	21.6	8.40***
Fossil fuels electricity share (%)	2030	-15.0	28894***	**-1.55**	-10.1	-17.03***
	2050	-25.1	27557***	**-1.82**	-14.6	-28.52***
	2100	-5.5	54476***	**-1.07**	-2.7	-19.77***
Solar + Wind electricity share (%)	2030	14.9	322915***	**1.61**	9.9	18.40***
	2050	31.4	271304***	**2.33**	23.6	29.93***
	2100	20.9	255861***	**1.70**	17.3	24.25***
Carbon Capture and Storage ($GtCO_2$/yr)	2030	0.2	152627***	-0.06	0.1	-0.73
	2050	0.9	137883*	-0.10	0.5	-1.36
	2100	-3.6	90756***	-0.46	-2.5	-6.28***
H_2 from fossil fuels (PWh)	2030	0.0	41135	-0.12	0.0	-0.71
	2050	0.2	64113***	-0.02	0.1	-0.17
	2100	-3.9	42151***	**-0.57**	-1.7	-5.33***
H_2 from solar (PWh)	2030	0.0	1231***	**0.68**	0.0	1.58
	2050	0.3	565**	**1.37**	0.3	2.87*
	2100	-0.0	445	0.43	-0.0	0.84
GHG emissions ($GtCO_2eq$/yr)	2030	-6.2	79056***	**-0.79**	-4.9	-9.88***
	2050	-14.9	80568***	**-0.92**	-7.8	-13.51***
	2100	-2.7	110209***	**-0.66**	-1.9	-10.88***
CO_2 emissions ($GtCO_2$/yr)	2030	-4.7	80564***	**-0.74**	-4.1	-9.30***
	2050	-12.6	81753***	**-0.90**	-6.9	-12.91***
	2100	-1.6	110010***	**-0.64**	-0.6	-10.26***
Net-zero CO_2 year (year)	2030	-2.5	52203**	-0.25	-2.0	-2.31*
	2050	-2.5	52203**	-0.25	-2.0	-2.31*
	2100	-2.5	52203**	-0.25	-2.0	-2.31*
Surface warming (°C)	2030	-0.0	113410***	-0.39	-0.0	-4.42***
	2050	-0.1	71427***	**-0.95**	-0.1	-12.78***
	2100	-0.3	88378***	**-0.81**	-0.2	-12.22***

Note: Comprehensive comparison for 10 metrics (6 energy + 4 climate) across three time periods. Left columns: Probable vs Improbable trajectories with full statistical testing. Right columns: Probable vs All AR6 scenarios showing median difference and t-test (parametric test for mean differences). ΔMedian = difference in median values (Probable minus comparison group). MW stat = Mann-Whitney U statistic (non-parametric, tests median differences). t-stat = Welch's t-test statistic (parametric, tests mean differences). All statistics shown with superscript significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Cohen's d = effect size (**bold** if $|d| > 0.5$, indicating medium to large effect).

Comparison with the larger set of AR6 database scenarios Here we recreate the main paper findings using a larger set of over 2000 AR6 database scenarios, that is, including the scenarios present in the AR6 database but not included in the vetted AR6 set used for the WGIII report. Figure 5.7 complements Figure 5.2. These results are similar to the main text (Figure 5.2), with some differences in the distribution of probable scenarios and the statistical significance of certain metrics. These differences likely reflect the inclusion of additional scenarios that failed historical vetting. Notably,

Table 5.7: Comprehensive statistical test results: methodological transparency for all metrics

Metric (Unit)	Year	n_L	n_{NL}	ΔMedian	Cohen's d	MW stat	t-stat	KS stat
Total electricity generation (PWh)	2030	491	695	1.0	0.21	194827***	3.65***	0.19***
	2050	337	849	10.9	**1.06**	224599***	17.18***	0.55***
	2100	334	852	24.2	**0.77**	199935***	11.57***	0.31***
Fossil fuels electricity share (%)	2030	493	696	-15.0	**-1.55**	28894***	-26.90***	0.79***
	2050	339	850	-25.1	**-1.82**	27557***	-35.87***	0.73***
	2100	336	853	-5.5	**-1.07**	54476***	-21.72***	0.56***
Solar + Wind electricity share (%)	2030	493	696	14.9	**1.61**	322915***	28.38***	0.94***
	2050	339	850	31.4	**2.33**	271304***	42.43***	0.80***
	2100	336	853	20.9	**1.70**	255861***	31.48***	0.70***
Carbon Capture and Storage ($GtCO_2$/yr)	2030	491	531	0.2	-0.06	152627***	-1.03	0.29***
	2050	339	742	0.9	-0.10	137883*	-1.70	0.21***
	2100	336	752	-3.6	-0.46	90756***	-7.86***	0.32***
H_2 from fossil fuels (PWh)	2030	339	244	0.0	-0.12	41135	-1.51	0.22***
	2050	273	406	0.2	-0.02	64113***	-0.24	0.30***
	2100	272	431	-3.9	**-0.57**	42151***	-7.58***	0.38***
H_2 from solar (PWh)	2030	20	68	0.0	**0.68**	1231***	2.10*	0.80***
	2050	6	111	0.3	**1.37**	565**	3.02*	0.69**
	2100	6	119	-0.0	0.43	445	0.88	0.42
GHG emissions ($GtCO_2eq$/yr)	2030	336	853	-6.2	**-0.79**	79056***	-12.98***	0.40***
	2050	336	853	-14.9	**-0.92**	80568***	-16.61***	0.36***
	2100	336	853	-2.7	**-0.66**	110209***	-12.49***	0.23***
CO_2 emissions ($GtCO_2$/yr)	2030	336	853	-4.7	**-0.74**	80564***	-12.16***	0.38***
	2050	336	853	-12.6	**-0.90**	81753***	-16.02***	0.36***
	2100	336	853	-1.6	**-0.64**	110010***	-11.94***	0.23***
Net-zero CO_2 year (year)	2030	249	488	-2.5	-0.25	52203**	-3.22**	0.13**
	2050	249	488	-2.5	-0.25	52203**	-3.22**	0.13**
	2100	249	488	-2.5	-0.25	52203**	-3.22**	0.13**
Surface warming (°C)	2030	336	853	-0.0	-0.39	113410***	-5.97***	0.21***
	2050	336	853	-0.1	**-0.95**	71427***	-16.37***	0.43***
	2100	336	853	-0.3	**-0.81**	88378***	-14.75***	0.32***

Note: Complete statistical test results for methodological transparency (Probable vs Improbable comparison). n_L = sample size probable trajectories, n_{NL} = sample size improbable trajectories. ΔMedian = difference in medians (Probable minus Improbable). Cohen's d = effect size (**bold** if $|d| > 0.5$, indicating medium to large effect). MW stat = Mann-Whitney U statistic (non-parametric test for median differences). t-stat = Welch's t-test (parametric test for mean differences). KS stat = Kolmogorov-Smirnov statistic (tests whether entire distributions differ). All statistics shown with superscript significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

probable scenarios in the larger AR6 database set diverge from improbable scenarios in CCS deployment at 2060, suggesting the vetting process excludes scenarios with higher CCS deployment that are inconsistent with historical growth trajectories.

Figure 5.8 complements Figures 5.3, showing that applying the methodology to the larger set yields similar results for warming, emission, and net-zero outcomes. Figure 5.8 also complements Figure 5.4 by comparing the distribution of SSPs in the larger AR6 database set to those that are probable. Findings in both cases are similar to the results presented in the main text for the vetted scenarios. A slightly larger proportion of SSP4 scenarios are in the probable subset (blue). This indicates that socio-economic pathways

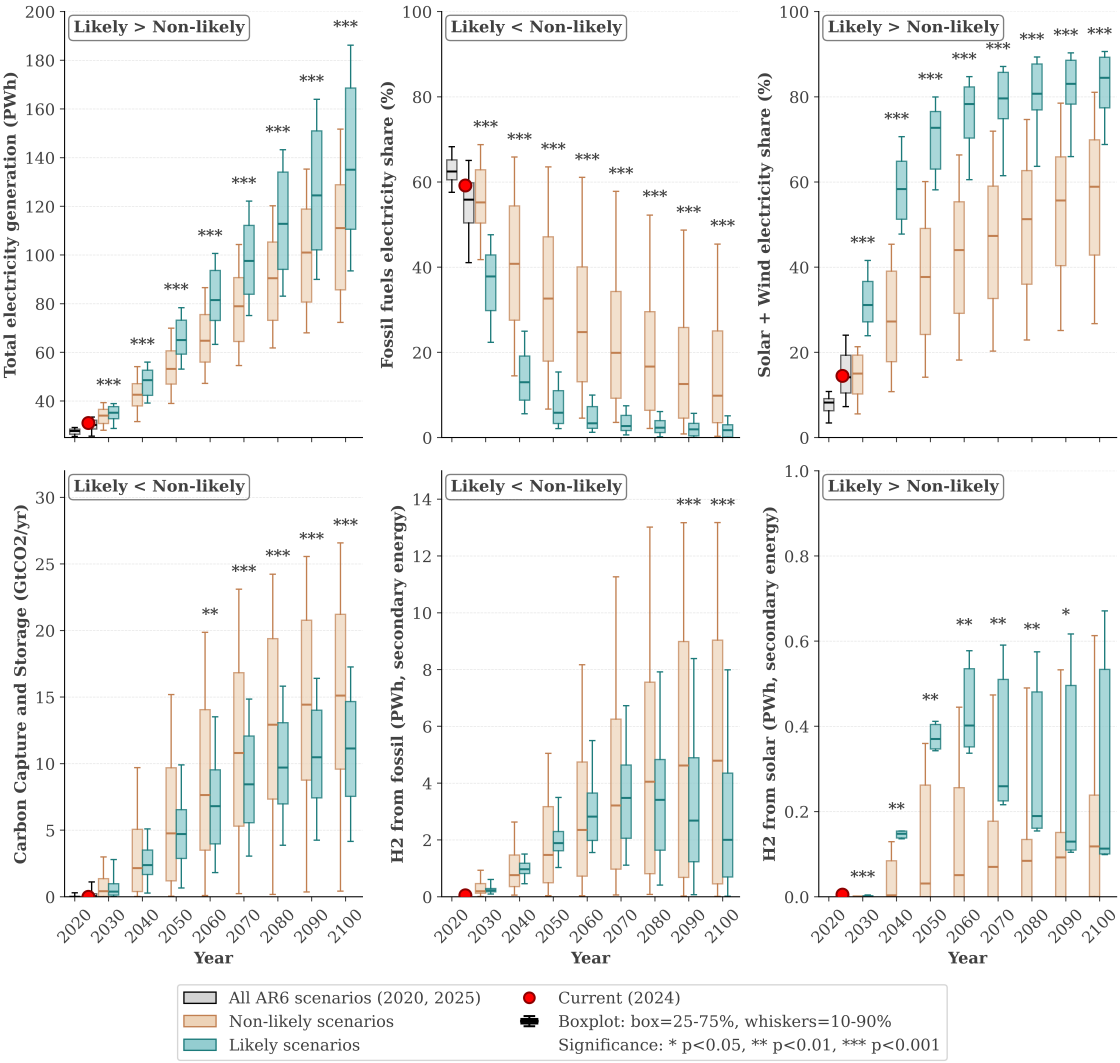


Figure 5.7: Energy transitions for probable and improbable scenarios using the full set of AR6 database scenarios with the binary likelihood method used in the main text. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** Electricity share. **(b)** Solar PV share of electricity. **(c)** Wind share of electricity. **(d)** Fossil fuel share of electricity. **(e)** Carbon Capture and Storage ($GtCO_2$/yr). **(f)** Hydrogen share of final energy from solar PV. **(g)** Hydrogen share of final energy from fossil fuels.

with heightened inequality and regional disparities can also produce scenarios consistent with historical solar and wind growth trajectories, despite being excluded from the vetted set analysed in the main text.

For the larger set of AR6 database scenarios, REMIND is the most common model (marginally ahead of WITCH) in the probable subset, while for the vetted set, REMIND and WITCH trade places. In both sets, these two models produce by far the largest

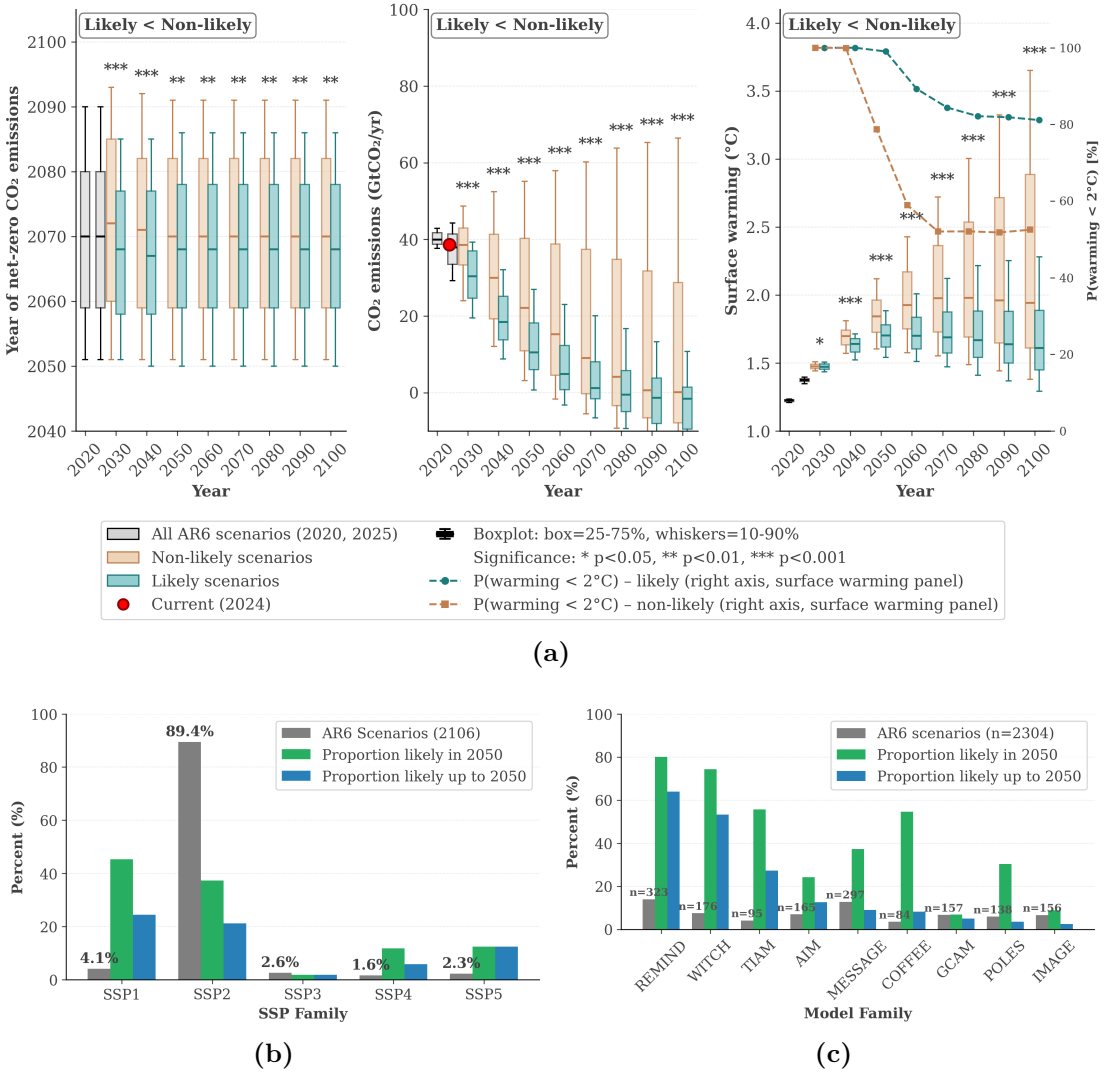


Figure 5.8: **(a)** Comparison of warming, emissions, and net-zero outcomes for larger AR6 database set. Colours are consistent with main text. Teal indicates probable scenarios, tan indicates improbable scenarios. Grey box-plots represent AR6 scenarios in 2020 and 2025, with the red dot marking the 2024 observed levels. **(b)** Distribution of SSPs in the full set of AR6 scenarios. **(c)** Distribution of SSPs in the probable subset of the full AR6 set. Grey is the proportion of each SSP in the full AR6 set, green the proportion within each SSP that is probable in 2050, and blue the proportion within each SSP that is probable across the trajectory up to 2050. Numbers in brackets are different, as some scenarios do not contain SSP information.

proportion of probable scenarios. The distribution of SSPs and models in the probable subset differs from the full AR6 database set, suggesting that certain SSPs and models are more likely to produce scenarios consistent with historical solar and wind growth trajectories.

5.13.3 Results using alternative likelihood formulations (M4–M8)

Here we show results using the continuous likelihood formulations **M4**–**M8** (Table 5.2, methods in Appendix 5.9). These results provide a robustness check and show how different likelihood formulations can produce different distributions of probable scenarios while generally supporting similar conclusions about differences between probable and improbable scenarios. The choice of likelihood formulation should be based on the specific research question and context.

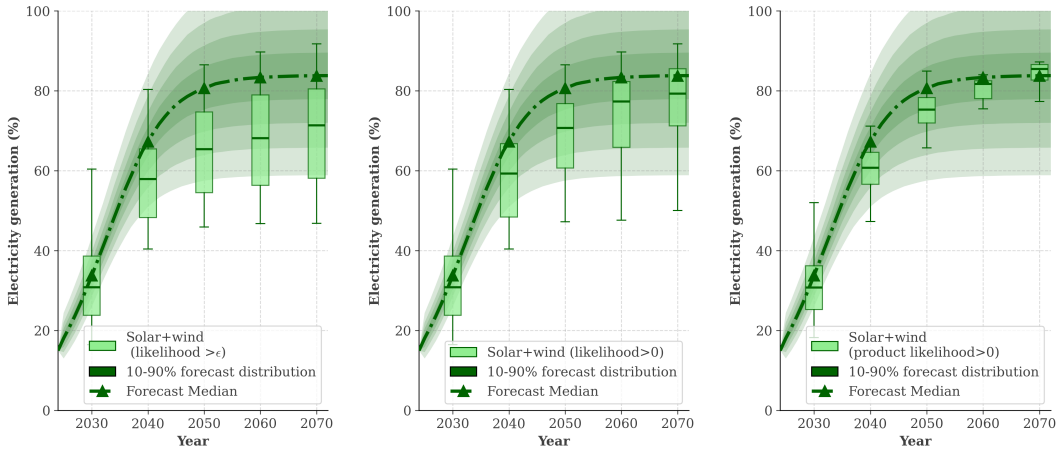
We show results using likelihood-weighted distributions of probable scenarios measured per year, as well as product and trajectory likelihoods.

Figure 5.9 shows the forecast of solar+wind electricity shares compared to the probable AR6 scenarios using alternative likelihood formulations.

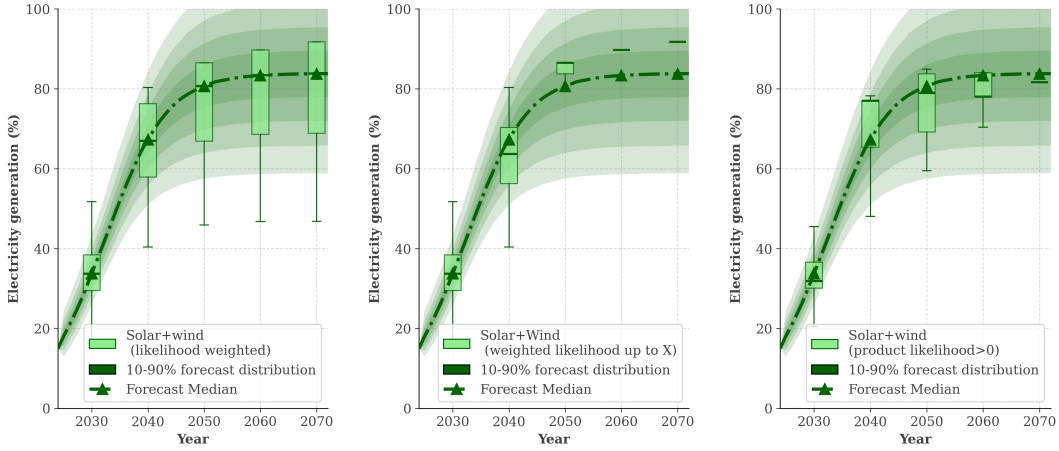
Table 5.8 mirrors Table 5.5 from the main text but uses alternative likelihood formulations.

A notable finding is that for the metric *Carbon Capture and Storage ($GtCO_2/yr$)*, the weighted trajectory likelihood (**M6**) shows significantly lower CCS in probable scenarios compared to improbable scenarios, while the unweighted trajectory likelihood (**M4**) shows a less significant difference. This suggests that scenarios with higher CCS deployment have comparatively lower consistency scores; weighting by likelihood can therefore change the distribution of the probable subset and reveal differences not apparent when treating all probable scenarios equally. This is potentially a result of CCS being used to compensate for constrained solar deployment in ambitious scenarios, or to generate overshoot solutions using negative emissions technologies such as bioenergy carbon capture and storage.

Table 5.9 compares the distribution of metrics shown in Table 5.5 in the main paper to the same metrics under alternative likelihood formulations, testing whether distributions



(a) Unweighted yearly likelihood **(b)** Unweighted trajectory likelihood **(c)** Unweighted product trajectory likelihood



(d) Weighted yearly likelihood **(e)** Weighted trajectory likelihood **(f)** Weighted product trajectory likelihood

Figure 5.9: Comparison of solar + wind electricity generation shares for AR6 scenarios to the probabilistic forecast, using alternative likelihood formulations. AR6 scenarios shown as box plots in 10-year increments from 2030 to 2070. Boxes are the IQR and whiskers the $10-90\%$. Green fans are the forecasts from $10-90\%$. Line is the median forecast. **(a)** AR6 scenarios probable in a given year. **(b)** AR6 scenarios with probable trajectory for sum of solar and wind electricity share. **(c)** AR6 scenarios with probable trajectories for solar and wind electricity shares independently. **(d)** AR6 scenarios weighted yearly likelihood. **(e)** AR6 scenarios weighted trajectory likelihood. **(f)** AR6 scenarios weighted product likelihood.

and means differ significantly across formulations. The results show that distributions expectedly differ when using alternative likelihood formulations, as the weighting of scenarios changes the composition of the probable subset. However, the direction of differences between probable and improbable scenarios is consistent across formulations, and statistical significance is often preserved. That is, while the mean values of metrics for the probable

Table 5.8: Robustness of energy metric comparisons (probable vs. improbable) across five likelihood methods (M4–M8). Direction of ΔMedian is consistent across all five methods in 83% of energy and 72% of supplementary year×metric cells, confirming findings are not sensitive to the choice of likelihood methodology.

		M4		M5		M6		M7		M8	
Metric (Unit)	**Year**	ΔMed	d	ΔMed	d	ΔMed	d	ΔMed	d	ΔMed	d
Total electricity generation (PWh)	2030	0.4	0.08^{ns}	0.5	0.05^{ns}	0.4	0.08^{ns}	0.5	0.05^{ns}	1.0	0.27^{***}
	2050	11.8	$\mathbf{1.05}^{***}$	30.2	$\mathbf{2.83}^{***}$	9.1	$\mathbf{0.83}^{***}$	25.0	$\mathbf{1.62}^{***}$	18.6	$\mathbf{1.66}^{***}$
	2100	19.1	0.46^{***}	31.2	$\mathbf{1.14}^{***}$	11.7	0.46^{***}	33.6	$\mathbf{0.64}^{***}$	57.3	$\mathbf{2.00}^{***}$
Fossil fuels electricity share (%)	2030	-16.1	$\mathbf{-1.55}^{***}$	-19.2	$\mathbf{-2.08}^{***}$	-16.1	$\mathbf{-1.55}^{***}$	-19.2	$\mathbf{-2.08}^{***}$	-11.5	$\mathbf{-0.77}^{***}$
	2050	-22.3	$\mathbf{-1.49}^{***}$	-26.1	$\mathbf{-1.91}^{***}$	-28.8	$\mathbf{-1.92}^{***}$	-33.8	$\mathbf{-2.36}^{***}$	-16.0	$\mathbf{-1.37}^{***}$
	2100	-3.4	$\mathbf{-0.87}^{***}$	-5.1	$\mathbf{-1.10}^{***}$	-15.7	$\mathbf{-1.13}^{***}$	-18.0	$\mathbf{-1.24}^{***}$	-4.5	$\mathbf{-1.03}^{***}$
Solar + Wind electricity share (%)	2030	14.1	$\mathbf{1.50}^{***}$	17.0	$\mathbf{2.11}^{***}$	14.1	$\mathbf{1.50}^{***}$	17.0	$\mathbf{2.11}^{***}$	10.9	$\mathbf{0.90}^{***}$
	2050	26.5	$\mathbf{1.55}^{***}$	42.3	$\mathbf{2.83}^{***}$	29.7	$\mathbf{2.14}^{***}$	45.0	$\mathbf{2.89}^{***}$	26.1	$\mathbf{1.71}^{***}$
	2100	17.3	$\mathbf{1.12}^{***}$	27.1	$\mathbf{2.25}^{***}$	27.3	$\mathbf{1.17}^{***}$	37.8	$\mathbf{1.50}^{***}$	21.6	$\mathbf{1.83}^{***}$
Carbon Capture and Storage ($GtCO_2$/yr)	2030	0.4	0.29^{ns}	0.3	0.22^{ns}	0.4	0.29^{ns}	0.3	0.22^{ns}	0.1	-0.06^{ns}
	2050	1.6	0.13^{ns}	0.0	-0.36^{ns}	0.9	-0.15^{ns}	0.1	-0.30^{ns}	2.0	0.38^{ns}
	2100	-1.6	-0.19^{*}	-10.0	$\mathbf{-1.70}^{***}$	-0.7	0.05^{ns}	-6.2	-0.34^{***}	-0.8	-0.29^{ns}
H_2 from fossil fuels (PWh)	2030	0.1	0.16^{ns}	0.0	0.10^{ns}	0.1	0.16^{ns}	0.0	0.10^{ns}	0.1	-0.07^{ns}
	2050	0.2	0.17^{ns}	-1.4	$\mathbf{-0.89}^{***}$	0.4	0.07^{ns}	-1.1	-0.36^{***}	-0.2	-0.12^{ns}
	2100	-2.1	-0.34^{***}	-4.1	$\mathbf{-1.39}^{***}$	-2.4	-0.20^{**}	-5.5	$\mathbf{-0.64}^{***}$	-3.7	$\mathbf{-0.91}^{*}$
H_2 from solar (PWh)	2030	0.0	0.50^{***}	0.0	$\mathbf{0.66}^{***}$	0.0	0.50^{***}	0.0	$\mathbf{0.66}^{***}$	0.0	0.47^{***}
	2050	0.3	$\mathbf{1.20}^{***}$	-0.0	0.11^{***}	0.3	$\mathbf{2.26}^{***}$	0.3	$\mathbf{2.16}^{***}$	–	–
	2100	0.4	$\mathbf{1.03}^{**}$	-0.0	0.09^{ns}	0.0	$\mathbf{0.72}^{***}$	0.1	$\mathbf{0.67}^{***}$	–	–

Methods: **M4**: cumulative unweighted trajectory likelihood using sum-based solar+wind likelihood, equal weights. **M5**: as M4 but resampled proportional to normalized cumulative likelihood. **M6**: binary split at each year independently (likelihood > 0 vs $= 0$), equal weights. **M7**: as M6 but resampled proportional to that year's likelihood. **M8**: cumulative trajectory likelihood using product-based likelihood ($P(\text{solar}) \times P(\text{wind})$), equal weights.

ΔMed = median(probable) − median(improbable) in reported units. d = Cohen's d (**bold** if $|d| > 0.5$). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\text{ns}}$ not significant (Mann–Whitney U, one-sided).

subset may change with different formulations, the overall conclusions are relatively robust.

A majority of the differences in mean (measured by Cohen's D) are negligible or small ($|d| < 0.2$ or $0.2 < |d| < 0.5$). Across the 6 metrics and 5 methods, 15 out of 18 comparisons show the same direction of difference between probable and improbable scenarios. H_2 from fossil fuels (PWH) in 2050 in **M5**[4], **M6**[5], and **M7**[6] show small negative differences between probable and improbable scenarios, while **M4** shows a small but insignificant positive difference (see Table 5.8).

Table 5.10 shows the same comparison as Table 5.8 but for the additional metrics shown in Figure 5.6 and Table 5.6.

[4]Weighted trajectory sum likelihood
[5]Unweighted yearly sum likelihood
[6]Weighted yearly sum likelihood

Table 5.9: Pairwise comparison of probable distributions for energy metrics across likelihood methods. Cohen's d comparing the raw (unweighted) probable scenarios of each method against M4. Values near zero confirm that the probable sets are robust to the method choice.

		Cohen's d (M4 vs. Mx)			
Metric (Unit)	**Year**	d_{45}	d_{46}	d_{47}	d_{48}
Total electricity generation (PWh)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	-0.18^{*}
	2050	0.00^{ns}	0.37^{***}	0.37^{***}	$\mathbf{-0.80}^{***}$
	2100	0.00^{ns}	0.26^{**}	0.26^{**}	$\mathbf{-1.28}^{*}$
Fossil fuels electricity share (%)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	-0.26^{*}
	2050	0.00^{ns}	-0.36^{***}	-0.36^{***}	0.43^{ns}
	2100	0.00^{ns}	$\mathbf{-0.56}^{***}$	$\mathbf{-0.56}^{***}$	$\mathbf{1.04}^{**}$
Solar + Wind electricity share (%)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	0.13^{ns}
	2050	0.00^{ns}	0.30^{***}	0.30^{***}	$\mathbf{-0.73}^{**}$
	2100	0.00^{ns}	$\mathbf{0.73}^{***}$	$\mathbf{0.73}^{***}$	$\mathbf{-1.40}^{**}$
Carbon Capture and Storage ($GtCO_2$/yr)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	0.20^{**}
	2050	0.00^{ns}	0.23^{***}	0.23^{***}	-0.33^{*}
	2100	0.00^{ns}	-0.19^{*}	-0.19^{*}	0.12^{ns}
H_2 from fossil fuels (PWh)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	0.13^{ns}
	2050	0.00^{ns}	0.09^{ns}	0.09^{ns}	0.24^{ns}
	2100	0.00^{ns}	-0.24^{*}	-0.24^{*}	$\mathbf{0.58}^{ns}$
H_2 from solar (PWh)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	-0.05^{ns}
	2050	0.00^{ns}	-0.04^{ns}	-0.04^{ns}	–
	2100	0.00^{ns}	$\mathbf{0.75}^{*}$	$\mathbf{0.75}^{*}$	–

d_{45}: M4 vs. M5 (pathway weighted). d_{46}: M4 vs. M6 (yearly unweighted). d_{47}: M4 vs. M7 (yearly weighted). d_{48}: M4 vs. M8 (product likelihood).

Computed on raw (pre-resampling) probable distributions using a two-sided Mann–Whitney U test. **bold** if $|d| > 0.5$. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\text{ns}}$ not significant.

Table 5.11 shows the same comparison as Table 5.9 but for the additional metrics shown in Figure 5.6 and Table 5.6.

A majority of the differences in mean (measured by Cohen's D) are negligible or small ($|d| < 0.2$ or $0.2 < |d| < 0.5$). There are 13/18 cases in the same direction.

Most to least stringent likelihood formulations As shown in Table 5.2, the product trajectory likelihood (**M8**) is the most stringent, as it requires scenarios to be probable at each time step for both solar and wind electricity shares separately. Yearly likelihoods (**M6** and **M7**) are less stringent than trajectory likelihoods, as they consider scenarios probable at each time step independently, without requiring consistency across time steps.

Table 5.10: Robustness of supplementary metric comparisons (probable vs. improbable) across five likelihood methods (M4–M8). Direction of ΔMedian is consistent across all five methods in 83% of energy and 72% of supplementary year×metric cells, confirming findings are not sensitive to the choice of likelihood methodology.

		M4		**M5**		**M6**		**M7**		**M8**	
Metric (Unit)	**Year**	ΔMed	d	ΔMed	d	ΔMed	d	ΔMed	d	ΔMed	d
CO_2 emissions ($GtCO_2$/yr)	2030	-10.2	**-1.42*****	-11.9	**-1.79*****	-10.2	**-1.42*****	-11.9	**-1.79*****	–	–
	2050	-15.4	**-1.17*****	-21.5	**-1.69*****	-18.7	**-1.13*****	-27.0	**-1.52*****	–	–
	2100	-3.4	**-0.68*****	-1.8	**-0.64*****	-7.0	**-0.75*****	-7.6	**-0.78*****	–	–
GHG emissions ($GtCO_2$eq/yr)	2030	-12.8	**-1.46*****	-14.1	**-1.84*****	-12.8	**-1.46*****	-14.1	**-1.84*****	–	–
	2050	-18.0	**-1.15*****	-23.7	**-1.60*****	-21.3	**-1.14*****	-29.4	**-1.49*****	–	–
	2100	-3.5	**-0.68*****	-1.8	**-0.64*****	-8.0	**-0.76*****	-8.0	**-0.79*****	–	–
Warming (°C)	2030	0.0	0.04^{ns}	0.0	0.04^{ns}	0.0	0.04^{ns}	0.0	0.04^{ns}	0.0	0.18^{ns}
	2050	-0.2	**-1.27*****	-0.3	**-2.11*****	-0.1	**-0.95*****	-0.3	**-1.50*****	-0.2	**-1.23*****
	2100	-0.4	**-0.92*****	-0.5	**-1.35*****	-0.5	**-0.68*****	-0.6	**-0.86*****	-0.4	**-1.16****
Net-zero CO_2 year (year)	2030	-7.0	-0.50***	-8.0	**-0.50*****	-7.0	-0.50***	-8.0	**-0.50*****	1.0	0.05^{ns}
	2050	-8.0	**-0.51*****	-16.0	**-1.25*****	-8.0	-0.43***	-19.0	**-0.90*****	-9.0	**-0.66****
	2100	-5.5	-0.43***	-13.0	**-1.41*****	3.0	0.26^{ns}	-10.0	-0.23**	-12.5	**-0.53**^{ns}

Methods: M4: cumulative unweighted trajectory likelihood using sum-based solar+wind likelihood, equal weights. **M5**: as M4 but resampled proportional to normalized cumulative likelihood. **M6**: binary split at each year independently (likelihood > 0 vs $= 0$), equal weights. **M7**: as M6 but resampled proportional to that year's likelihood. **M8**: cumulative trajectory likelihood using product-based likelihood (P(solar) $\times$ P(wind)), equal weights.

ΔMed = median(probable) − median(improbable) in reported units. d = Cohen's d (**bold** if $|d| > 0.5$). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\text{ns}}$ not significant (Mann–Whitney U, one-sided).

Table 5.11: Pairwise comparison of probable distributions for supplementary metrics across likelihood methods. Cohen's d comparing the raw (unweighted) probable scenarios of each method against M4. Values near zero confirm that the probable sets are robust to the method choice.

		Cohen's d (M4 vs. Mx)			
Metric (Unit)	**Year**	d_{45}	d_{46}	d_{47}	d_{48}
CO_2 emissions ($GtCO_2$/yr)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	–
	2050	0.00^{ns}	-0.38***	-0.38***	–
	2100	0.00^{ns}	-0.38**	-0.38**	–
GHG emissions ($GtCO_2$eq/yr)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	–
	2050	0.00^{ns}	-0.36***	-0.36***	–
	2100	0.00^{ns}	-0.38**	-0.38**	–
Warming (°C)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	-0.13^{ns}
	2050	0.00^{ns}	**-0.53*****	**-0.53*****	0.24^{ns}
	2100	0.00^{ns}	**-0.63*****	**-0.63*****	**0.51**^{ns}
Net-zero CO_2 year (year)	2030	0.00^{ns}	0.00^{ns}	0.00^{ns}	-0.30**
	2050	0.00^{ns}	-0.20**	-0.20**	0.30^{ns}
	2100	0.00^{ns}	-0.42***	-0.42***	0.23^{ns}

d_{45}: M4 vs. M5 (pathway weighted). d_{46}: M4 vs. M6 (yearly unweighted). d_{47}: M4 vs. M7 (yearly weighted). d_{48}: M4 vs. M8 (product likelihood).

Computed on raw (pre-resampling) probable distributions using a two-sided Mann–Whitney U test. **bold** if $|d| > 0.5$. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\text{ns}}$ not significant.

In terms of difference from improbable scenarios (measured by Cohen's D), the weighted trajectory likelihood (**M5**) shows the largest differences for most metrics, while the unweighted trajectory likelihood (**M4**) shows the smallest differences for energy metrics (Table 5.8) and middle values for socioeconomic metrics (Table 5.10).

Comparing yearly and trajectory likelihoods Figure 5.10 complements Figure 5.1 by comparing the binary trajectory likelihood measure used in the main text to the yearly consistency measure (the main text method without requiring consistency across time steps). Differences between probable and improbable scenarios are similar to the main text.

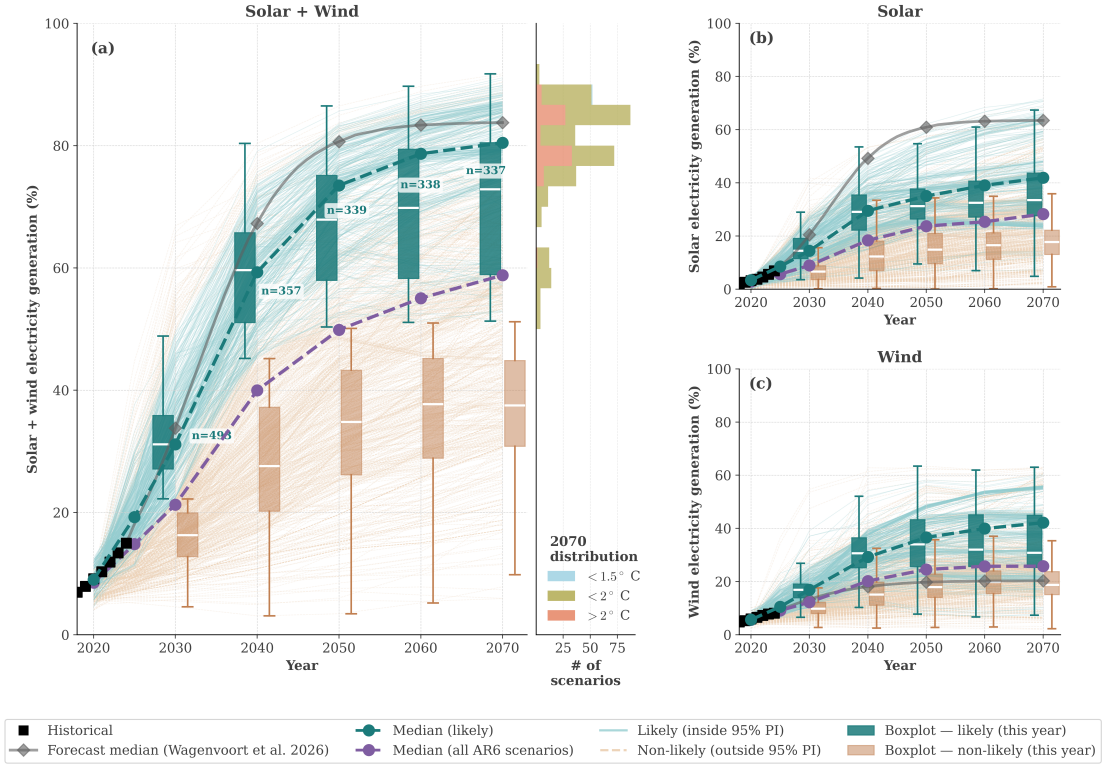


Figure 5.10: Comparison of solar + wind electricity generation shares for AR6 scenarios to the probabilistic forecast, using yearly consistency measure. probable scenarios are shown in teal, improbable scenarios are shown in tan. **(a)** Solar PV share of electricity. **(b)** Wind share of electricity. **(c)** Solar + wind share of electricity.

Notably, the yearly likelihood measure allows scenarios to leave the 95% prediction interval at certain time steps while being consistent at others, whereas the trajectory measure requires consistency across the full trajectory. The yearly measure therefore produces more probable scenarios. In particular, the number of probable scenarios increases from 2040

to 2100, indicating that some scenarios reach a probable level of solar + wind electricity share but assume lower growth rates than the forecast.

Trajectory likelihood measures Figure 5.11 complements Figure 5.1 by comparing the binary trajectory likelihood measure used in the main text to the unweighted trajectory likelihood measure presented in Appendix 5.9.

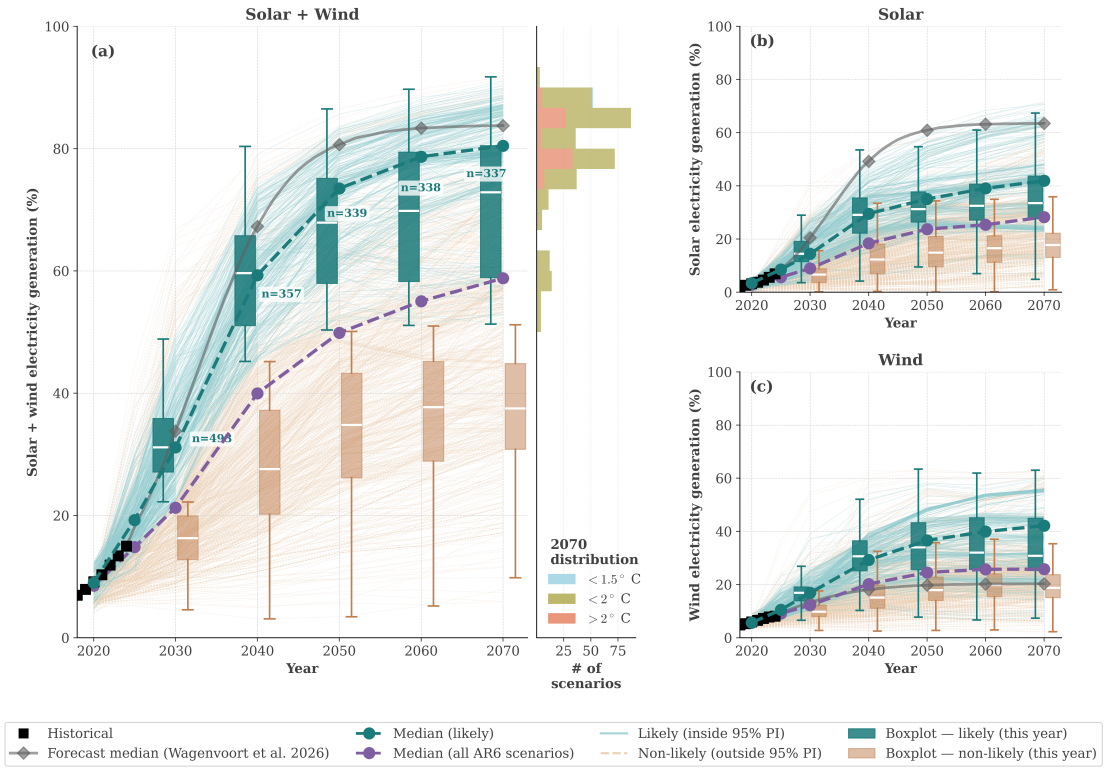


Figure 5.11: Comparison of solar + wind electricity generation shares for AR6 scenarios to the probabilistic forecast, using unweighted trajectory likelihood measure. probable scenarios are shown in teal, improbable scenarios are shown in tan. **(a)** Solar PV share of electricity. **(b)** Wind share of electricity. **(c)** Solar + wind share of electricity.

There are slightly fewer probable scenarios here compared to the main text. Overall conclusions are similar, and the result holds across all formulations presented. 18% of probable scenarios under this metric achieve net-zero by 2050, compared to 10% for the main text binary trajectory measure. Similarly, the probability of limiting warming to $< 2°C$ is $> 90\%$ for probable scenarios under this metric, compared to $> 80\%$ for the main text binary trajectory measure. CCS deployment between the two subsets is less significantly different under the unweighted trajectory measure: probable scenarios show

significantly lower median CCS deployment from 2080 onward, compared to 2060 under the main text measure.

5.13.4 Results using electricity generation (PWh)

The main text focuses on shares. Here we show results using electricity generation levels (using the main text method of distinguishing probable vs. improbable scenarios), complementing the main text results using shares. We use an exogenously determined 2% p.a. growth rate as a minimum, to account for growing energy demand after technology-driven growth has stalled [284]. The resulting S-curve is thus growing in absolute terms even when the share of solar and wind electricity generation is constant, as total electricity generation grows. The S-curve (in analytic form, excluding noise terms) is defined as

$$Q_i(t) = \frac{L_i e^{gt}}{\left(1 + e^{-k_i \beta (t - t_{0,i})}\right)^{1/\beta}} \tag{5.30}$$

where $Q_i(t)$ is the electricity generation level of technology i at time t, L_i is the maximum electricity generation level, k_i is the growth rate , β is a shape parameter, and $t_{0,i}$ is the inflection point, and g is the growth rate of total electricity generation. We set $g = 0.02$ following [284]. This is a conservative assumption for near-term projections, as it is below the historical growth rate of total electricity generation of 3% p.a. (Based on analysis of data from OWID [210]). Nonetheless, it allows us to test whether the main text results using shares are robust to differences in total electricity generation across scenarios. Forecasts are made using electricity generation data from OWID [210].

Figure 5.12 compares solar and wind electricity generation levels in AR6 scenarios to the probabilistic forecast, complementing Figure 5.1 in the main text, which uses shares. Comparing them shows how scenarios can have similar shares but different levels of solar and wind electricity generation, depending on the total electricity generation. For example,

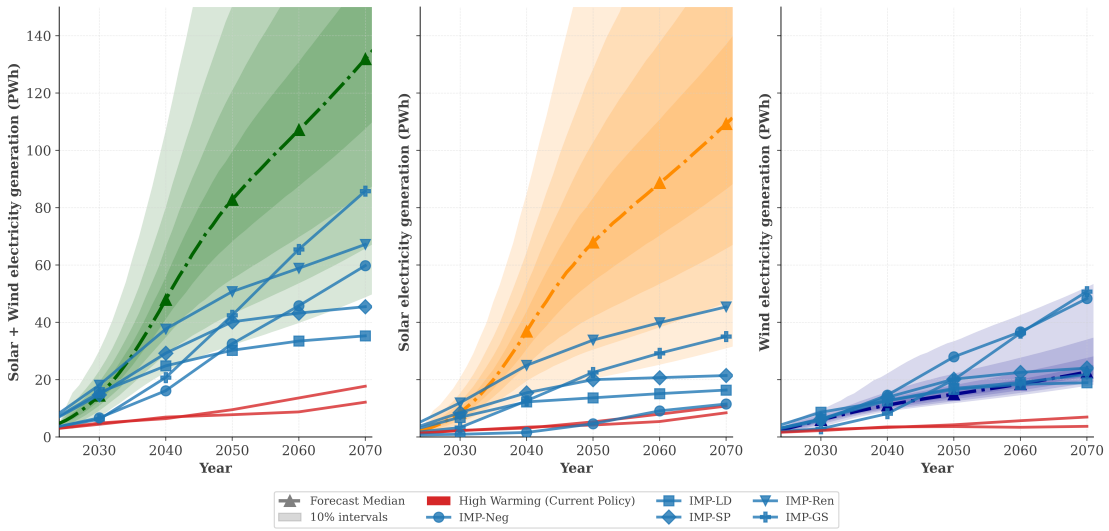
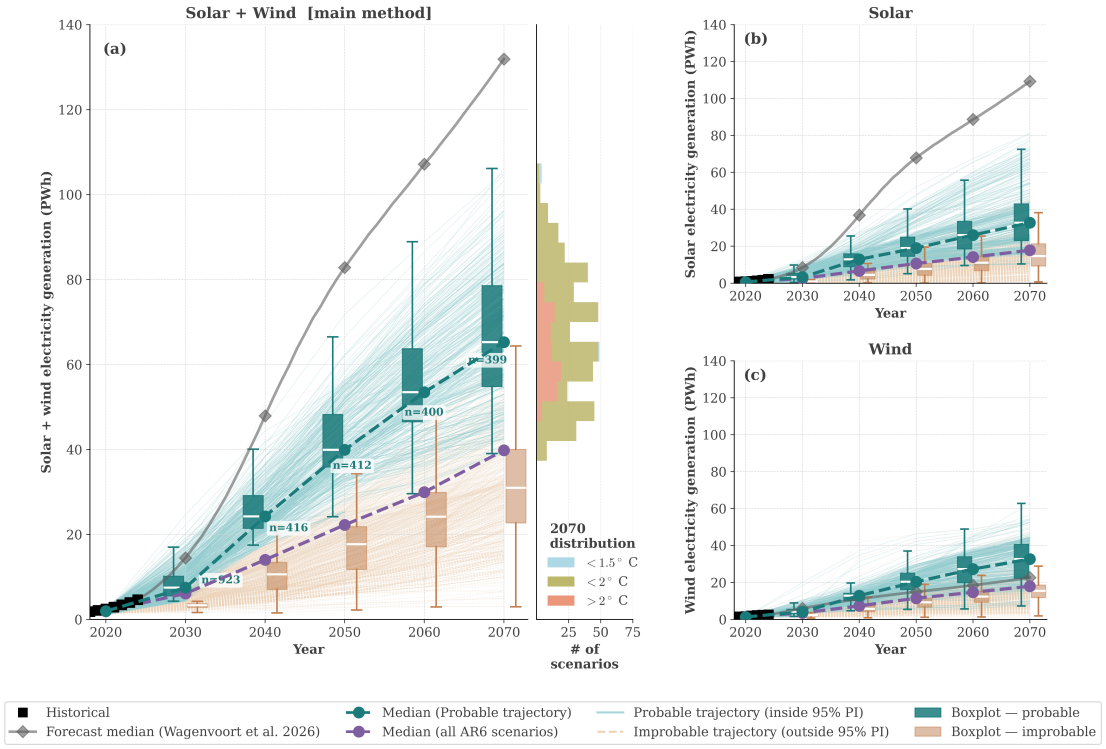


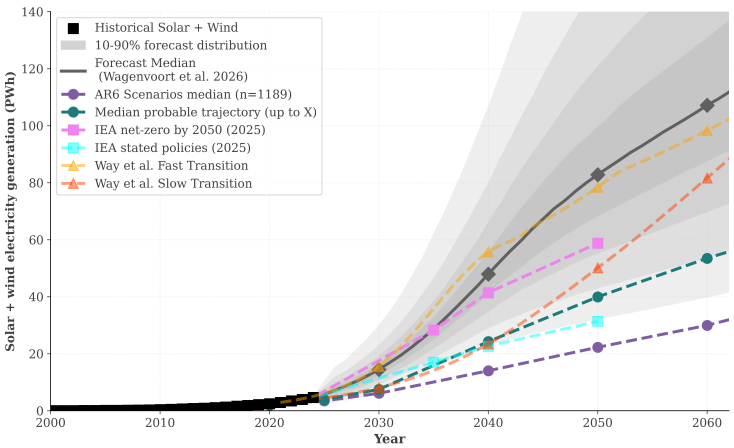
Figure 5.12: Comparison of IMP's solar + wind electricity generation levels for AR6 scenarios to the probabilistic forecast. Left panel shows solar + wind, middle panel solar, and right panel wind. Shaded regions show 80% prediction interval (10% steps). Red lines indicate high-warming scenarios; blue lines indicate illustrative mitigation pathways.

IMP-LD (low demand; blue line with square) scenario show high levels (near 75%) of solar and wind electricity shares (See main text Figure 5.1), but low levels of solar and wind electricity generation, due to low total electricity generation. Hence, in the main text the IMP-LD scenario falls well within the 95% prediction interval for solar and wind electricity shares, but in Figure 5.12, it falls below the 10% prediction interval for solar and wind electricity generation levels after 2050. This illustrates how using shares can mask differences in total electricity generation across scenarios, which may differ due to factors such as demand growth and energy efficiency improvements.

Results using electricity generation (PWh) Figure 5.13 compares solar and wind electricity generation levels in probable and improbable scenarios, complementing Figure 5.1 in the main text, which uses shares. The patterns align with the main text, with probable scenarios characterised by higher solar and wind electricity generation levels compared to improbable scenarios. Similar to the main text, many scenarios lie well below the median forecast. Here, scenarios lie well below the median, while in the main text, some scenarios lie above the median. The discrepancy may indicate that historical growth



(a) Probable vs. improbable AR6 scenarios



(b) Comparison to other scenarios

Figure 5.13: Comparison of solar PV, wind, and solar PV + wind electricity generation for probable (teal) and improbable (tan) AR6 scenarios. Historical data (black squares) shown for 2020–2024. Forecast median (grey diamonds), median of all AR6 scenarios (purple), median of probable scenarios (teal). **(a–c)** Progression of probable and improbable AR6 scenarios over time. Histogram indicates solar+wind electricity share in 2070 coloured by surface warming. **(a)** Solar PV + wind (main method), and the disaggregation into **(b)** Solar PV, **(c)** wind electricity shares. The number of probable scenarios decreases over time to 399 in 2070. **(d)** Median probable scenario trajectories (teal) compared to IEA scenarios (pink, cyan) and Way et al. transitions (yellow, red).

of electricity generation (PWh) is faster than what the AR6 scenarios project. As such, shares may be more consistent with the forecast than levels.

The median forecast for solar and wind electricity generation levels initially lies below the Way et al. [29] fast transition, but around 2040, the paths cross, unlike in the main text where the median forecast lies consistently below the Way et al. fast transition. The Way et al. fast transition assumes faster solar + wind growth from 2021-2024 than what was realized. Hence, there may be a cascading effect where the forecast is lower than the Way et al. fast transition in the near term. Furthermore, the Way et al. fast transition assume a lower long-term growth rate of total electricity demand, which may explain why the median forecast in levels lies above it from 2040 onwards, while in the main text, the median shares forecast lies below it from 2040 onwards.

Figures 5.14 and 5.15 compare the evolution of key energy transition metrics for probable and improbable scenarios using electricity generation levels, complementing main paper Figures 5.2 and 5.6, which use shares.

These findings are consistent with the main text, with probable scenarios characterised by higher total electricity generation, higher solar and wind electricity generation levels, and lower fossil fuel electricity generation levels. CCS deployment differs significantly between probable and improbable scenarios from 2070 onward, while in the main text, a significant difference emerges at 2060. In levels, probable scenarios do not have data/do not generate a lot of hydrogen from solar PV, while in shares they did. Surface warming is limited to below 2°C with approx. 80% probability in probable scenarios, compared to approx. 50–60% in improbable scenarios, consistent with the main text results using shares. Under levels, median net-zero year is marginally lower than for shares (main text).

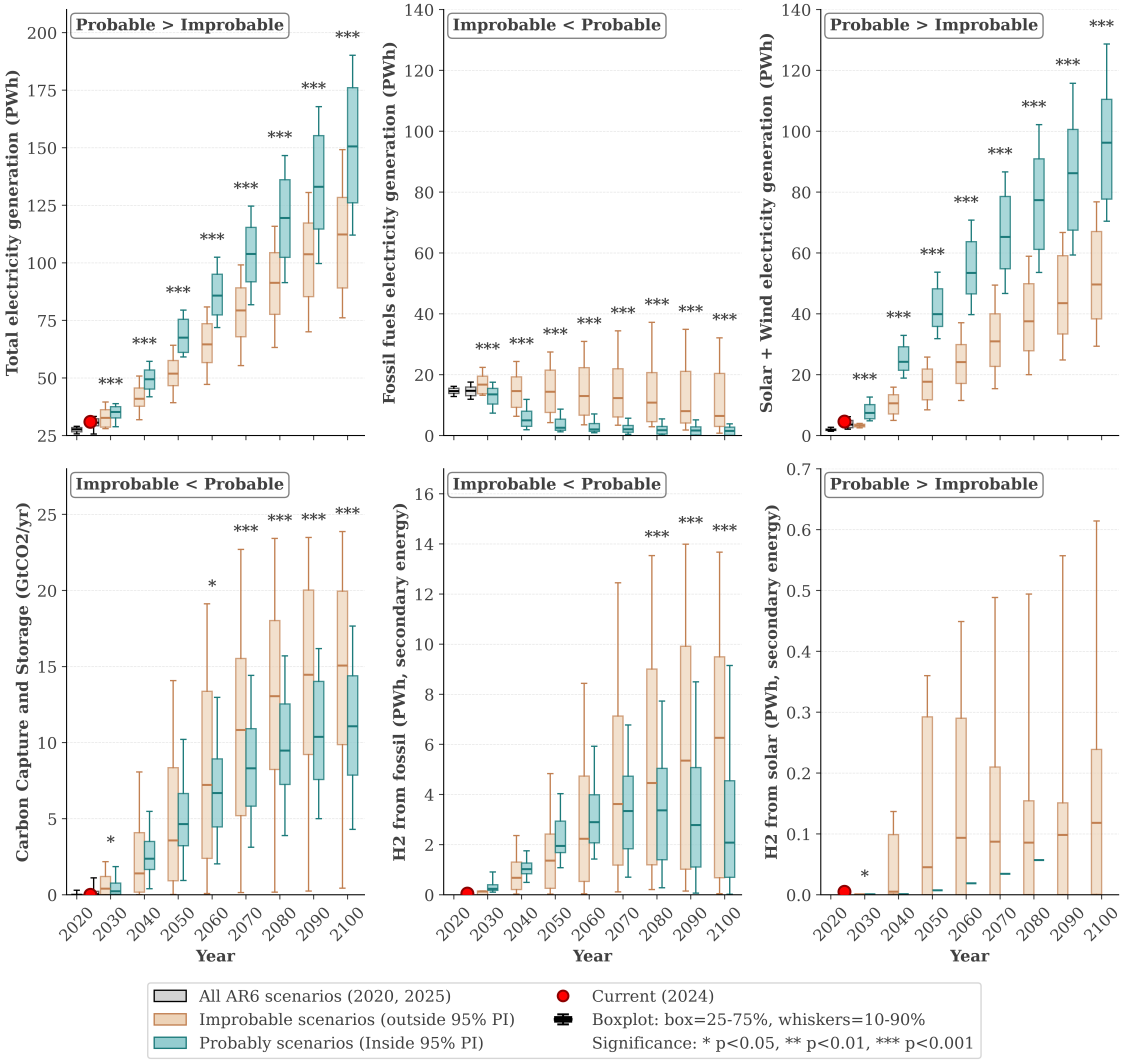


Figure 5.14: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan). Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. Grey box-plots represent AR6 scenarios in 2020 and 2025, with the red dot marking the 2024 observed levels. **(a)** Total electricity generation. **(b)** Fossil fuel electricity generation. **(c)** Solar and wind electricity generation share. **(e)** Yearly carbon sequestration from CCS. **(d)** Blue hydrogen (from fossil fuel incl. CCS) production (secondary energy). **(f)** Hydrogen production from solar (secondary energy).

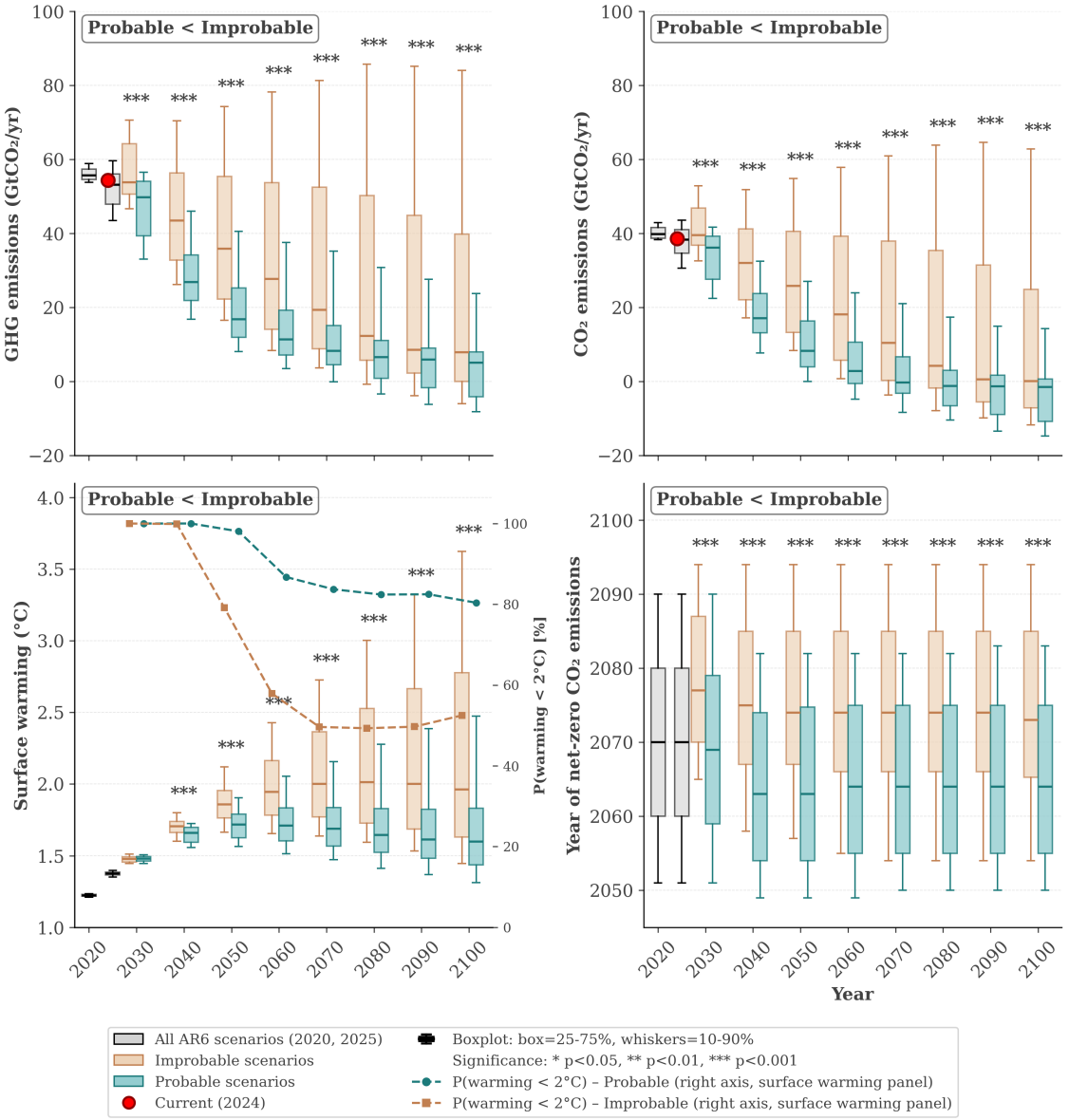


Figure 5.15: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan) using electricity generation levels. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** GHG emissions. **(b)** CO_2 emissions. **(c)** Surface warming. **(d)** Year of net-zero.

5.13.5 Results using useful energy shares

Electricity accounts for only a subset of total energy demand, and also excludes end-use energy losses. Useful energy can account for energy losses, while also providing a more comprehensive measure of energy demand across sectors. Appendix 5.12 presents the method for calculating useful energy shares. Here we show results using useful energy shares, complementing the main text results using electricity generation shares.

Figure 5.16 compares solar and wind useful energy shares in AR6 scenarios to the probabilistic forecast, complementing Figure 5.1 in the main text, which uses electricity generation shares. Results are broadly consistent with the main text but the number

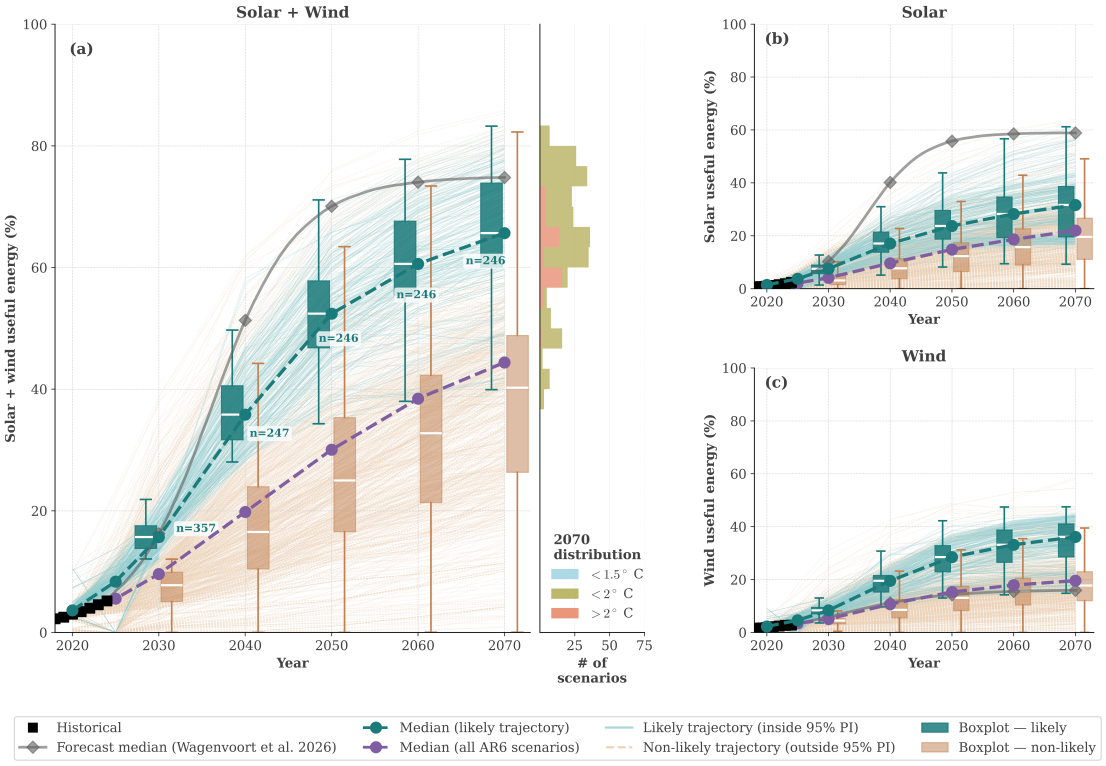


Figure 5.16: Comparison of solar + wind useful energy shares for AR6 scenarios to the probabilistic forecast. probable scenarios are shown in teal, improbable scenarios are shown in tan. **(a)** Solar + wind share of useful energy. **(b)** Solar share of useful energy. **(c)** Wind share of useful energy.

of probable scenarios is slightly smaller. For example, in 2050, there are 246 scenarios with trajectories inside the 95% prediction interval for solar + wind useful energy shares, compared to 339 for electricity shares.

Comaring to IMP scenarios using useful energy shares Figure 5.17 compares solar and wind useful energy shares of the IMP scenarios, complementing Figure 5.4 in the main text, which uses electricity generation shares.

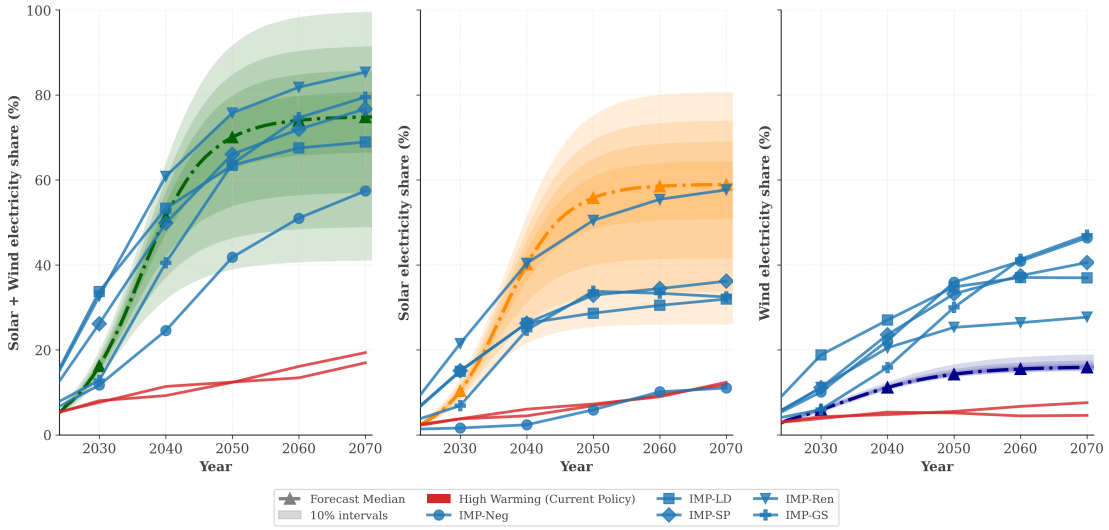


Figure 5.17: Comparison of solar + wind useful energy shares for AR6 scenarios to the probabilistic forecast, using useful energy shares. IMP scenarios are shown as lines, coloured by SSP. Shaded regions show 80% prediction interval (10% steps). Red lines indicate high-warming scenarios; blue lines indicate illustrative mitigation pathways.

Energy evolution in probable vs. improbable scenarios using useful energy shares Figure 5.18 compares the evolution of key energy transition metrics for probable and improbable scenarios using useful energy shares, complementing main paper Figure 5.2. The patterns echo the main text, with probable scenarios characterised by higher electrification, higher solar and wind useful energy shares, and lower fossil fuel shares. Fossil fuels retain a small share of useful energy, likely because some end-uses are difficult to electrify (e.g. hard-to-abate sectors such as heavy industry and transport) or are coupled with CCS. CCS deployment does not differ significantly between probable and improbable scenarios at any time point, unlike the main text result where a significant difference emerges at 2060.

Supplementary comparison using useful energy shares Figure 5.19 complements Figure 5.6 by comparing additional metrics for probable and improbable scenarios

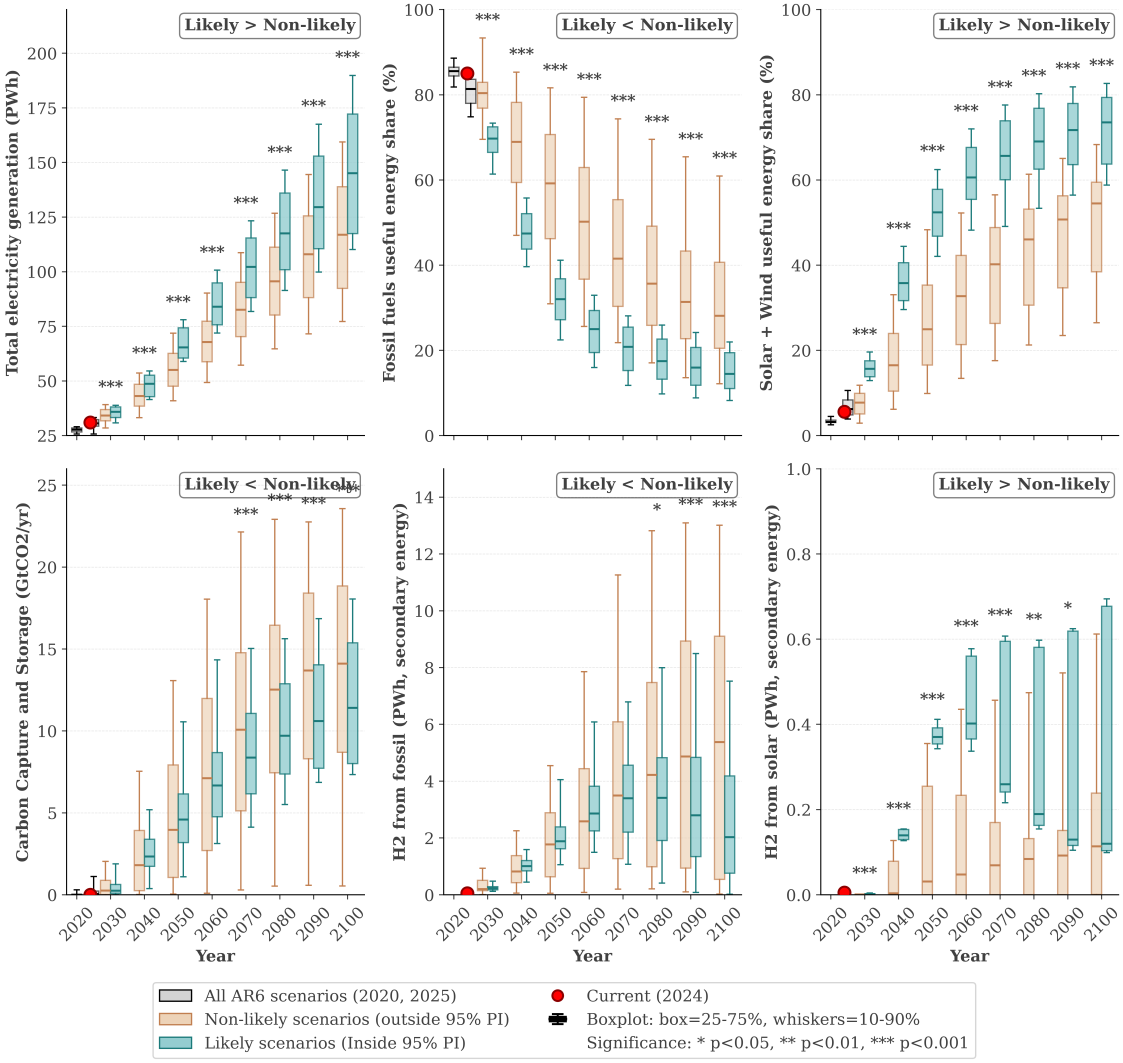


Figure 5.18: Energy transitions for probable and improbable AR6 scenarios using useful energy shares. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** Electricity share. **(b)** Solar PV share of electricity. **(c)** Wind share of electricity. **(d)** Fossil fuel share of electricity. **(e)** Carbon Capture and Storage ($GtCO_2$/yr). **(f)** Hydrogen share of final energy from solar PV. **(g)** Hydrogen share of final energy from fossil fuels.

using useful energy shares. Notably, the probability of limiting warming to below 2°C is much higher in probable than in improbable scenarios, consistent with the main result. Probable scenarios under the useful energy share measure show near 90% probability of limiting warming to below 2°C, while improbable scenarios show near 60%. Both values exceed those in the main text using electricity generation shares.

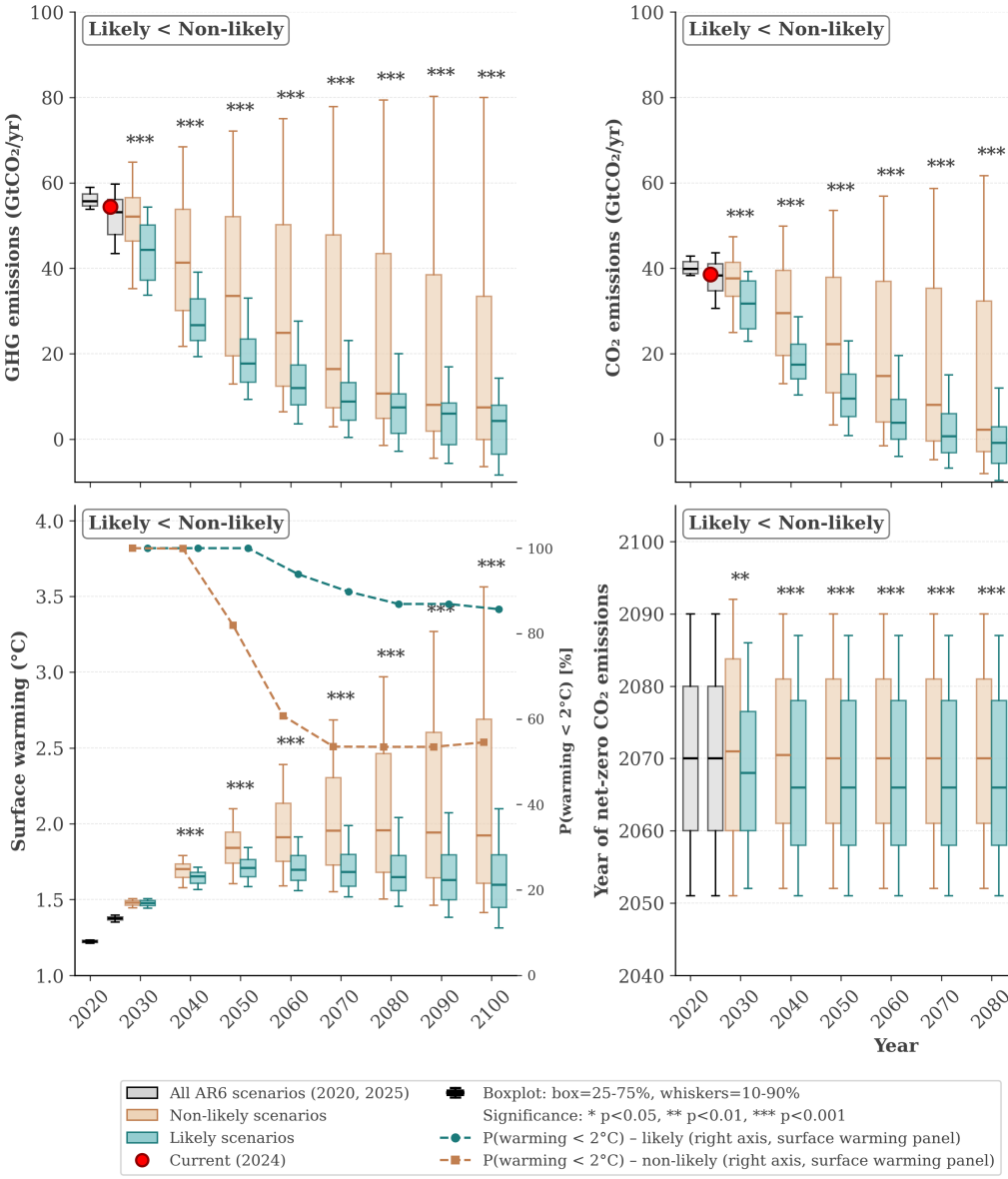


Figure 5.19: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan) using useful energy shares. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** CO_2 emissions. **(b)** GHG emissions. **(c)** Peak warming. **(d)** GDP. **(e)** Population. **(f)** Year of net-zero.

5.13.6 Useful energy in levels.

As a robustness check, we also compare scenarios using useful energy levels. As for electricity generation levels, here we use an exogenously determined 2% p.a. growth rate as a minimum, to account for growing energy demand after technology-driven growth has stalled (see Section 5.13.4 for details).

Comparing probable vs. improbable scenarios using useful energy levels

Figure 5.20 shows that probable scenarios have higher solar and wind useful energy than improbable scenarios, consistent with the main text results.

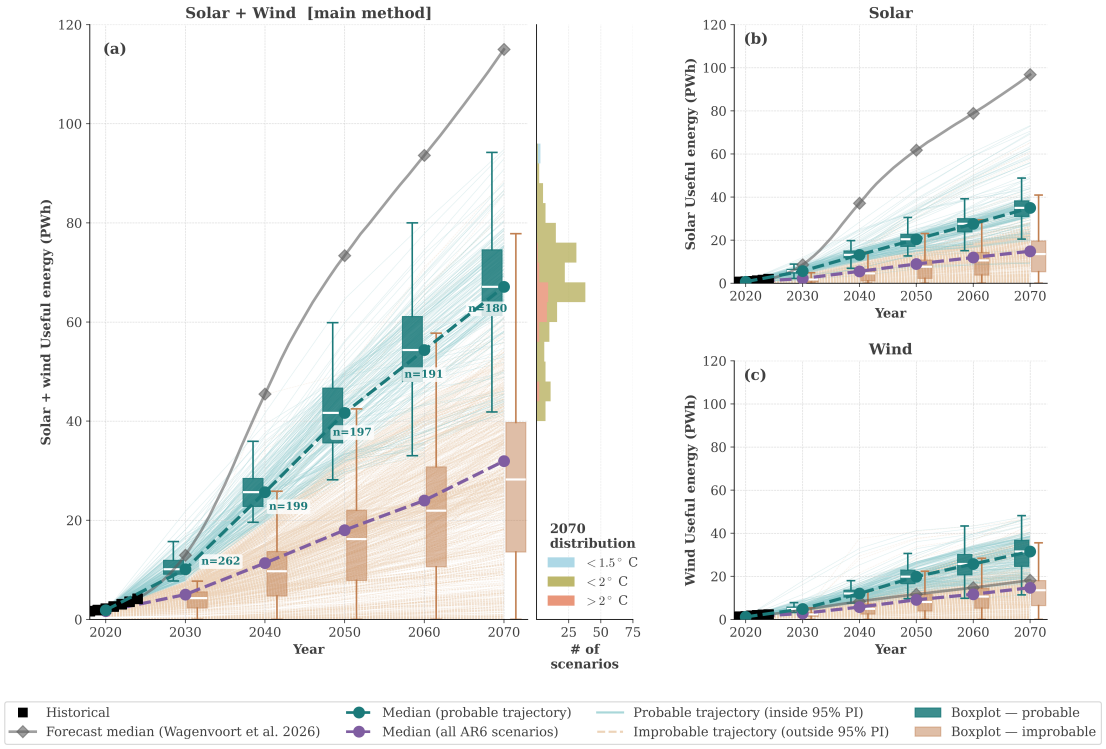
Comaring to IMP scenarios using useful energy generation Figure 5.21 compares solar and wind useful energy generation of the IMP scenarios, complementing Figure 5.4 in the main text, which uses electricity generation shares.

Figure 5.22 compares the evolution of key energy transition metrics for probable and improbable scenarios using useful energy levels, complementing the main paper Figure 5.2. Results are broadly consistent with the main text. Probable scenarios project substantially higher electrification than improbable scenarios, and the difference is more pronounced than in the main text result using electricity shares, indicating that probable scenarios feature not only higher shares but also higher absolute levels of solar and wind useful energy.

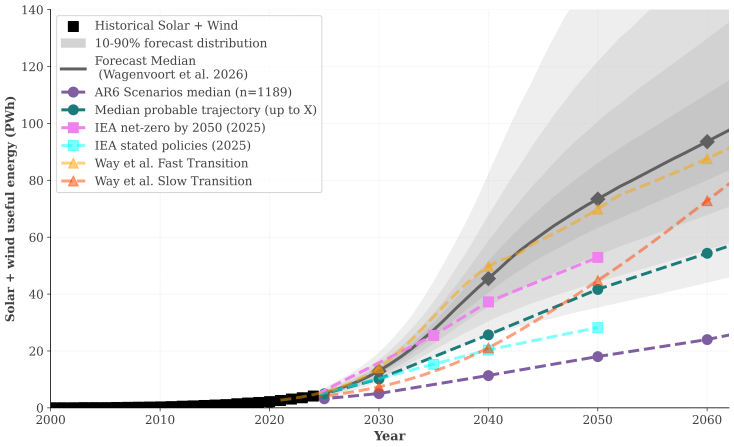
In contrast to the main text, probable scenarios deploy marginally more CCS than improbable scenarios in the median. Hydrogen from fossil fuels is also higher in probable scenarios, while data on hydrogen from solar PV are absent in this subset.

Figure 5.23 shows additional metrics, complementing Figure 5.6 in the main text. Scenarios consistent with solar + wind useful energy levels through 2070 reach a median net-zero just after 2060, slightly lower than probable scenarios as measured by electricity levels in Appendix 5.13.4, useful energy shares in Appendix 5.13.5, and electricity shares in the main text. The probability of limiting warming to below 2°C is slightly larger than in the main text results, using electricity shares.

Small differences between electricity shares and levels, between electricity and useful energy shares, and between useful energy shares and levels, suggest that the main text results are robust to different measures of energy transition pathways. Yet the differences



(a) Probable vs. improbable AR6 scenarios



(b) Comparison to other scenarios

Figure 5.20: Comparison of solar PV, wind, and solar PV + wind useful energy for probable (teal) and improbable (tan) AR6 scenarios. Historical data (black squares) shown for 2020–2024. Forecast median (grey diamonds), median of all AR6 scenarios (purple), median of probable scenarios (teal). **(a–c)** Progression of probable and improbable AR6 scenarios over time. Histogram indicates solar+wind useful energy share in 2070 coloured by surface warming. **(a)** Solar PV + wind (main method), and the disaggregation into **(b)** Solar PV, **(c)** wind useful energy shares. The number of probable scenarios decreases over time to 180 in 2070. **(d)** Median probable scenario trajectories (teal) compared to IEA scenarios (pink, cyan) and Way et al. transitions (yellow, red).

that do exist highlight the importance of considering multiple measures of energy transition pathways, as they can provide complementary insights and reveal different aspects of the

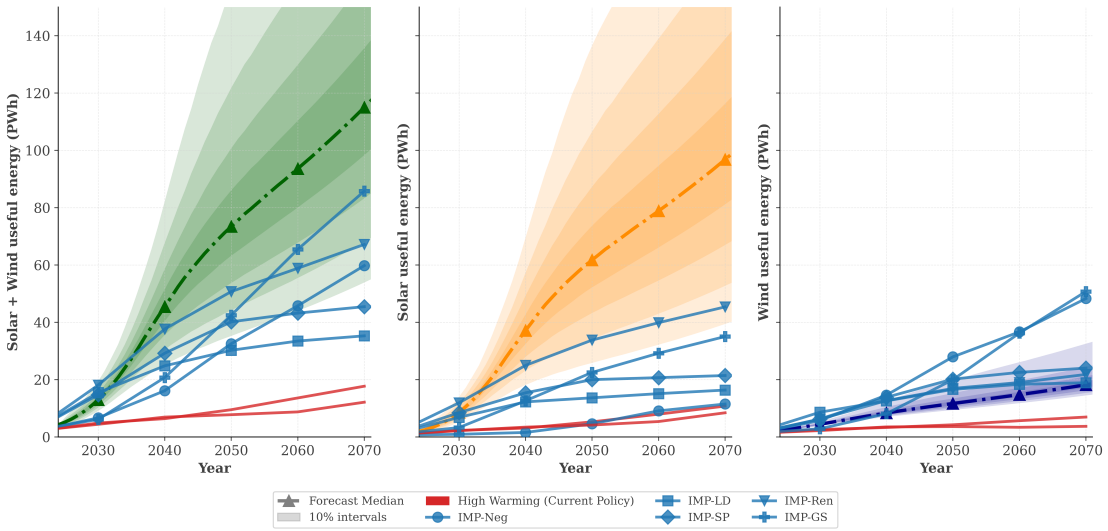


Figure 5.21: Comparison of solar + wind useful energy generation for AR6 scenarios to the probabilistic forecast, using useful energy generation. IMP scenarios are shown as lines, coloured by SSP. Shaded regions show 80% prediction interval (10% steps). Red lines indicate high-warming scenarios; blue lines indicate illustrative mitigation pathways.

transition. For example, shares can mask differences in total energy demand and generation levels across scenarios, while levels can provide a more comprehensive picture of the scale of the transition. On the other hand, levels can be influenced by assumptions about total energy demand growth, which may differ across scenarios. Useful energy can account for energy losses and provide a more comprehensive measure of energy demand across sectors, but it also requires additional assumptions and data to calculate.

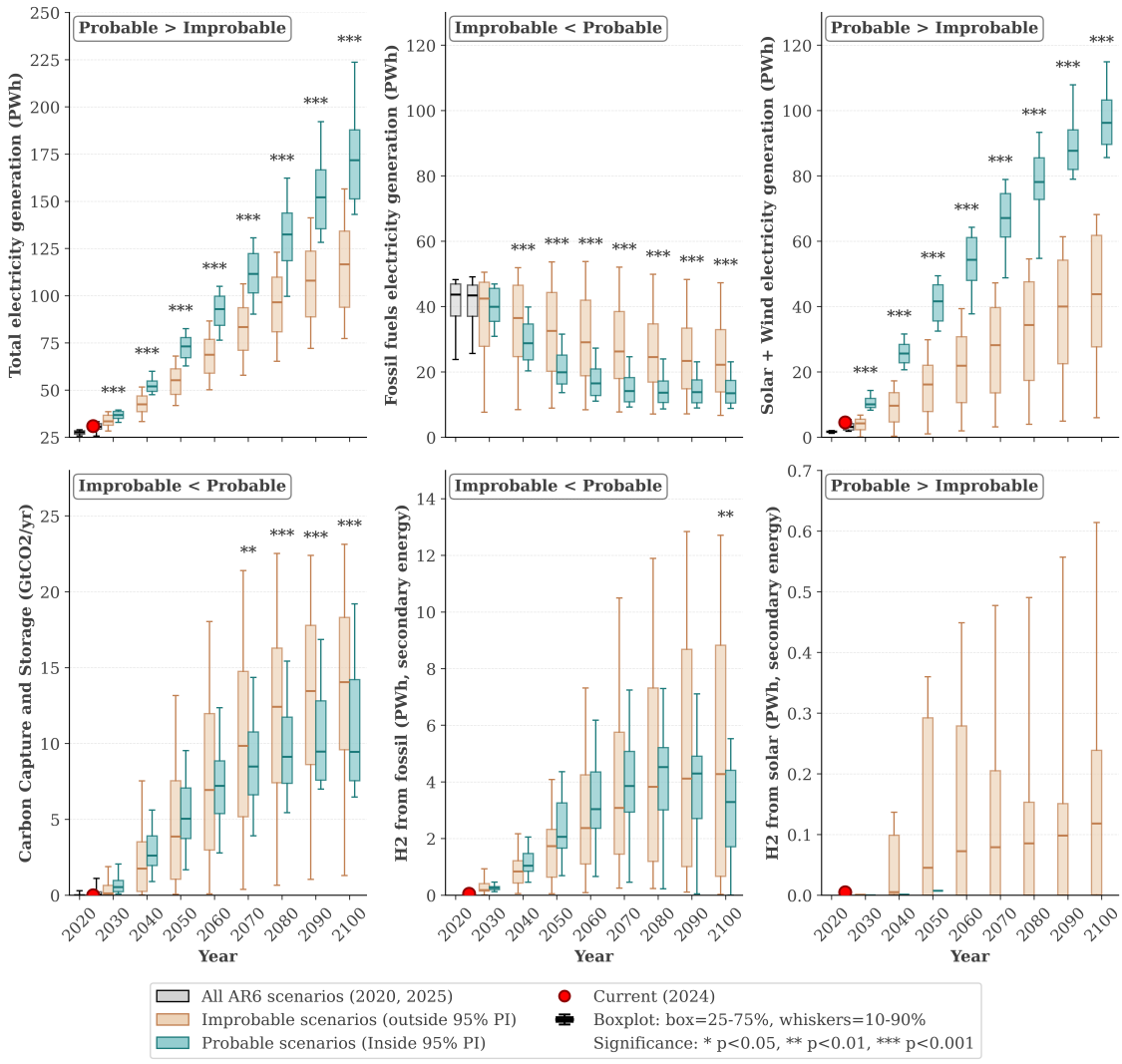


Figure 5.22: Energy transitions for probable and improbable AR6 scenarios using useful energy levels. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** Electricity share. **(b)** Solar PV share of electricity. **(c)** Wind share of electricity. **(d)** Fossil fuel share of electricity. **(e)** Carbon Capture and Storage ($GtCO_2$/yr). **(f)** Hydrogen share of final energy from solar PV. **(g)** Hydrogen share of final energy from fossil fuels.

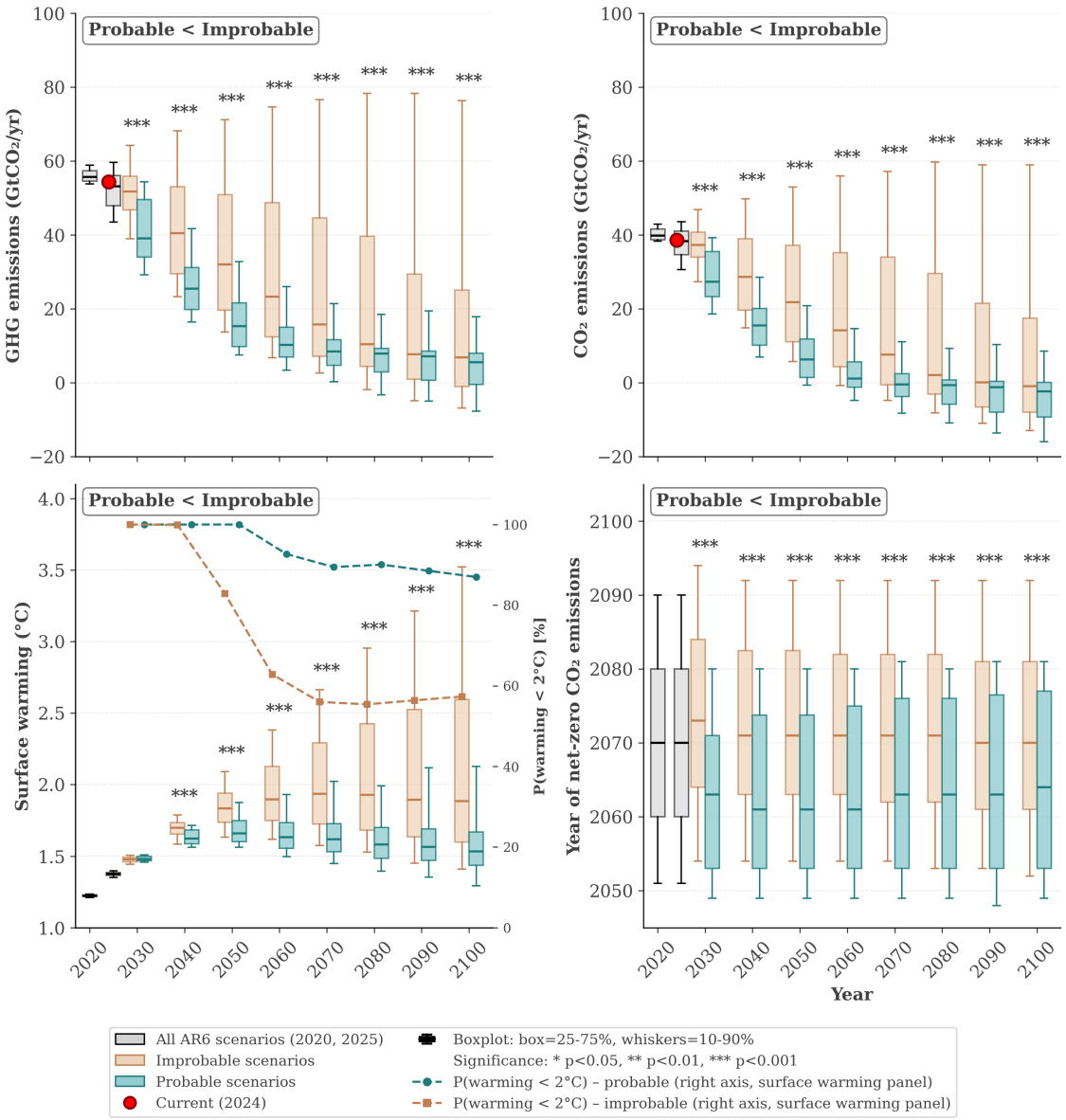


Figure 5.23: Boxplots comparing probable (teal) to improbable AR6 scenarios (tan) using useful energy levels. Boxplots show median, 25th and 75th percentiles, and 10th and 90th percentiles. Stars indicate significant differences based on Mann-Whitney U test. **(a)** GHG emissions. **(b)** CO_2 emissions. **(c)** Surface warming. **(d)** Year of net-zero.

Chapter 6

Discussion

Chapters 3–5 established that S-curve fitting with partial data produces systematic underestimation, that empirical regularities across 120 technologies enable well-calibrated Bayesian forecasts, and that these forecasts favour rapid decarbonisation pathways over the IAM consensus. This chapter interprets these findings, discusses their policy implications, identifies limitations, and outlines directions for future research.

6.1 Interpretation of Results

6.1.1 Are S-curves adequately usable?

Chapter 3 offers a statistical explanation for why renewable deployment forecasts may have been systematically underestimated [28], and why they vary so much [31]. Even when fitting the correct functional form, or even the correct noise model, the S-curve fitting process itself produces downward bias when only early diffusion data are observed. Consequently, fitting S-curves to early data, for example as done in energy forecasting and projections, can systematically mislead and underestimate future adoption.

The underestimation bias identified in Chapter 3 is a statistical property of S-curve fitting with partial data that applies to any diffusion process. This has a subtle but important consequence: if practitioners across different domains have historically fitted

S-curves to early data and anchored their expectations to the resulting (downward-biased) projections, this creates a systematic tendency to underestimate the speed and scale of diffusion processes more generally.

The implications cut both ways. For beneficial processes, such as renewable energy deployment, vaccine adoption, or electric vehicle uptake, underestimation leads to underinvestment in supporting infrastructure and missed opportunities for early action. If forecasts systematically underestimate renewable deployment, policymakers may underinvest in grid infrastructure, such as additional transmission and storage to accommodate more variable renewables, and overinvest in fossil fuel capacity and infrastructure that becomes stranded (Section 6.2 discusses the policy consequences in detail). For harmful processes, such as epidemic spread, invasive species, or misinformation, the same bias produces dangerous complacency. For example, public health responses calibrated to underestimates of viral spread lead to insufficient healthcare capacity and delayed interventions. Similarly, underestimating the adoption rate of AI could lead to insufficient investment in education and workforce retraining. Addressing underestimation bias is therefore not merely an academic exercise but has tangible consequences for how societies plan and invest. The debiasing method proposed in Chapter 3 offers a general diagnostic for any field that relies on S-curve extrapolation from early-stage data, while the Bayesian framework developed in Chapter 4 provides a more robust alternative that explicitly quantifies uncertainty rather than relying on potentially misleading point estimates.

From structural bias to calibrated inference. The existence of structural underestimation bias does not render S-curves unusable; rather, it changes how they must be interpreted. Deterministic early-stage S-curve fits should not be treated as neutral forecasts but considered with appropriate caution in light of downward bias and high variance when using simple NLLS fitting procedures. Chapter 3 shows that the problem is

not incorrect model choice but incomplete information: asymptotic parameters are weakly identified during exponential growth, creating shrinkage toward conservative estimates. This reframes S-curve practice from curve-fitting toward uncertainty quantification. The appropriate response is not abandonment, but calibration.

Chapter 4 demonstrates that by explicitly modelling parameter uncertainty within a Bayesian framework, S-curves can transition from fragile point forecasts to probabilistically robust inference tools. Empirically grounded regularities across technologies, combined with out-of-sample validation and calibration diagnostics, enable clear quantification of forecast quality that reflects both structural uncertainty and historical diffusion patterns. Wide credible intervals during early exponential growth are not forecasting failures but honest representations of limited information. As new data accumulate, posterior distributions update coherently, making the framework adaptive and self-correcting rather than anchored to a single trajectory.

6.1.2 What probabilistic forecasts reveal

Chapter 4 analysed empirical regularities across 120 technologies and showed that technologies across sectors and time periods share common diffusion properties. Technology diffusion is systematically asymmetric, with adoption initially growing exponentially from early uptake. The practical consequence of the findings is that the exponential growth phase, where deployment increases fastest and uncertainty is greatest, is also the phase where forecasts matter most for planning. Technologies observed only during this phase are simultaneously the hardest to forecast and the most consequential to get right.

Solar and wind sit precisely in this critical window. Chapter 4 showed that both remain in early exponential growth, with median projections exceeding IEA Net Zero 2050 targets. The wide credible intervals (median ∼80 PWh for solar by 2050, 90% interval 13–470 PWh) do not indicate poor forecasting but rather honestly quantify the range of

outcomes consistent with historical diffusion patterns at this stage. Deterministic point forecasts, whether from IAMs or expert panels, project comparable uncertainty but conceal it behind a single number. The Bayesian framework makes this uncertainty explicit, enabling decision-makers to plan across the distribution rather than anchoring to a single trajectory. Crucially, probabilistic forecasts with stated credible intervals are inherently falsifiable: as new data accumulate, one can directly test whether calibration holds, making the methodology self-correcting in a way that conditional what-if scenarios are not.

Interpreting the apparent acceleration of diffusion over time. Chapter 4 finds that fitted growth rates k rise approximately exponentially with the year of first observation, suggesting that more recent technologies diffuse faster. This finding admits at least two non-mutually-exclusive interpretations. The first is mechanistic: globalisation, denser communication networks, and faster information flow plausibly accelerate social contagion [439], so later technologies genuinely diffuse faster because the underlying network supports faster spread. The second is compositional: Wilson et al. [211] argue that more granular technologies (small unit size, modular, mass-manufacturable) diffuse faster than bulky technologies that require large lump-sum investment. Historical samples, including ours, are dominated by bulkier infrastructure-heavy technologies, while recent samples include more granular consumer products and modular renewables. If granular technologies systematically diffuse faster and their share of the dataset rises with time, the time trend in k partly reflects this compositional shift rather than a mechanism in which calendar time itself accelerates diffusion. Distinguishing these two explanations would require a dataset stratified by granularity and time, which is an open empirical task. The practical implication is that extrapolating the time trend in k forward should be done with care: it may be partly driven by which technologies entered the historical record rather than by a universal trend in adoption speed.

Empirical grounding versus expert judgment. IAM renewable deployment assumptions often reflect expert judgment about resource constraints and integration challenges rather than extrapolation from empirical learning curves and diffusion patterns. As documented in Section 2.2, experts systematically underestimate uncertainty and exhibit anchoring bias [25], leading to conservative growth assumptions that underweight exponential dynamics. Chapter 4 addressed this by grounding forecast methods in empirical regularities and implementing quasi-empirical priors to inform the asymptote rather than eliciting expert beliefs. Prior specification still requires choices and thus does not eliminate judgment, but it makes those choices transparent and empirically grounded rather than implicit in expert-driven scenario assumptions.

6.1.3 Probabilistic forecasts as complement to IAM scenarios

The literature review (Section 2.3) established that IAM scenarios present deterministic pathways without probabilistic uncertainty quantification, which systematically underestimate renewable deployment, and lack empirical validation against historical diffusion patterns. This thesis develops Bayesian S-curve forecasting as a complementary tool addressing these limitations while acknowledging fundamental differences in purpose and scope.

From what-if to how-likely. IAM scenarios explore conditional pathways under different assumptions [38], answering questions like *what happens if we impose carbon pricing of $100/ton?* They deliberately vary technological, economic, and policy assumptions to explore possible futures [233]. Probabilistic S-curve forecasts instead quantify likelihood based on empirical diffusion regularities, answering *how probable is continued exponential growth?* or *what is the 90% credible interval for 2050 solar capacity given historical patterns?* One is a what-if exploration tool; the other is a data-driven likelihood assessment.

Combining them allows one to evaluate which what-if scenarios are also "likely" given historical evidence, and which require unprecedented departures from historical patterns. It is worth emphasising that the Bayesian forecasts of Chapter 4 are themselves *conditional*: they are conditional on the historical data, the chosen functional form, and the prior. In the IPCC sense (see introduction terminology), they are therefore closer to probabilistic projections than to unconditional predictions, with the important difference that the conditioning is on empirical regularities rather than on stipulated policy or socio-economic assumptions. This positions them as a probabilistically calibrated complement to IAM scenarios rather than as a categorically different object.

The probabilistic assessment allows one to evaluate which IAM assumptions align with empirical evidence (likely) versus which embody speculative departures (non-likely). For instance, when IAM scenarios impose exogenous constraints such as floor costs, limiting solar growth rates despite empirical learning curves showing consistent cost reductions, probabilistic forecasts quantify how this assumption departs from historical regularities. The framework does not replace scenarios but disciplines them. Transparently distinguishing scenarios that *extend* historical trends from those that *contradict* can help reduce systematic bias in how scenario ensembles are used for policy guidance.

Complementarity with IAM scenarios. The approaches serve different roles. IAM scenarios inform which futures are *technically feasible* given policy interventions and behavioural change; probabilistic forecasts assess which are *historically plausible* given observed diffusion patterns. IAM scenarios can model policy discontinuities, behavioural transformations like the IMP-LD 40% energy demand reduction [68], structural breaks, and technology interactions that probabilistic S-curve forecasts cannot capture (Section 6.3 discusses these methodological boundaries in detail). Conversely, IAMs appear to lack sufficient empirical grounding and uncertainty quantification that probabilistic forecasts

provide. Together the two approaches identify futures that are both technically feasible *and* empirically supported—a narrower, more policy-relevant set than either produces alone (Section 6.2 discusses how this filtered subset enables more targeted policy analysis). More broadly, this complementarity extends to socio-technical transitions theory: the multi-level perspective [199] explains *how* and *why* transitions unfold through niche-regime-landscape interactions, while S-curve forecasting quantifies *how fast* and *how far* the aggregate outcome proceeds. Models such as the FTT framework [259, 260] already demonstrate that logistic substitution dynamics can operate within system-level scenario models, though the trade-offs involved are discussed in Section 6.3.

6.1.4 Ensemble composition as anchoring bias

Chapter 5 revealed that scenarios most consistent with empirically grounded solar and wind forecasts sit at the optimistic tail of the AR6 ensemble. Although the IPCC explicitly states that AR6 scenarios carry no attached likelihoods, in practice policymakers and investors likely anchor their expectations to the ensemble centre and treat outcomes near the tails as implausible. Rapid decarbonisation pathways, which this thesis identifies as the most empirically supported, are therefore likely to be dismissed as unrealistically ambitious, even when historical diffusion patterns suggest otherwise.

This finding invites a meta-interpretation: the perceived implausibility of rapid decarbonisation is partly an artefact of ensemble composition rather than evidence against such pathways. If scenario modellers investigated more high-growth pathways, the ensemble centre would shift. Scenarios previously labelled as optimistic outliers appear central, despite no change to other individual scenarios. The scenarios do not become more or less plausible; only their position within the ensemble changes. This is anchoring bias operating not through individual expert judgment but through the structural composition of the scenario ensemble that decision-makers treat as the range of reasonable futures.

Section 6.2 discusses the policy consequences of this effect. The same reflexivity applies to probabilistic forecasts: if decision-makers anchor to our finding that rapid deployment is empirically likely, this could become self-reinforcing, just as pessimistic IAM scenarios may have discouraged ambition. Forecasts shape the expectations that produce outcomes, and this performativity should be acknowledged even when correcting a systematically pessimistic anchor is beneficial. The IPCC is explicit that the AR6 scenario set is an ensemble of opportunity rather than a probability distribution, and cautions against treating its spread as such [234]. In practice, however, ensembles assembled to span the literature can still be treated as approximate distributions by downstream users; if the underlying literature is itself systematically skewed, those approximate distributions inherit the skew. Acknowledging this in scenario communication, as the IPCC does, is necessary but is unlikely to be sufficient on its own to neutralise the anchoring effect.

6.2 Policy Implications

The results in this thesis have direct consequences for energy policy, infrastructure planning, and how scenario ensembles inform decision-making.

6.2.1 Policy Agnosticism Does Not Mean Policy Irrelevance

The methods employed in this thesis are policy-agnostic. The forecasts are not conditional on specific policy assumptions but reflect the distribution of outcomes consistent with historical diffusion patterns. This raises an important question: if rapid renewable deployment is empirically plausible without specifying which policies enable it, does policy still matter?

It does. As the Lucas critique suggests, policies that affect the parameters of the diffusion model also affect the forecasts. Historical diffusion patterns already embed the

effects of past policies, such as subsidies, mandates, R&D investment, so the forecast distribution reflects outcomes under the historical range of policy environments. Where future policies fall within that range determines where the realised trajectory sits within the forecast distribution. Policies supporting rapid growth push the realised trajectory toward the upper tail, while insufficient policies may push it toward the lower tail. The climate consequences differ substantially: results from Chapters 4 and 5 suggest that the AR6 scenarios in the lower tail of the forecast distribution could still produce 2.5°C warming, while the upper tail limits warming well below 2°C.

Specific policy levers may map onto model parameters. Renewable energy subsidies and deployment mandates affect near-term growth rates (k_i in the model of Chapter 4). Governmental support for grid expansion may increase the renewable capacity the grid can accommodate, affecting the asymptote parameter (L_i). Baumgartner and Farmer [410] demonstrate that learning rates vary by country, particularly for wind, suggesting that country-specific effects and policies can accelerate deployment and cost reductions. Tankwa et al.[450] show that policies have had a significant effect on solar PV growth rates.

When Chapter 5 finds that scenarios consistent with empirically grounded solar and wind forecasts indicate a >80% probability of limiting warming to 2°C, this should not be interpreted as *no policy action needed*. Rather, it means *if policies sustain historical deployment trends, rapid transitions are empirically probable*. Empirical plausibility establishes feasibility but does not guarantee realisation without appropriate institutional and political conditions. Institutional barriers, such as fossil fuel industry resistance, regulatory inertia, or financing constraints, may prevent realisation even when rapid deployment is empirically probable. Probabilistic forecasts quantify *technical* probability given historical diffusion patterns but remain silent on the *political* feasibility of sustaining those patterns.

6.2.2 Risk-Based Infrastructure Planning

Imagine a scenario-based planning approach that designs infrastructure for discrete pathways: *build for 30% renewables by 2030 under Scenario A*. Probabilistic forecasts enable risk-based design using credible intervals. For example, one may plan transmission networks for 25–45% renewables to be robust across plausible outcomes. Chapter 4 demonstrated that solar forecasts provide a median of ∼80 PWh useful energy generation by 2050 (90% interval 13–470 PWh), allowing grid planners to assess which investments yield benefits across the full predictive distribution rather than betting on a single scenario. If the 10th percentile forecast exceeds current expansion plans, planners face measurable underinvestment risk. Chapter 5 showed that scenarios consistent with empirically grounded forecasts imply a significant decline in fossil fuel share by 2050, informing stranded asset risk assessment for pension funds and fossil-dependent economies.

6.2.3 Cost Implications and the Economics of Transition

Chapter 4 found that solar and wind growth rates consistent with historical diffusion patterns would exceed Net Zero 2050 targets, while Chapter 5 showed these rapid deployment pathways are consistent with limiting warming below 2°C. This directly suggests that the observed learning-curve-driven cost reductions make rapid transition *likely* (in the sense of empirically consistent with solar PV and wind forecasts). Further, combining our results with those from Way et al. [29] challenges the IAM consensus that mitigation is necessarily costly, instead suggesting that rapid transitions are economically favourable; a conclusion grounded in empirical cost data rather than expert priors about future technology limits.

6.2.4 The Ensemble Anchoring Problem

The ensemble anchoring bias identified in Section 6.1 has direct policy consequences. Infrastructure investment, national emissions pledges, and corporate strategy are all shaped by which futures the scenario ensemble treats as mainstream.

Chapter 5 already provides a practical tool for addressing this: by filtering the AR6 ensemble against empirical forecast distributions, it narrows hundreds of scenarios to an empirically grounded subset. This subset enables more targeted policy analysis than the full ensemble permits; specific investment requirements, infrastructure expansion timelines, and stranded asset exposure can be assessed across a range of futures that historical diffusion patterns actually support, rather than across the entire (structurally biased) ensemble.

At the ensemble level, scenario modellers and assessment bodies should actively solicit scenarios built on empirically grounded diffusion assumptions. Way et al.[29] demonstrate that including rapid economic growth combined with rapid renewable deployment produces scenarios where ambitious decarbonisation is both economically and empirically plausible. Including more such pathways would recentre the ensemble, making rapid-deployment scenarios appear central rather than extreme. Policymakers and investors need to be informed that the current ensemble centre reflects which scenarios modellers happened to submit, not a probability-weighted assessment of likely futures. This matters because countries that lead in renewable deployment also tend to lead more broadly in technology adoption and benefit most economically from transitions [204]; the current ensemble structure may actively discourage the early-mover strategies that historical evidence suggests would be most rewarding.

6.3 Limitations

This thesis makes several methodological and empirical choices that constrain scope and introduce limitations. Understanding these boundaries is essential for interpreting results and identifying avenues for future research.

6.3.1 Data Limitations

Selection bias toward successful technologies. The multi-technology dataset used in Chapter 4 consists predominantly of technologies that successfully diffused to maturity, creating survivorship bias: failed technologies that follow different diffusion patterns are underrepresented. Nonetheless, some technologies that start diffusion but fail to become as successfully adopted as one may expect (such as Nuclear power), are still included in the dataset. Many candidate technologies fail before reaching widespread adoption, and these failures plausibly contain information about the conditions, network structures, and economic regimes under which diffusion stalls. Studying failed technologies, alongside successful ones, would inform the boundary between conditions that produce S-curve growth and those that do not, and may sharpen forecasts for nascent technologies that have not yet declared themselves successful or failed.

Limited data in tails. Many technologies in the dataset have sparse observations during the critical early exponential growth phase, or later diffusion stages. This means that the empirical regularities identified, such as the superior fit of an asymmetric s-curve are primarily informed by data from the middle diffusion stages, where most technologies have more observations. The early and late stages, which are crucial for forecasting, defining patterns, and understanding saturation mechanisms, are less well represented. Better data on the very earliest stages of diffusion, before a technology declares itself a clear winner, remains an open empirical priority: it is precisely in this regime that forecasts are most

consequential and least well constrained.

Data quality and consistency. Technology adoption time-series exhibit inconsistencies across data sources, measurement methodologies, and geographic aggregation levels. A single technology may be represented in absolute terms or relative shares. For example, solar PV deployment can be measured in installed capacity (GW), generation (TWh), or share of electricity. These measurement choices influence the shape and estimated parameters of fitted S-curves, complicating cross-technology comparisons and model specification. Mixing data from different sources with varying quality and definitions may introduce additional noise or systematic bias into the analysis, raising the question of how much of the observed regularities reflect underlying diffusion dynamics rather than artefacts of measurement heterogeneity.

Due to data scarcity, this thesis does not attempt to harmonize all datasets exhaustively. However, extensive backtesting, sensitivity analyses, and robustness checks indicate that the core findings are not sensitive to reasonable variations in prior, training-test split, or small changes in model parameters. The presence of consistent cross-technology regularities despite heterogeneous measurement conventions suggests that the results are unlikely to be driven solely by measurement artefacts. In particular, strong out-of-sample performance indicates that the models capture stable predictive structure rather than overfitting idiosyncratic measurement noise.

Analyzing only AR6 scenarios. In Chapter 5, the thesis evaluates the plausibility of IPCC AR6 scenarios against historical diffusion patterns. This focus on a single scenario ensemble limits generalizability: other scenario sets (e.g., IEA, Shell, Greenpeace) may have different assumptions and distributions of technology trajectories. While we briefly compare to IEA scenarios, a more comprehensive multi-ensemble analysis would strengthen conclusions about the general plausibility of rapid decarbonisation pathways. However, the

AR6 ensemble is the most widely used for policy guidance, making it a natural starting point for this analysis.

6.3.2 Methodological Limitations

S-curve fitting bias Chapter 3 identifies a systematic downward bias in S-curve fitting when only early diffusion data are observed. This bias is a statistical property of the fitting process rather than a reflection of any particular technology or domain. The debiasing method offers a remedy, but relies on critical assumptions about the noise model and functional form. If these assumptions are violated, the correction may be inaccurate. In high noise low data environments, the bias may be more severe and harder to correct, making early-stage forecasts particularly challenging for the debiasing method.

Furthermore, this begs the question, to what extent are the results in Chapter 4 affected by this bias. The Bayesian framework partially mitigates this by incorporating uncertainty and allowing for a range of trajectories. The backtesting validation also provides some reassurance that forecasts are not systematically biased. However, the bias may still be present for forecasts made for solar PV and wind (especially solar PV), which remain in early exponential growth phases.

Single S-curves and the assumption of technology independence. The framework forecasts each technology along a single S-curve, treating its growth as independent of other technologies in the system. This is a strong simplification: solar PV and wind compete for shares of the same electricity market, while also acting as complements (e.g. diurnal/seasonal balancing, both benefiting from grid expansion and storage). Substitution-based frameworks such as the Fisher–Pry logistic substitution model [37] and the FTT framework [259, 260] represent these interactions endogenously, generating coupled dynamics that single-technology S-curves cannot. The independence assumption is

therefore most defensible during the early growth phase examined here, where the relevant constraint is the asymptote L of each technology rather than competition for a shared share. As renewables approach saturation in the electricity mix, however, substitution effects with each other and with fossil incumbents will become first-order, and the single-S-curve framing will break down. Coupling distributional S-curve forecasts with substitution dynamics is a natural extension (Section 6.4).

Computational inference method. Chapter 4 employs Metropolis-Hastings MCMC, which converges more slowly than modern gradient-based samplers (HMC, NUTS) for high-dimensional or strongly correlated parameter spaces. For the 3–5 parameter models used here this is adequate, but extensions to fully hierarchical multi-technology models would benefit from probabilistic programming frameworks (e.g. PyMC) that implement more efficient samplers and facilitate model specification. The choice of MCMC method does not affect the validity of results but may limit scalability and computational efficiency for more complex models.

Model specification and functional form. The thesis focuses on parametric S-curve models (logistic, Gompertz, B-R) that assume monotonic growth toward a fixed saturation level. This excludes multi-stage dynamics. Chapter 4 finds that the B-R curve with $\beta \approx 2/3$ provides superior fit and backtesting-performance across 120 technologies compared to logistic ($\beta = 1$) or Gompertz ($\beta \to 0$). Nonetheless, individual technology variation in β is not fully explained, and the thesis selects B-R based on fit rather than implementing Bayesian model averaging (BMA) across specifications.

Single-model selection and model uncertainty. Selecting a single functional form rather than implementing BMA [153, 451] risks underestimating forecast uncertainty: the posterior predictive distributions quantify parameter uncertainty conditional on the

chosen model but ignore model uncertainty. When different S-curves fit historical data similarly yet imply divergent long-term trajectories, selecting only one fails to reflect this structural uncertainty. Comparing three candidate models and selecting the best-fitting one also introduces specification search bias [180]: the reported fit may overstate true predictive accuracy. Chapter 4 partially mitigates both concerns through extensive out-of-sample hindcasting validation, but results of cross-technological validation may be subject to average performance across technologies rather than reflecting the true uncertainty for any given technology.

Prior specification and sensitivity. The quasi-empirical Bayesian approach developed in Chapter 4 grounds priors in data together with some subjective judgment. For well-observed technologies, posterior distributions are data-dominated and prior-insensitive. For nascent technologies with limited data, however, prior choices substantially influence forecasts. Hence, forecasts for technologies at early diffusion stages (e.g., solar PV, wind) are more sensitive to prior specification, and the wide credible intervals reflect this uncertainty. While Chapter 4 includes sensitivity analyses showing that forecasts are robust to reasonable variations in prior parameters, it also shows that extreme prior choices can produce implausible forecasts, underscoring the importance of careful prior specification and transparency about assumptions.

The empirical Bayes approach also makes double use of the data. For example we use the last observed data point to normalize the time-series to contain values between 0 and 1 in the training set. Then we set the location parameter of the prior distribution for the asymptote parameter L, and use the same data point again in the likelihood when fitting the model to that technology. This can lead to overfitting and underestimation of uncertainty[145], since the data inform both the prior and the likelihood [137]. A fully Bayesian hierarchical model would address this by jointly estimating hyperparameters and

technology-level parameters, but at greater computational cost.

Conditionality of scenario plausibility assessments. The plausibility assessment of IPCC AR6 scenarios in Chapter 5 is inherently conditional on the forecasts produced in Chapter 4: because the filtering criterion uses solar and wind forecasts as the empirical benchmark, scenarios identified as most plausible are by construction those with high solar and wind deployment. Chapter 4 demonstrates well-calibrated forecasts across 120 technologies, providing empirical justification for this benchmark, but the analysis should be understood as a first step applying the methodology to two critical technologies. Extending the framework to additional technologies and incorporating system-level dynamics, as outlined in Section 6.4, may reveal additional intricacies beyond those captured by solar and wind alone.

Warming outcomes are inherited from AR6, not independently modelled. The probability of limiting warming below 2°C reported in Chapter 5 is conditional on more than the renewable forecasts. We filter AR6 scenarios by their solar and wind trajectories and then read off the warming outcomes those scenarios already report; we do not separately model the carbon cycle, non-CO_2 forcing, climate sensitivity, or the deployment trajectories of negative-emission technologies (BECCS, DAC, afforestation). The likely subset's warming distribution therefore inherits the AR6 ensemble's representation of these processes. If AR6 systematically misrepresents CDR feasibility or carbon-cycle feedbacks, the warming outcomes within the likely subset would shift accordingly. The diffusion-based filter is informative about the empirical plausibility of the renewable side of each scenario, not about the climate-system response, and the warming claims should be interpreted with this division of labour in mind.

6.3.3 Scope Limitations

Energy system complexity. The thesis focuses on technology-level S-curve forecasting without explicitly modelling energy system feedbacks, infrastructure constraints, or policy dynamics. Chapter 5 applies solar + wind S-curve forecasts to assess IPCC scenario plausibility, but does not account for: (1) other novel and potentially disruptive technologies; (2) grid integration constraints (transmission capacity, storage requirements, grid stability); (3) supply chain and manufacturing scale-up (mineral supply chains, skilled labour); (4) cost-deployment feedbacks (learning curves creating positive feedbacks not endogenised in the forecasting framework); (5) policy endogeneity (successful technologies attracting subsidies and mandates, creating two-way causality); or (6) social and behavioural dynamics (heterogeneous consumer preferences, social network effects, and threshold/cascade dynamics). These factors may crucially determine real-world diffusion trajectories and thus the applicability of S-curve forecasts for energy transition planning. The thesis establishes a baseline for technology diffusion forecasting and assessing scenario plausibility through determining feasibility of AR6 scenarios. This approach relies on the AR6 scenarios sufficiently spanning the space of plausible futures.

Geographic aggregation. Most analyses operate at global aggregation levels following Way et al. [29], maximising sample size for cross-technology comparisons. This obscures geographic variation: Denmark has achieved 40% wind penetration and Norway 80% EV market share, far exceeding global averages, while renewables deployment in some developing nations lags behind. Global aggregation also conflates demand-side adoption with supply-side manufacturing (e.g., Chinese solar PV production vs. global deployment) and obscures policy shocks in major economies. This thesis prioritises establishing universal diffusion properties globally before addressing geographic heterogeneity.

Technology scope. The thesis focuses primarily on electricity generation technologies (solar PV, wind) and electric vehicles. Other critical decarbonisation technologies such as energy storage, industrial decarbonisation (steel, cement, chemicals), negative emissions (direct air capture, BECCS), and emerging technologies (advanced nuclear, enhanced geothermal, green hydrogen), receive less attention. This selective focus reflects data availability and the centrality of renewable electricity to energy transitions, but means conclusions about overall transition pathways should be interpreted with caution. The diffusion patterns of other technologies may differ, and their interactions with solar and wind could crucially shape transition trajectories.

6.3.4 Structural Stability and Future Growth Behaviour

Mentioned previously, Lucas [452] argues that forecasting methods relying on historical data assume structural stability: that the underlying processes generating the data remain consistent over time. If future diffusion dynamics differ fundamentally from historical patterns due to novel technologies, unprecedented policy interventions, or transformative social changes, then forecasts based on past data may be systematically inaccurate. As Paltsev [246] observes, *if history is any guide, energy scenarios overestimate the extent to which the future will look like the recent past.*

Four specific challenges arise. (1) *saturation mechanisms*: renewable saturation could be driven by electricity demand limits, land availability, grid stability, or political acceptability. (2) *non-stationarity*: the models assume stationary parameters (L, k, β), but structural breaks, such as China's solar PV manufacturing scale-up or future policy shocks, could induce regime shifts; or as per Tankwa et al. [204] policy changes yearly growth rates (i.e. affecting k_i). Chapter 4's oil crisis backtest showed some robustness to historical shocks, but unprecedented disruptions may differ in magnitude. (3) *technology interactions*: the thesis forecasts individual trajectories without modelling complementarities (solar

and storage) or substitution effects (renewables displacing coal), which may crucially determine outcomes. Network substitution models such as FTT [259, 260] address this by modelling competitive dynamics endogenously (Section 2.3). This thesis instead prioritises parsimony and calibration over system-level complexity. (4) *unprecedented diffusion speed*: solar PV and wind are diffusing faster than any historical energy technology, resembling consumer electronics more than traditional energy transitions [211]. If these technologies are genuinely unprecedented, their resulting adoption shape may differ fundamentally from historical patterns, making forecasts based on past data less reliable, and cross-technological regularities less applicable.

This limitation is shared by all data-driven forecasting methods. The approach adopted in this thesis is more empirically grounded than IAM expert judgment and quantifies uncertainty probabilistically rather than deterministically [23, 30]. Nonetheless, the assumption of structural stability represents an epistemological limitation that should inform how results are applied under deep uncertainty.

6.3.5 Aggregation Level and the S-Curve Assumption

This thesis relies on the assumption that technologies, broadly defined to include social, economic, and institutional innovations, follow S-curve diffusion patterns at some aggregation level. In the appendix to Chapter 5, we demonstrate that the sum of two or more S-curves most likely does **not** follow an S-curve. This is an important limitation: if some aggregation level, be it cross-country, cross-technology, or cross-sector, presents the *true* unit at which diffusion follows an S-curve, then any other aggregation level would likely not.

The implications reach across the whole thesis. Chapter 3 relies on the assumption that some underlying diffusion process follows an S-curve, though not on a specific functional form. Chapter 4 gathers data on 120 technologies across multiple sectors, countries, and

aggregation levels. For example, the dataset contains mobile phones and computers, which share underlying components, raising the question of whether the components or the final products represent the appropriate S-curve unit. Country-level and global-level data are also mixed, raising the question of whether diffusion follows S-curves at the country level, global level, or some intermediate level. Chapter 5 evaluates IPCC AR6 scenarios using individual technology forecasts, then sums them. If solar PV and wind individually follow S-curves, their combined generation likely does not. For the sake of applying forecasts, Chapter 5 makes the strong and unrealistic assumption that the technologies are independent, or that any interdependencies are captured in the historical data.

Are S-curves a good approximation? At what level of aggregation do S-curves emerge, and can we distinguish levels where S-curves are a good approximation from those where they are not? Noisy data may make it empirically difficult to distinguish between these levels. Developing methods to identify the appropriate aggregation level for S-curve modelling would enhance the applicability of the framework and represents an important avenue for future research.

The thesis limits itself to parametric S-curve models and does not explore alternative functional forms. It therefore makes no claim that S-curves are the *right* model for all technology diffusion. Rather, it documents the remarkable cross-technological empirical regularities and predictive power that S-curves provide within the domains examined.

6.3.6 Mechanistic Understanding: Observation versus Explanation

This thesis observes empirical S-curve patterns, characterises regularities across technologies, and develops probabilistic forecasting methods based on these regularities. It does not explain *why* technologies follow specific S-curve shapes. Formal derivation of specific

functional forms (e.g., B-R curve with $\beta \approx 2/3$) from underlying social, economic, or network processes remains an open challenge (Section 2.1 reviews candidate mechanisms). Testing whether frameworks like complex contagion or economic transition theories can formally predict observed diffusion parameters would advance understanding beyond empirical description toward causal explanation, but exceeds the scope of this thesis.

6.4 Future Directions

Sections 6.1–6.3 suggest three broad avenues for future work: extending the scope of the forecasting approach, refining the inference methodology, and integrating the framework with system-level energy transition models.

6.4.1 Extending Scope and Granularity

Regional heterogeneity. This thesis's global focus obscures substantial geographic variation in diffusion rates and saturation levels. Hierarchical models that nest country-level S-curves within global distributions could share information across regions while permitting heterogeneous parameters driven by resource endowments, policy environments, and infrastructure constraints. Such models would enable more granular policy analysis. For example, it could be used to assess the plausibility of rapid solar deployment in India or offshore wind expansion in Europe, and reveal how regional trajectories combine to produce global outcomes.

Broader applications and model adequacy. This thesis focuses its application on energy generation, but the S-curve framework may extend to demand-side transformations such as heat pumps and energy efficiency measures, enabling a better assessment of scenarios like the IMP-LD 40% which assumes major technology induced demand reductions [68]. S-curve dynamics also appear in behavioural adoption (e.g., electric vehicles), social

movements (e.g., divestment campaigns), and epidemiological processes (e.g., vaccination uptake). These domains involve fundamentally different generative mechanisms. Network externalities and supply-side constraints for technologies, and social learning for behavioural norms, or biological contagion for epidemics. This raises a deeper question: when do S-curves emerge from first principles, and when do they fail? Whether the B-R curve with $\beta \approx 2/3$ finding generalises beyond technology would clarify whether it reflects a universal property of diffusion or a contingent feature of the technology dataset. Answering this requires both theoretical work, such as deriving S-curve parameters from micro-level interaction models, and empirical testing across non-technological domains.

Richer data. Many limitations identified in Section 6.3 are ultimately data constraints. Expanding the dataset with early-stage technologies, failed technologies that plateaued at low adoption, and consistently aggregated measurements would strengthen the cross-technology regularities and reduce survivorship bias. Finer-grained data, disaggregated by country or sector, would test whether S-curve regularities hold at sub-global levels or emerge only at sufficiently aggregate scales where heterogeneity averages out. Standardised, open-access technology diffusion databases, analogous to climate data repositories, would accelerate progress across the field.

6.4.2 Methodological Extensions

Probabilistic programming languages such as Stan [453] implement gradient-based samplers (HMC, NUTS) that explore posterior geometries more efficiently than Metropolis-Hastings. These tools would enable fully hierarchical multi-technology models that simultaneously estimate technology-level parameters and hyperparameters, addressing the empirical Bayes double-use-of-data concern (Section 6.3) and improving sampling efficiency for more complex specifications such as coupled S-curve models or cost-diffusion feedback models.

Sequential Bayesian updating, where each period's posterior becomes the next period's prior [150], would transform the static forecasts in this thesis into operational monitoring tools. Simulation-based calibration [322] would provide rigorous diagnostics for these more complex specifications.

6.4.3 Towards Probabilistic Transition Models

The most consequential extension of this work could be to move from forecasting individual technologies in isolation toward modelling energy transitions as integrated systems with calibrated uncertainty. However, the increase in complexity this would generate is potentially intractable.

Technology interactions. This thesis forecasts each technology independently, ignoring complementarities (solar, storage, and wind) and substitution effects (also possible with solar and wind) that shape transition trajectories. Existing models address this from different directions: the FTT framework [259, 260] couples logistic substitution dynamics with macroeconomic feedbacks, while agent-based models (Section 2.3) represent heterogeneous actors and the network dynamics that drive diffusion. Both generate richer transition dynamics than isolated S-curves but do not directly quantify the uncertainty in technological growth. Coupling distributional S-curve forecasts with interaction models, such as FTT-type or agent-based models, would capture how, for example, storage deployment accelerates solar adoption, while calibrating the uncertainty in that interaction. Integrating cost-diffusion feedbacks would further enable modelling of endogenous learning effects.

Probabilistic calibration of complex models. Approximate Bayesian Computation (ABC) [454] offers a natural bridge between the parsimonious framework of this thesis and system-level models. For models where likelihoods are intractable, ABC can calibrate complex simulators against empirical diffusion patterns while simulation-based calibration

methods verify the resulting posteriors.

Extending scenario evaluation. The probabilistic filtering method of Chapter 5 can be applied to other scenario ensembles as we have to the IEA scenarios, and extended to evaluate cost assumptions alongside technology diffusion, enabling multi-dimensional plausibility assessment with probabilistic scenario weights. More broadly, complex models such as IAMs, ABMs, and FTT-type frameworks could be calibrated against the empirical regularities and forecasts of Chapter 4, embedding these as empirical priors on diffusion rates within scenario frameworks. A natural near-term extension is to incorporate energy storage into the likelihood calculation in Chapter 5. The current implementation evaluates scenarios against solar PV and wind growth alone, but storage deployment is increasingly load-bearing for high-renewables pathways, both as a complement to variable generation and as an indicator of system readiness. Including storage forecasts in the likelihood would tighten the empirical constraint on scenarios that assume rapid renewable shares without commensurate storage build-out, and would mitigate the bias against scenarios that rely on flexibility resources beyond solar and wind.

Decoupling warming outcomes from the AR6 climate response. A second extension addresses the limitation that warming statements in Chapter 5 are inherited from each AR6 scenario rather than independently derived. Coupling the diffusion-based filter with a transparent reduced-complexity climate-system module (such as MAGICC or FaIR) would let the warming distribution be re-derived from the diffusion-filtered emissions trajectories under explicit, varied assumptions about carbon-cycle feedbacks, non-CO_2 forcing, climate sensitivity, and the scale of negative-emission technologies (BECCS, DAC, afforestation). This would convert the warming claims from *conditional on AR6's climate response* to *conditional on transparently chosen carbon-cycle and CDR assumptions*, and would quantify how sensitive the $< 2°C$ probability is to those assumptions independently

of the renewable filter.

A generative probabilistic model of energy transitions. The most ambitious integration would model the joint distribution of all key technologies including complementarities, substitution effects, and fossil fuel phase-out, with endogenous cost-diffusion feedbacks, producing fully probabilistic transition trajectories consistent with historical diffusion patterns. This would build on FTT-type models that already use logistic dynamics endogenously [259], but add calibrated uncertainty bounds informed by the cross-technology regularities of Chapter 4. Chapter 5 filters existing scenarios; a fully generative approach would produce probabilistic scenarios directly, moving beyond deterministic IAM projections and the conditional scenarios of Way et al. [29] towards a comprehensive probabilistic model of energy system evolution. In the terminology adopted in Chapter 1, this would shift the assessment of energy transitions from deterministic scenarios and projections towards fully probabilistic, conditionally calibrated forecasts of the integrated system, providing a single coherent object that decision-makers can interrogate for likelihood as well as for what-if exploration.

Chapter 7

Conclusion

7.1 Summary of Research

This thesis set out to answer the research question: *How can S-curves be used to better understand and forecast technological change?* Three papers together identify and explain systematic underestimation bias in S-curve fitting, develop a Bayesian forecasting framework grounded in empirical regularities across 120 technologies, and apply these forecasts to evaluate the plausibility of IPCC AR6 scenarios against historical diffusion patterns.

Chapter 3 answered the first sub-question: *Why have S-curves failed to accurately predict diffusion patterns?* It demonstrated that standard S-curve fitting with partial data produces systematic downward bias and highly varied forecasts, providing a statistical explanation for why S-curve forecasts have historically underestimated renewable deployment. The debiasing method developed offers a practical correction applicable to any S-curve forecasting context, shifting assessments of feasibility for rapid technology growth.

Chapter 4 answered the second sub-question: *How universal and predictable is technological growth?* It compiled a database of 120 historical growth trajectories, uncovering remarkable empirical regularities across diverse technologies. These regularities enabled the development of a technology-agnostic Bayesian forecasting framework that produces well-calibrated distributional forecasts. The Richards curve with $\beta \approx 2/3$ emerged as the significantly better-fitting functional form, establishing mild asymmetry and early exponen-

tial growth as typical of technology diffusion. Hindcasting validation demonstrated that empirically calibrated forecasts reasonably capture technologies experiencing exogenous shocks. For solar, median projections exceed IEA Net Zero 2050 targets and nearly all AR6 scenarios; for wind, the opposite holds.

Chapter 5 answered the third sub-question: *What are likely future energy transitions?*, by applying forecasts from Chapter 4 to evaluate the plausibility of IPCC AR6 scenarios. A likelihood-based evaluation method narrows the full AR6 ensemble to a smaller, more meaningful subset. Rapid decarbonisation pathways align most closely with empirical regularities, while many scenarios assume solar growth rates below historical precedent. Under trajectories consistent with observed diffusion, SSP1, SSP2, and SSP5 scenarios are most and almost equally represented. Warming is likely limited to 2°C with $>$ 80% probability, with future energy transitions characterised by rapid electrification and renewable deployment, a significant decline in fossil fuel share by 2050, and a non-trivial role for carbon capture and storage.

7.2 Key Contributions

This thesis makes the following contributions to the fields of technology forecasting, energy systems analysis, and climate policy:

1. **Identification and correction of underestimation bias.** Chapter 3 provided the first systematic statistical explanation for the persistent underestimation of renewable energy deployment in S-curve forecasts, showing that fitting with partial data produces downward-biased and highly variable projections. The parametric bootstrap debiasing method offers a practical correction applicable to any S-curve forecasting context.

2. **Empirically grounded Bayesian S-curve framework.** Chapter 4 demonstrated

that technologies across sectors, eras, and geographic scales exhibit remarkable regularities in their diffusion patterns, and developed a Bayesian forecasting framework that leverages these regularities to produce well-calibrated distributional forecasts.

3. **Forecasting technologies.** The B-R curve with $\beta \approx 2/3$ provides superior out-of-sample fit across 120 technologies, suggesting it as a robust default model for technology diffusion forecasting. The finding that $\beta < 1$ implies that growth rates tend to slow more gradually than in the logistic model, with important implications for long-term projections.

4. **Forecasts for solar and wind energy.** Chapter 4 showed that rapid solar growth exceeding Net Zero 2050 targets and nearly all AR6 scenarios is consistent with historical diffusion patterns, while wind growth consistent with historical patterns falls below many AR6 scenarios. These forecasts provide an empirical benchmark for assessing transition pathway plausibility.

5. **Probabilistic assessment of IPCC scenario likelihood.** Chapter 5 introduced a likelihood-based method for evaluating scenario ensembles against empirically grounded forecasts, adding an empirical-plausibility dimension to scenario interpretation.

6. **Complementarity between forecasts and scenarios.** The thesis establishes probabilistic S-curve forecasts as a complement to IAM scenarios: scenarios explore what policy *could* achieve; forecasts assess what historical trends *suggest*. This framing enables risk-based planning using credible intervals rather than discrete scenario pathways.

7.3 Final Remarks

Energy forecasting faces a fundamental tension: scenarios are needed to guide policy under deep uncertainty, yet the tools producing those scenarios have systematically underestimated the technologies most critical to decarbonisation. This thesis investigated technological change, how S-curves can be used to understand and forecast it, and the implications for energy transitions. The key finding is that grounding forecasts in empirical diffusion regularities can yield well-calibrated forecasts across technologies, which can be used to assess the plausibility of energy transition scenarios.

Three broader insights emerge. First, S-curves have a poor track record as forecasting tools, but this reflects methodological shortcomings rather than inherent limitations; with appropriate methods they become powerful instruments for understanding and forecasting technological change. Second, the historical record of technology diffusion contains exploitable regularities, such as technologies diffusing asymmetrically. These cross-technological regularities enable well-calibrated, technology-agnostic forecasts. Third, rapid energy transitions are consistent with historical diffusion patterns, challenging the conventional view that rapid renewable growth is unprecedented or economically prohibitive.

The methods and findings of this thesis provide a foundation for more empirically grounded, probabilistic approaches to understanding and forecasting technological change, with potential applications wherever adoption follows Sigmoidal trajectories, from the diffusion of behaviours and social norms to institutional change. Moreover, the thesis offers a framework for letting historical data discipline forward-looking projections: not replacing expert-driven models but providing an empirical benchmark against which their assumptions can be transparently evaluated. This complementarity, between what is empirically grounded and what is captured by scenarios, offers a more honest foundation for the urgent decisions that will shape the global energy system for decades to come.

Bibliography

[1] GP Pokhariyal and AJ Rodrigues. An accurate epidemiological model. *Applied mathematics and computation*, 53(1):1–12, 1993.

[2] Angela R McLean. Infectious disease modeling. In *Infectious diseases*, pages 9–20. Springer, 2023.

[3] Anna Kane Laird. Dynamics of tumour growth. *British journal of cancer*, 18(3):490, 1964.

[4] Frank M Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969.

[5] Everett M Rogers. Elements of diffusion. *Diffusion of innovations*, 5(1.38):453, 2003.

[6] Thomas W Valente. Social network thresholds in the diffusion of innovations. *Social networks*, 18(1):69–89, 1996.

[7] Arnulf Grübler. Diffusion: long-term patterns and discontinuities. In *Diffusion of technologies and social behavior*, pages 451–482. Springer, 1991.

[8] Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522, 1957.

[9] Derek de Solla Price. *Little Science, Big Science*. Columbia University Press, New York, 1963.

[10] Jean Tague, Jamshid Beheshti, and Lorna Rees-Potter. The law of exponential growth: evidence, implications and forecasts, 1981.

[11] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies: in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, 115:513–585, 1825.

[12] Verhulst Pierre-François. Notice sur la loi que la population poursuit dans son accroissement. *Corresp. Math. Phys*, 10:113–121, 1838.

[13] Francis J Richards. A flexible growth function for empirical use. *Journal of experimental Botany*, 10(2):290–301, 1959.

[14] Anastasios Tsoularis and James Wallace. Analysis of logistic growth models. *Mathematical biosciences*, 179(1):21–55, 2002.

[15] Arnulf Grübler. Time for a change: on the patterns of diffusion of innovation. *Daedalus*, 125(3):19–42, 1996.

[16] Arnulf Grübler. *Technology and Global Change*. Cambridge University Press, 1998.

[17] Hanna Cardol, Ingrid Mignon, and Björn Lantz. Rethinking the forecasting of innovation diffusion: A combined actor-and system approach. *Technological Forecasting and Social Change*, 214:124058, 2025.

[18] Christophe Van den Bulte and Gary L Lilien. Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science*, 16(4):338–353, 1997.

[19] Nigel Meade. The use of growth curves in forecasting market development—a review and appraisal. *Journal of Forecasting*, 3(4):429–451, 1984.

[20] Nigel Meade and Towhidul Islam. Technological forecasting—model selection, model stability, and combining models. *Management science*, 44(8):1115–1130, 1998.

[21] Giovani L Vasconcelos, Antônio MS Macêdo, Raydonal Ospina, Francisco AG Almeida, Gerson C Duarte-Filho, Arthur A Brum, and Inês CL Souza. Modelling fatality curves of covid-19 and the effectiveness of intervention strategies. *PeerJ*, 8:e9421, 2020.

[22] Stacey Kung, Marjan Doppen, Melissa Black, Irene Braithwaite, Ciléin Kearns, Mark Weatherall, Richard Beasley, and Nethmi Kearns. Underestimation of covid-19 mortality during the pandemic. *ERJ open research*, 7(1), 2021.

[23] Béla Nagy, J Doyne Farmer, Quan M Bui, and Jessika E Trancik. Statistical basis for predicting technological progress. *PloS one*, 8(2):e52669, 2013.

[24] Matthew Ives, Luca Righetti, Johanna Schiele, Kris De Meyer, Lucy Hubble-Rose, Fei Teng, Lucas Kruitwagen, Leah Tillmann-Morris, Tianpeng Wang, Rupert Way, et al. A new perspective on decarbonising the global energy system. *Smith School of Enterprise and the Environment, University of Oxford*, 2021.

[25] Philip E Tetlock. Expert political judgment: How good is it? how can we know?-new edition, 2017.

[26] Paris agreement, 2015.

[27] Core Writing Team, H. Lee, and J. (eds.) Romero. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 2023.

[28] Felix Creutzig, Peter Agoston, Jan Christoph Goldschmidt, Gunnar Luderer, Gregory Nemet, and Robert C Pietzcker. The underestimated potential of solar energy to mitigate climate change. *Nature Energy*, 2(9):1–9, 2017.

[29] Rupert Way, Matthew C Ives, Penny Mealy, and J Doyne Farmer. Empirically grounded technology forecasts and the energy transition. *Joule*, 6(9):2057–2082, 2022.

[30] J Doyne Farmer and François Lafond. How predictable is technological progress? *Research Policy*, 45(3):647–665, 2016.

[31] C Lennart Baumgärtner, Rupert Way, Matthew C Ives, and J Doyne Farmer. The need for better statistical testing in data-driven energy technology modeling. *Joule*, 8(9):2453–2466, 2024.

[32] Stefan Pfenninger, Adam Hawkes, and James Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, 2014.

[33] Hans-Kristian Ringkjøb, Peter M. Haugan, and Ida Marie Solbrekke. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96:440–459, 2018.

[34] Evelina Trutnevyte. Does cost optimization approximate the real-world energy

transition? *Energy*, 106:182–193, 2016.

[35] Arnulf Grübler and Nebojša Nakićenović. Long waves, technology diffusion, and substitution. *Review (Fernand Braudel Center)*, 14(2):313–343, 1991.

[36] Arnulf Grübler, Nebojša Nakićenović, and David G Victor. Dynamics of energy technologies and global change. *Energy policy*, 27(5):247–280, 1999.

[37] John C Fisher and Robert H Pry. A simple substitution model of technological change. *Technological forecasting and social change*, 3:75–88, 1971.

[38] John Bistline, Morgan Browning, Julianne DeAngelo, Daniel Huppmann, Ryan Jones, James McFarland, Anahi Molar-Cruz, Steven Rose, and Steven J Davis. Uses and limits of national decarbonization scenarios to inform net-zero transitions. *Joule*, 8(10):2721–2726, 2024.

[39] Céline Guivarch, Thomas Le Gallic, Nico Bauer, Panagiotis Fragkos, Daniel Huppmann, Marc Jaxa-Rozen, Ilkka Keppo, Elmar Kriegler, Tamás Krisztin, Giacomo Marangoni, et al. Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change*, 12(5):428–435, 2022.

[40] Everett M. Rogers. *Diffusion of Innovations*. Free Press, 5th edition, 2003.

[41] Vijay Mahajan, Eitan Muller, and Frank M Bass. New product diffusion models in marketing: A review and directions for research. *Journal of marketing*, 54(1):1–26, 1990.

[42] Mohsen Bahmani-Oskooee and Scott W Hegerty. The j-and s-curves: A survey of the recent literature. *Journal of Economic Studies*, 37(6):580–596, 2010.

[43] Maria Priestley, Timothy J Sluckin, and Thanassis Tiropanis. Innovation on the web: the end of the s-curve? *Internet Histories*, 4(4):390–412, 2020.

[44] Dmitry Kucharavy and Roland De Guio. Application of s-shaped curves. *Procedia Engineering*, 9:559–572, 2011.

[45] Diana Crane. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. University of Chicago Press, Chicago, 1972.

[46] Subhash C Bhargava. A generalized form of the fisher-pry model of technological substitution. *Technological Forecasting and Social Change*, 49(1):27–33, 1995.

[47] Roger Fouquet. The slow search for solutions: lessons from historical energy transitions by sector and service. *Energy Policy*, 38(11):6586–6596, 2010.

[48] Diego Comin and Bart Hobijn. Cross-country technology adoption: making the theories face the facts. *Journal of monetary Economics*, 51(1):39–83, 2004. data retrieved from: https://www.nber.org/research/data/historical-cross-country-technology-adoption-hccta-dataset.

[49] Arnulf Grubler. *The rise and fall of infrastructures: dynamics of evolution and technological change in transport*. Physica-Verlag, 1990.

[50] Paul A Geroski. Models of technology diffusion. *Research policy*, 29(4-5):603–625, 2000.

[51] BP Hills and KM Wright. A new model for bacterial growth in heterogeneous systems. *Journal of theoretical biology*, 168(1):31–41, 1994.

[52] A Pilkington. Modelling the diffusion of innovation management theory using s-curves. In *2008 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 920–924. IEEE, 2008.

[53] Dominique Foray and Arnulf Grübler. Technology and the environment: an overview. *Technological forecasting and social change*, 53(1):3–13, 1996.

[54] Bo Carlsson. *Technological systems and industrial dynamics*, volume 10. Springer Science & Business Media, 1997.

[55] Bo Carlsson. Four technological systems: What have we learned? In *Technological Systems and Industrial Dynamics*, pages 279–298. Springer, 1997.

[56] Stephen Jay Gould. The structure of evolutionary theory. In *The Structure of Evolutionary Theory*. Harvard university press, 2002.

[57] W Brian Arthur. Complexity and the economy. In *Handbook of Research on Complexity*. Edward Elgar Publishing, 2009.

[58] John Kruschke. *Doing Bayesian data analysis*. Academic Press, 2014.

[59] Irina Kareva and Georgy Karev. *Modeling Evolution of Heterogeneous Populations: Theory and Applications*. Academic Press, 2019.

[60] M Heinen. Analytical growth equations and their genstat 5 equivalents. *Netherlands Journal of Agricultural Science*, 47(1):67–89, 1999.

[61] Malcolm E Turner Jr, Edwin L Bradley Jr, Katherine A Kirk, and Kenneth M Pruitt. A theory of growth. *Mathematical Biosciences*, 29(3-4):367–373, 1976.

[62] Malcolm E Turner Jr, Brent A Blumenstein, and Jeanne L Sebaugh. 265 note: A generalization of the logistic law of growth. *Biometrics*, pages 577–580, 1969.

[63] JA Nelder. The fitting of a generalization of the logistic curve. *Biometrics*, 17(1):89–110, 1961.

[64] Ludwig von Bertalanffy. Quantitative Laws in Metabolism and Growth. *The Quarterly Review of Biology*, 32(3):217–231, September 1957. Publisher: The University of Chicago Press.

[65] William S Hoar, David J Randall, and Edward M Donaldson. *Fish physiology*. Academic Press, 1983.

[66] Wenxin Ye, Kangxin An, Shihui Zhang, Zihan Zhen, Wenjia Cai, and Can Wang. Probabilistic analysis of tripling global renewables based on bayesian-inferred growth dynamics model. *Renewable Energy*, page 124496, 2025.

[67] Eric W Weisstein. Logistic equation. *https://mathworld. wolfram. com/*, 2003.

[68] Arnulf Grubler, Charlie Wilson, Nuno Bento, Benigna Boza-Kiss, Volker Krey, David L McCollum, Narasimha D Rao, Keywan Riahi, Joeri Rogelj, Simon De Stercke, et al. A low energy demand scenario for meeting the 1.5 c target and sustainable development goals without negative emission technologies. *Nature energy*, 3(6):515–527, 2018.

[69] Aleh Cherp, Vadim Vinichenko, Jale Tosun, Joel A Gordon, and Jessica Jewell. National growth dynamics of wind and solar power compared to the growth required for global climate targets. *Nature Energy*, 6(7):742–754, 2021.

[70] Birgitte Andersen. The hunt for s-shaped growth paths in technological innovation: a patent study. *Journal of evolutionary economics*, 9(4):487–526, 1999.

[71] LCM Miranda and CAS Lima. On the logistic modeling and forecasting of evolutionary processes: Application to human population dynamics. *Technological Forecasting and Social Change*, 77(5):699–711, 2010.

[72] Trina Roy, Sinchan Ghosh, Bapi Saha, and Sabyasachi Bhattacharya. A noble

extended stochastic logistic model for cell proliferation with density-dependent parameters. *Scientific Reports*, 12(1):8998, 2022.

[73] Efim Pelinovsky, Andrey Kurkin, Oxana Kurkina, Maria Kokoulina, and Anastasia Epifanova. Logistic equation and covid-19. *Chaos, Solitons & Fractals*, 140:110241, 2020.

[74] Masayuki Kakehashi and Shoko Kawano. Fundamentals of mathematical models of infectious diseases and their application to data analyses. In *Handbook of statistics*, volume 36, pages 3–45. Elsevier, 2017.

[75] IS Shruti. A biological growth curve is a sum of two distinct s-curves. *bioRxiv*, pages 2025–02, 2025.

[76] Tomonori Kawano, Nigel Wallbridge, and Carrol Plummer. Logistic models for simulating the growth of plants by defining the maximum plant size as the limit of information flow. *Plant Signaling & Behavior*, 15(2):1709718, 2020.

[77] Md Aktar Ul Karim, Vikram Aithal, and Amiya Ranjan Bhowmick. Random variation in model parameters: A comprehensive review of stochastic logistic growth equation. *Ecological Modelling*, 484:110475, 2023.

[78] Robert U Ayres. Technological forecasting and long-range planning. *McGraw-Hill*, 1969.

[79] R Coenen. The use of technological forecasts in government planning. *Research Policy*, 1(2):156–172, 1972.

[80] E Pelinovsky, M Kokoulina, A Epifanova, A Kurkin, O Kurkina, M Tang, E Macau, and M Kirillin. Gompertz model in covid-19 spreading simulation. *Chaos, Solitons & Fractals*, 154:111699, 2022.

[81] EO Alzahrani, Asim Asiri, MM El-Dessoky, and Yang Kuang. Quiescence as an explanation of gompertzian tumor growth revisited. *Mathematical biosciences*, 254:76–82, 2014.

[82] Praveen Asthana. Jumping the technology s-curve. *Ieee Spectrum*, 32(6):49–54, 1995.

[83] MIE Halmi, MS Shukor, WLW Johari, and MY Shukor. Evaluation of several mathematical models for fitting the growth of the algae dunaliella tertiolecta. *Asian Journal of Plant Biology*, 2(1):1–6, 2014.

[84] Kathleen MC Tjørve and Even Tjørve. The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family. *PloS one*, 12(6):e0178691, 2017.

[85] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John ML Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8):e1003800, 2014.

[86] CE Timothy Paine, Toby R Marthews, Deborah R Vogt, Drew Purves, Mark Rees, Andy Hector, and Lindsay A Turnbull. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods in Ecology and Evolution*, 3(2):245–256, 2012.

[87] SE Aggrey. Comparison of three nonlinear and spline regression models for describing chicken growth curves. *Poultry science*, 81(12):1782–1788, 2002.

[88] J Matthias Starck and Robert E Ricklefs. *Avian growth and development: evolution within the altricial-precocial spectrum*. Oxford University Press, 1998.

[89] Guy E. Skinner, John W. Larkin, and E. Jeffery Rhodehamel. Mathematical modeling of microbial growth: a review. *Journal of food safety*, 14(3):175–217, 1994.

[90] Marcel H Zwietering, Ida Jongenburger, Frank M Rombouts, and Kjaem Van't Riet. Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881, 1990.

[91] Diego Comin and Marti Mestieri. Technology diffusion: Measurement, causes, and consequences. In *Handbook of economic growth*, volume 2, pages 565–622. Elsevier, 2014.

[92] Mohammad Bakher Naseri and Greg Elliott. The diffusion of online shopping in australia: Comparing the bass, logistic and gompertz growth models. *Journal of Marketing Analytics*, 1:49–60, 2013.

[93] Daisuke Satoh. Model selection among growth curve models that have the same number of parameters. *Cogent Mathematics & Statistics*, 6(1):1660503, 2019.

[94] Tatcha Sudtasan and Hitoshi Mitomo. Comparison of diffusion models for forecasting the growth of broadband markets in thailand. In *14th ITS Asia-Pacific Regional Conference, Kyoto 2017: Mapping ICT into Transformation for the Next Information Society*. International Telecommunications Society (ITS), 2017.

[95] S Vieira and R Hoffmann. Comparison of the logistic and the gompertz growth functions considering additive and multiplicative error terms. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 26(2):143–148, 1977.

[96] M Dhar and P Bhattacharya. Comparison of the logistic and the gompertz curve under different constraints. *Journal of Statistics and Management Systems*, 21(7):1189–1210, 2018.

[97] Elvan Akın, Neslihan Nesliye Pelen, Ismail Uğur Tiryaki, and Fusun Yalcin. Parameter identification for gompertz and logistic dynamic equations. *PLoS One*, 15(4):e0230582, 2020.

[98] Manohar Nagula. Forecasting of fuel cell technology in hybrid and electric vehicles using gompertz growth curve. *Journal of Statistics and Management Systems*, 19(1):73–88, 2016.

[99] Robert U Ayres. Turning point: The end of exponential growth? *Technological Forecasting and Social Change*, 73(9):1188–1203, 2006.

[100] Dennis Meadows and Jorgan Randers. *The limits to growth: the 30-year update*. Routledge, 2012.

[101] Chauncey Starr and Richard Rudman. Parameters of technological growth. *Science*, 182(4110):358–364, 1973.

[102] Fred Hirsch. *Social limits to growth*. Taylor & Francis, 2025.

[103] Clayton M Christensen, M Raynor, and R McDonald. 17. disruptive innovation. *Harvard business review*, 93(12):44–53, 2015.

[104] Joe Anderson. Technology foresight for competitive advantage. *Long Range Planning*, 30(5):665–677, 1997.

[105] Fred Phillips. On s-curves and tipping points. *Technological Forecasting and Social Change*, 74(6):715–730, 2007.

[106] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–

59, 2009.

[107] Moshe Elitzur, Scott Kaplan, Željko Ivezić, and David Zilberman. The impact of policy timing on the spread of covid-19. *Infectious Disease Modelling*, 6:942–954, 2021.

[108] Stijn Baert, Louis Lippens, Eline Moens, Philippe Sterkens, and Johannes Weytjens. The covid-19 crisis and telework: A research survey on experiences, expectations and hopes. Technical report, IZA Discussion Papers, 2020.

[109] Xiang-Sheng Wang, Jianhong Wu, and Yong Yang. Richards model revisited: Validation by and application to infection dynamics. *Journal of theoretical biology*, 313:12–19, 2012.

[110] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2002.

[111] Pedro Domingos. The role of occam's razor in knowledge discovery. *Data mining and knowledge discovery*, 3(4):409–425, 1999.

[112] George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[113] Avi Jakhmola, Jessica Jewell, Vadim Vinichenko, and Aleh Cherp. Probabilistic projections of global wind and solar power growth based on historical national experience. *Nature Energy*, 11(5):743–755, 2026.

[114] Nik Zielonka and Evelina Trutnevyte. Probabilities of reaching required diffusion of granular energy technologies in european countries. *iScience*, 28(2), 2025.

[115] Geoffrey A Moore. *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers*. HarperBusiness, New York, 3rd edition, 2014.

[116] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[117] Malcolm Gladwell. *The tipping point: How little things can make a big difference*. Little, Brown, 2000.

[118] Diego Comin, Bart Hobijn, and Emilie Rovito. Five facts you need to know about technology diffusion. NBER Working Paper 11928, National Bureau of Economic Research, 2006.

[119] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.

[120] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.

[121] Douglas Guilbeault, Joshua Becker, and Damon Centola. Complex contagions: A decade in review. *Complex spreading phenomena in social systems: Influence and contagion in real-world social networks*, pages 3–25, 2018.

[122] H Peyton Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American economic review*, 99(5):1899–1924, 2009.

[123] W Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):116–131, 1989.

[124] Michael L Katz and Carl Shapiro. Network externalities, competition, and compatibility. *The American Economic Review*, 75(3):424–440, 1985.

[125] M Granger Morgan. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National academy of Sciences*, 111(20):7176–7184, 2014.

[126] Tamara Savage, Alex Davis, Baruch Fischhoff, and M Granger Morgan. A strategy to improve expert technology forecasts. *Proceedings of the National Academy of Sciences*, 118(21):e2021558118, 2021.

[127] Don A Moore and Paul J Healy. The trouble with overconfidence. *Psychological review*, 115(2):502, 2008.

[128] Philip E Tetlock. *Expert political judgment: How good is it? How can we know?* Princeton University Press, 2005.

[129] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974.

[130] Joseph Engelberg, Charles F Manski, and Jared Williams. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41, 2009.

[131] Jing Meng, Rupert Way, Elena Verdolini, and Laura Diaz Anadon. Comparing expert elicitation and model-based probabilistic technology cost forecasts for the energy transition. *Proceedings of the National Academy of Sciences*, 118(27):e1917165118, 2021.

[132] Paul S Meyer, John W Yung, and Jesse H Ausubel. "bi-logistic" growth. *Technological Forecasting and Social Change*, 47(1):89–102, 1994.

[133] Charles P Winsor. The gompertz curve as a growth curve. *Proceedings of the national academy of sciences*, 18(1):1–8, 1932.

[134] George AF Seber and Christopher J Wild. *Nonlinear regression*. John Wiley & Sons, 2003.

[135] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360–1383, 2008.

[136] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 3rd edition, 2013.

[137] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC Press, 3rd edition, 2008.

[138] Andrew Gelman and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. *Cambridge University Press*, 2006.

[139] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534, 2006.

[140] Rens Van De Schoot, Joris J Broere, Koen H Perryck, Mariëlle Zondervan-Zwijnenburg, and Nancy E Van Loey. Analyzing small data sets using bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European journal of psychotraumatology*, 6(1):25216, 2015.

[141] Patrice Dion, Nora Galbraith, and Elham Sirag. Using expert elicitation to build long-term projection assumptions. *Developments in demographic forecasting*, pages 43–62, 2020.

[142] Jeremy E Oakley and Anthony O'Hagan. Uncertain judgements: Eliciting experts' probabilities. *John Wiley & Sons*, 2010.

[143] Daniel Kahneman. Fast and slow thinking. *Allen Lane and Penguin Books, New York*, 2011.

[144] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55, 1983.

[145] James Berger. The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385 – 402, 2006.

[146] Tihomir Asparouhov and Bengt Muthén. Bayesian analysis of latent variable models using mplus, 2010.

[147] Bengt Muthén. Bayesian analysis in mplus: A brief introduction, 2010.

[148] John Geweke and Charles Whiteman. Bayesian forecasting. *Handbook of economic forecasting*, 1:3–80, 2006.

[149] Ronald D Lee and Shripad Tuljapurkar. Probabilistic prediction of population futures. *Demography*, 36(4):791–804, 1999.

[150] Andy Pole, Mike West, and Jeff Harrison. *Applied Bayesian forecasting and time series analysis*. CRC Press, 1994.

[151] Sumio Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594, 2010.

[152] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432, 2017.

[153] Jennifer A Hoeting, David Madigan, Adrian E Raftery, and Chris T Volinsky. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *Statistical science*, 14(4):382–417, 1999.

[154] Adrian E Raftery, David Madigan, and Jennifer A Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

[155] J Scott Armstrong. Standards and practices for forecasting. In *Principles of forecasting: A handbook for researchers and practitioners*, pages 679–732. Springer, 2001.

[156] Leonard J Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450, 2000.

[157] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

[158] Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[159] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, 2007.

[160] A Philip Dawid. Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 147(2):278–290, 1984.

[161] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

[162] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

[163] François Lafond, Aimee Gotway Bailey, Jan David Bakker, Dylan Rebois, Rubina Zadourian, Patrick McSharry, and J Doyne Farmer. How well do experience curves predict technological progress? a method for making distributional forecasts. *Technological Forecasting and Social Change*, 128:104–117, 2018.

[164] François Lafond, Diana Seave Greenwald, and J Doyne Farmer. Can stimulating demand drive costs down? world war ii as a natural experiment. *World War II as a Natural Experiment (June 1, 2020)*, 2020.

[165] Christian P Robert and George Casella. *Monte Carlo statistical methods*. Springer, 1999.

[166] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[167] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[168] Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

[169] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

[170] Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[171] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2, 2011.

[172] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[173] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

[174] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[175] Charlie Wilson, Arnulf Grubler, Nico Bauer, Volker Krey, and Keywan Riahi. Future capacity growth of energy technologies: are scenarios consistent with historical evidence? *Climatic Change*, 118(2):381–395, 2013.

[176] Simon Evans. How the uk transformed its electricity supply in just a decade, 2019.

[177] Adrian E Raftery, Alec Zimmer, Dargan MW Frierson, Richard Startz, and Peiran Liu. Less than 2 c warming by 2100 unlikely. *Nature climate change*, 7(9):637–641, 2017.

[178] D Hendry. Forecasting pitfalls. *Bulletin EU&US inflation and macroeconomic analysis (BIMA)*, 100, 2003.

[179] James D Hamilton. *Time series analysis*. Princeton University Press, 1994.

[180] Edward E Leamer. Specification searches: Ad hoc inference with nonexperimental data. *Wiley New York*, 53, 1978.

[181] Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, pages 733–760, 1996.

[182] Nassim Nicholas Taleb. Black swans and the domains of statistics. *The american statistician*, 61(3):198–200, 2007.

[183] Roger Fouquet. Historical energy transitions: Speed, prices and system transformation. *Energy research & social science*, 22:7–12, 2016.

[184] International Energy Agency. Global ev outlook 2021, 2021. Accessed on 2 October, 2025 at https://www.iea.org/reports/global-ev-outlook-2021,.

[185] Robert J Lempert, Steven W Popper, and Steven C Bankes. *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*. Rand Corporation, 2003.

[186] Frank W. Geels, Benjamin K. Sovacool, Tim Schwanen, and Steve Sorrell. The socio-technical dynamics of low-carbon transitions. *Joule*, 1(3):463–479, 2017.

[187] Peter Schwartz. *The art of the long view: Planning for the future in an uncertain world*. Currency Doubleday, 1991.

[188] John Weyant, Ogunlade Davidson, Hadi Dowlatabadi, Jae Edmonds, Michael Grubb, EA Parson, R Richels, Jan Rotmans, PR Shukla, Richard SJ Tol, et al. Integrated assessment of climate change: an overview and comparison of approaches and results. *Climate change*, 3, 1995.

[189] Shinichiro Fujimori, Tomoko Hasegawa, Toshihiko Masui, Kiyoshi Takahashi, Diego Silva Herran, Hancheng Dai, Yasuaki Hijioka, and Mikiko Kainuma. Ssp3: Aim implementation of shared socioeconomic pathways. *Global Environmental Change*, 42:268–283, 2017.

[190] Katherine Calvin, Pralit Patel, Leon Clarke, Ghassem Asrar, Ben Bond-Lamberty, Ryna Yiyun Cui, Alan Di Vittorio, Kalyn Dorheim, Jae Edmonds, Corinne Hartin, et al. Gcam v5. 1: representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, 12(2):677–698, 2019.

[191] Elke Stehfest, Detlef van Vuuren, Tom Kram, Lex Bouwman, Rob Alkemade, Michel Bakkenes, Hester Biemans, Alphons Bouwman, Michel Den Elzen, Jan Janse, et al. Integrated assessment of global environmental change with image 3.0: Model description and policy applications. Technical report, Netherlands Environmental Assessment Agency (PBL), 2014.

[192] Daniel Huppmann, Joeri Rogelj, Elmar Kriegler, Volker Krey, and Keywan Riahi. Iamc 1.5 c scenario explorer and data hosted by iiasa. *Integrated Assessment Modeling Consortium & International Institute for Applied Systems Analysis*, 2019.

[193] Gunnar Luderer, Marian Leimbach, Nico Bauer, Elmar Kriegler, Lavinia Baumstark, Christoph Bertram, Anastasis Giannousakis, Jerome Hilaire, David Klein, Antoine Levesque, et al. Description of the remind model (version 1.6). Technical report, Potsdam Institute for Climate Impact Research (PIK), 2015.

[194] Richard Loulou, Gary Goldstein, and Ken Noble. Documentation for the markal family of models. Technical report, Energy Technology Systems Analysis Programme, 2004.

[195] Johannes Emmerling, Laurent Drouet, Lara Aleluia Reis, Marinella Bevione, Laurent Berger, Valentina Bosetti, Samuel Carrara, Enrica De Cian, Gauthier D'Aertrycke,

Tom Longden, et al. The witch 2016 model-documentation and implementation of the shared socioeconomic pathways. *Fondazione Eni Enrico Mattei Working Papers*, 42, 2016.

[196] Brian C O'Neill, Elmar Kriegler, Keywan Riahi, Kristie L Ebi, Stephane Hallegatte, Timothy R Carter, Ritu Mathur, and Detlef P van Vuuren. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, 122(3):387–400, 2014.

[197] Keywan Riahi, Detlef P Van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C O'neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environmental Change*, 42:153–168, 2017.

[198] IPCC. Climate change 2022: Mitigation of climate change. https://www.ipcc.ch/report/ar6/wg3/, 2022. Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

[199] Frank W Geels. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research policy*, 31(8-9):1257–1274, 2002.

[200] Cesare Marchetti and Nebojsa Nakicenovic. *The dynamics of energy systems and the logistic substitution model*. RR-79-013, 1979.

[201] Department for Energy Security and Net Zero. Uk energy in brief 2025. Technical report, UK Government, 2025. Accessed: 2025-02-05.

[202] Benjamin K Sovacool. How long will it take? conceptualizing the temporal dynamics of energy transitions. *Energy Research & Social Science*, 13:202–215, 2016.

[203] Robert Gross, Richard Hanna, Ajay Gambhir, Philip Heptonstall, and Jamie Speirs. How long does innovation and commercialisation in the energy sectors take? historical case studies of the timescale from invention to widespread commercialisation in energy supply and end use technology. *Energy Policy*, 123:682–699, 2018.

[204] Brendon Tankwa, Lucas Vazquez Bassat, Peter Barbrook-Johnson, and J Doyne Farmer. Technological progress at national level: Increasing diffusion speeds with ever-changing leaders and followers. Working paper no. 2025-01, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, UK, 2025.

[205] Arnulf Grubler. Energy transitions research: Insights and cautionary tales. *Energy policy*, 50:8–16, 2012.

[206] Nebojsa Nakicenovic. Dynamics and replacement of u.s. transport infrastructures. In Jesse Ausubel and Robert Herman, editors, *Cities and Their Vital Systems: Infrastructure Past, Present, and Future*, pages 175–221. The National Academies Press, Washington, D.C., 1988.

[207] Aleh Cherp, Vadim Vinichenko, Jessica Jewell, Masahiro Suzuki, and Miklós Antal. Comparing electricity transitions: A historical analysis of nuclear, wind and solar power in germany and japan. *Energy Policy*, 101:612–628, 2016.

[208] Joeri Rogelj, Alexander Popp, Katherine V Calvin, Gunnar Luderer, Johannes Emmerling, David Gernaat, Shinichiro Fujimori, Jessica Strefler, Tomoko Hasegawa, Giacomo Marangoni, et al. Scenarios towards limiting global mean temperature increase below 1.5 c. *Nature climate change*, 8(4):325–332, 2018.

[209] Max Roser, Hannah Ritchie, and Edouard Mathieu. Technological change. Our

World in Data, 2023. Accessed February 2026.

[210] Hannah Ritchie and Pablo Rosado. Electricity mix. *Our World in Data*, 2020. Accessed at https://ourworldindata.org/electricity-mix.

[211] Charlie Wilson, Arnulf Grubler, Nuno Bento, Stephen Healey, Simon De Stercke, and Caroline Zimm. Granular technologies to accelerate decarbonization. *Science*, 368(6486):36–39, 2020.

[212] Alexandra Krumm, Diana Süsser, and Philipp Blechinger. Modelling social aspects of the energy transition: What is the current representation of social factors in energy models? *Energy*, 239:121706, 2022.

[213] Joseph DeCarolis, Hannah Daly, Paul Dodds, Ilkka Keppo, Francis Li, Will McDowall, Steve Pye, Neil Strachan, Evelina Trutnevyte, Will Usher, Matthew Winning, Sonia Yeh, and Marianne Zeyringer. Formalizing best practice for energy system optimization modelling. *Applied Energy*, 194:184–198, 2017.

[214] Febin Kachirayil, Jann Michael Weinand, Fabian Scheller, and Russell McKenna. Reviewing local and integrated energy system models: insights into flexibility and robustness challenges. *Applied Energy*, 324:119666, 2022.

[215] Alexander Pfeiffer, Richard Millar, Cameron Hepburn, and Eric Beinhocker. Committed emissions from existing and planned power plants and asset stranding required to meet the paris agreement. *Environmental Research Letters*, 13(5):054019, 2018.

[216] Charlie Wilson and Arnulf Grubler. Historical characteristics and scenario analysis of technological change in the energy system. In Rob Vos and Diana Alarcón, editors, *Technology and Innovation for Sustainable Development*, pages 45–80. Bloomsbury Academic, London, 2016.

[217] Alexander Q Gilbert and Benjamin K Sovacool. Looking the wrong way: Bias, renewable electricity, and energy modelling in the united states. *Energy*, 94:533–541, 2016.

[218] Diana Süsser, Andrzej Ceglarz, Hannes Gaschnig, Vassilis Stavrakas, Alexandros Flamos, George Giannakidis, and Johan Lilliestam. Model-based policymaking or policy-based modelling? how energy models and energy policy interact. *Energy Research & Social Science*, 75:101984, 2021.

[219] Benjamin K Sovacool, Mari Martiskainen, Andrew Hook, and Lucy Baker. Decarbonization and its discontents: a critical energy justice perspective on four low-carbon transitions. *Climatic Change*, 155:581–619, 2019.

[220] Theodore Paul Wright. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, 3(4):122–128, 1936.

[221] Gregory F Nemet. Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy*, 34(17):3218–3232, 2006.

[222] Ryan Wiser and Mark Bolinger. Land-based wind market report: 2021 edition. Technical report, Lawrence Berkeley National Laboratory, 2021.

[223] Micah S Ziegler and Jessika E Trancik. Re-examining rates of lithium-ion battery technology improvement and cost decline. *Energy & Environmental Science*, 14(4):1635–1651, 2021.

[224] Marta Victoria, Nancy Haegel, Ian Marius Peters, Ron Sinton, Arnulf Jäger-Waldau, Carlos Del Cañizo, Christian Breyer, Matthew Stocks, Andrew Blakers, Izumi Kaizuka, et al. Solar photovoltaics is ready to power a sustainable future. *Joule*, 5(5):1041–1056, 2021.

[225] Volker Krey, Fei Guo, Peter Kolp, Wenji Zhou, Roberto Schaeffer, Aayushi Awasthy, Christoph Bertram, Harmen-Sytze de Boer, Panagiotis Fragkos, Shinichiro Fujimori, et al. Looking under the hood: a comparison of techno-economic assumptions across national and global integrated assessment models. *Energy*, 172:1254–1267, 2019.

[226] Andrii Gritsevskyi and Nebojša Nakićenovi. Modeling uncertainty of induced technological change. *Energy policy*, 28(13):907–921, 2000.

[227] William D Nordhaus. Rolling the 'dice': an optimal transition path for controlling greenhouse gases. *Resource and Energy Economics*, 15(1):27–50, 1993.

[228] Varun Sivaram. *Taming the sun: Innovations to harness solar energy and power the planet*. MIT Press, 2018.

[229] Richard Loulou, Gary Goldstein, Amit Kanudia, Antti Lettila, and Uwe Remme. Documentation for the times model part i. Technical report, IEA Energy Technology Systems Analysis Programme (ETSAP), 2016.

[230] Michael Grubb. *Planetary economics: Energy, climate change and the three domains of sustainable development*. Routledge, 2014.

[231] Richard M Cyert and James G March. *A behavioral theory of the firm*. Prentice-Hall, 1963.

[232] Daniel Scamman, Baltazar Solano-Rodríguez, Steve Pye, Lai Fong Chiu, Andrew Z. P. Smith, Tiziano Gallo Cassarino, Mark Barrett, and Robert Lowe. Heat decarbonisation modelling approaches in the uk: An energy system architecture perspective. *Energies*, 13(8), 2020.

[233] Elmar Kriegler, Massimo Tavoni, Tino Aboumahboub, Gunnar Luderer, Katherine Calvin, Gauthier Demaere, Volker Krey, and Keywan Riahi. Making or breaking climate targets: The ampere study on staged accession scenarios for climate policy. *Technological Forecasting and Social Change*, 90:24–44, 2015.

[234] Celine Guivarch, Elmar Kriegler, Joana Portugal-Pereira, Valentina Bosetti, James Edmonds, Manfred Fischedick, Petr Havlik, Paulina Jaramillo, Volker Krey, Franck Lecocq, et al. Annex iii: Scenarios and modelling methods. *IPCC 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1842–1908, 2023.

[235] K. Riahi, R. Schaeffer, J. Arango, K. Calvin, C. Guivarch, T. Hasegawa, K. Jiang, E. Kriegler, R. Matthews, G.P. Peters, A. Rao, S. Robertson, A.M. Sebbit, J. Steinberger, M. Tavoni, and D.P. Van Vuuren. Mitigation pathways compatible with long-term goals. In P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, book section 3. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.

[236] Edward Byers, Volker Krey, Elmar Kriegler, Keywan Riahi, Roberto Schaeffer, Jarmo Kikstra, Robin Lamboll, Zebedee Nicholls, Marit Sandstad, Chris Smith, Kaj van der Wijst, Franck Lecocq, Joana Portugal-Pereira, Yamina Saheb, Anders Stromann, Harald Winkler, Cornelia Auer, Elina Brutschin, Claire Lepault, Eduardo Müller-Casseres, Matthew Gidden, Daniel Huppmann, Peter Kolp, Giacomo Marangoni, Michaela Werning, Katherine Calvin, Celine Guivarch, Tomoko Hasegawa, Glen Peters, Julia Steinberger, Massimo Tavoni, Detlef van Vuuren, Alaa Al -Khourdajie,

Piers Forster, Jared Lewis, Malte Meinshausen, Joeri Rogelj, Bjorn Samset, and Ragnhild Skeie. Ar6 scenarios database, apr 2022.

[237] Tsimafei Kazlou, Aleh Cherp, and Jessica Jewell. Feasible deployment of carbon capture and storage and the requirements of climate targets. *Nature Climate Change*, 14(10):1047–1055, 2024.

[238] Zeke Hausfather and Glen P Peters. Emissions–the 'business as usual'story is misleading. *Nature*, 577(7792):618–620, 2020.

[239] Emily Ho, David V. Budescu, Valentina Bosetti, Detlef P. van Vuuren, and Klaus Keller. Not all carbon dioxide emission scenarios are equally likely: a subjective expert assessment. *Climatic Change*, 155:545–561, 2019.

[240] Hamish Beath, Chris Smith, Jarmo S. Kikstra, Mark M. Dekker, Matthew J. Gidden, and Joeri Rogelj. A weighting framework to improve the use of emissions scenario ensembles of opportunity. *Nature Climate Change*, feb 2026. Published 24 February 2026.

[241] John E. T. Bistline. Weighting for net zero. *Nature Climate Change*, feb 2026. News & Views; Published 24 February 2026.

[242] Nik Zielonka, Xin Wen, and Evelina Trutnevyte. Probabilistic projections of granular energy technology diffusion at subnational level. *PNAS Nexus*, 2(10):pgad321, oct 2023.

[243] Tomás Mac Uidhir, Brian Ó Gallachóir, John Curtis, and Fionn Rogan. Achieving the unprecedented: Modelling diffusion pathways for ambitious climate policy targets. *Energy and Climate Change*, 3:100073, dec 2022.

[244] Vassilis Stavrakas and Alexandros Flamos. An agent-based model to simulate technology adoption quantifying behavioural uncertainty of consumers. *Applied Energy*, 255:113795, 2019.

[245] Mathijs Harmsen, Detlef P van Vuuren, Maarten van den Berg, Andries F Hof, Chris Hope, Volker Krey, Jean-Francois Lamarque, Adriana Marcucci, Drew T Shindell, and Nicholas Stern. Integrated assessment model diagnostics: key indicators and model evolution. *Environmental Research Letters*, 16(5):054069, 2021.

[246] Sergey Paltsev. Energy scenarios: the value and limits of scenario analysis. *Wiley Interdisciplinary Reviews: Energy and Environment*, 6(4):e242, 2017.

[247] Johan Rockström, Owen Gaffney, Joeri Rogelj, Malte Meinshausen, Nebojsa Nakicenovic, and Hans Joachim Schellnhuber. A roadmap for rapid decarbonization. *Science*, 355(6331):1269–1271, 2017.

[248] Paul Denholm, Douglas J Arent, Samuel F Baldwin, Daniel E Bilello, Gregory L Brinkman, Jaquelin M Cochran, Wesley J Cole, Bethany Frew, Vahan Gevorgian, Jennifer Heeter, et al. Timescales of energy storage needed for reducing renewable energy curtailment. *Renewable Energy*, 130:388–404, 2019.

[249] Zahra Janipour, Reinier de Nooij, Peter Scholten, Mark AJ Huijbregts, and Heleen de Coninck. What are sources of carbon lock-in in energy-intensive industry? a case study into dutch chemicals production. *Energy Research & Social Science*, 60:101320, 2019.

[250] IPCC. Climate change 2022: Mitigation of climate change. summary for policymakers. https://www.ipcc.ch/report/ar6/wg3/, 2022. Summary for Policymakers, Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

[251] Gabriele Realmonte, Laurent Drouet, Ajay Gambhir, James Glynn, Adam Hawkes, Alexandre C Köberle, and Massimo Tavoni. An inter-model assessment of the role of direct air capture in deep mitigation pathways. *Nature Communications*, 10(1):1–12, 2019.

[252] Pete Smith, Steven J Davis, Felix Creutzig, Sabine Fuss, Jan Minx, Benoit Gabrielle, Etsushi Kato, Robert B Jackson, Annette Cowie, Elmar Kriegler, et al. Biophysical and economic limits to negative co2 emissions. *Nature Climate Change*, 6(1):42–50, 2016.

[253] Duncan McLaren and Andy Jarvis. Framing the carbon budget for 1.5 c: implications for policy and planning. *Wiley Interdisciplinary Reviews: Climate Change*, 7(6):834–851, 2016.

[254] Sabine Fuss, William F Lamb, Max W Callaghan, Jérôme Hilaire, Felix Creutzig, Thorben Amann, Tim Beringer, Wagner de Oliveira Garcia, Jens Hartmann, Tarun Khanna, et al. Betting on negative emissions. *Nature Climate Change*, 8(11):850–853, 2018.

[255] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287, 2002.

[256] J. Palmer, G. Sorda, and R. Madlener. Modeling the diffusion of residential photovoltaic systems in italy: An agent-based simulation. *Technological Forecasting and Social Change*, 99:106–131, 2015.

[257] Ardak Akhatova, Lukas Kranzl, Fabian Schipfer, and Charitha Buddhika Heendeniya. Agent-based modelling of urban district energy system decarbonisation —a systematic literature review. *Energies*, 15(2), 2022.

[258] Francis GN Li, Evelina Trutnevyte, and Neil Strachan. A review of socio-technical energy transition (stet) models. *Technological Forecasting and Social Change*, 105:26–44, 2016.

[259] J-F Mercure, Pablo Salas, Pim Vercoulen, Gregor Semieniuk, Aileen Lam, Hector Pollitt, Philip B Holden, Negar Vakilifard, Unnada Chewpreecha, Neil R Edwards, et al. Reframing incentives for climate policy action. *Nature Energy*, 6(12):1133–1143, 2021.

[260] Femke JMM Nijsse, Jean-Francois Mercure, Nadia Ameli, Francesca Larosa, Sumit Kothari, Jamie Rickman, Pim Vercoulen, and Hector Pollitt. The momentum of the solar energy transition. *Nature Communications*, 14(1):6542, 2023.

[261] Benjamin Wagenvoort. Why s-curves underestimate the future: Systematic bias in diffusion forecasting. Manuscript in preparation, 2026.

[262] International Energy Agency. Global ev outlook 2023, 2023. Accessed on 2 October, 2025 at https://www.iea.org/reports/global-ev-outlook-2023.

[263] Andrew Harvey and Paul Kattuman. Time series models based on growth curves with applications to forecasting coronavirus. *Harvard Data Science Review*, 1(S1), 2020.

[264] Hannah Ritchie. Tracking global data on electric vehicles. *Our World in Data*, 2024. https://ourworldindata.org/electric-car-sales.

[265] International Energy Agency. Global ev outlook 2025, 2025. Accessed on 2 October, 2025 at https://www.iea.org/reports/global-ev-outlook-2025.

[266] Patrick Hertzke, Nicolai Müller, Stephanie Schenk, and Ting Wu. The global electric-

vehicle market is amped up and on the rise. *McKinsey Cent. Futur. Mobil*, 1:1–8, 2018.

[267] Anna Murphy, Aileen Lam, Jean-Francois Mercure, and Simon Sharpe. Press release: Electric vehicles to pass two thirds of global car sales by 2030, 2025. Accessed on 2 October 2025, at https://eeist.co.uk/press-release-electric-vehicles-to-pass-two-thirds-of-global-car-sales-by-2030/.

[268] Roy Amara. What we have learned about forecasting and planning. *Futures*, 20(4):385–401, 1988.

[269] Charles P. Winsor. A comparison of certain symmetrical growth curves. *Journal of the Washington Academy of Sciences*, 22(4):73–84, 1932. Publisher: Washington Academy of Sciences.

[270] PM Shankar. Tutorial overview of simple, stratified, and parametric bootstrapping. *Engineering Reports*, 2(1):e12096, 2020.

[271] IEA. Solar pv power capacity in the net zero scenario, 2015-2030, 2023. Licence: CC BY 4.0.

[272] Minal Chandra, Pablo Busch, Francisco Parés Olguín, and Gil Tal. Paths of progress: Forecasting global electric vehicle demand amidst demographic and economic growth. *Transportation Research Part D: Transport and Environment*, 147:104928, 2025.

[273] Oluwadayomi Akinsooto, Enoch Oluwadunminu Ogunnowo, and Chukwuemeka Chukwuka Ezeanochie. The evolution of electric vehicles: A review of usa and global trends, 2025.

[274] Pablo Busch, Francisco Pares, Minal Chandra, Alissa Kendall, and Gil Tal. Future of global electric vehicle supply chain: Exploring the impact of global trade on electric vehicle production and battery requirements. *Transportation Research Record*, 2678(11):1468–1482, 2024.

[275] IRENA. Total solar capacity, 2025. Retrieved September 29, 2025 from https://archive.ourworldindata.org/20250909-093708/grapher/installed-solar-pv-capacity.html (archived on September 9, 2025).

[276] Gregory Nemet, Jenna Greene, Finn Müller-Hansen, and Jan C Minx. Dataset on the adoption of historical technologies informs the scale-up of emerging carbon dioxide removal measures. *Communications Earth & Environment*, 4(1):397, 2023. data retrieved from: https://www.nature.com/articles/s43247-023-01056-1.

[277] IRENA. Total wind capacity, 2025. Retrieved September 29, 2025 from https://archive.ourworldindata.org/20250909-093708/grapher/cumulative-installed-wind-energy-capacity-gigawatts.html (archived on September 9, 2025).

[278] IRENA. Total geothermal capacity, 2025. Retrieved September 19, 2026 from https://archive.ourworldindata.org/20260119-065148/grapher/installed-geothermal-capacity.html (archived on January 19, 2026).

[279] IRENA. Bioenergy and biofuels, 2025. https://www.irena.org/Energy-Transition/Technology/Bioenergy-and-biofuels (accessed on January 27, 2026).

[280] International Hydropower Association. 2025 world hydropower outlook, 2025. https://www.hydropower.org/publications/2025-world-hydropower-outlook (accessed on September 29, 2025).

[281] International Energy Agency. Copp28 tripling renewable capacity pledge, 2024. https://www.iea.org/reports/cop28-tripling-renewable-capacity-pledge (accessed on September 29, 2025).

[282] Philip Bradonjic, Nikolaus Franke, and Christian Lüthje. Decision-makers' underestimation of user innovation. *Research Policy*, 48(6):1354–1361, 2019.

[283] Gregory C Unruh. Understanding carbon lock-in. *Energy policy*, 28(12):817–830, 2000.

[284] Benjamin Wagenvoort, Joel C. Dyer, François Lafond, and J Doyne Farmer. Universality and predictability of technology diffusion. Manuscript in preparation, 2026.

[285] Edwin Mansfield. Technical change and the rate of imitation. *Econometrica*, 29(4):741–766, 1961.

[286] Nik Zielonka, Jenna Greene, Gregory Nemet, and Evelina Trutnevyte. Growth projections of hydrogen electrolysis worldwide with evidence from historical analogs and hindcasting, 2026.

[287] Nebojsa Nakicenovic and Arnulf Grübler. *Diffusion of technologies and social behavior*. Springer Science & Business Media, 2013.

[288] Richard M Swanson. A vision for crystalline silicon photovoltaics. *Progress in photovoltaics: Research and Applications*, 14(5):443–453, 2006.

[289] Joern Hoppmann, Joern Huenteler, and Bastien Girod. Compulsive policy-making—the evolution of the german feed-in tariff system for solar photovoltaic power. *Research policy*, 43(8):1422–1441, 2014.

[290] Christophe Ballif, Franz-Josef Haug, Mathieu Boccard, Pierre J Verlinden, and Giso Hahn. Status and perspectives of crystalline silicon photovoltaics in research and industry. *Nature Reviews Materials*, 7(8):597–616, 2022.

[291] Can Cui, Katherine Emma Lonergan, and Giovanni Sansavini. Policy-driven transformation of global solar pv supply chains and resulting impacts. *Nature Communications*, 16(1):6742, 2025.

[292] Diego Comin and Bart Hobijn. An exploration of technology diffusion. *American economic review*, 100(5):2031–2059, 2010.

[293] James D Murray. *Mathematical biology: I. An introduction*, volume 17. Springer Science & Business Media, 2007.

[294] Giacomo Masali, Maria Elena Bruni, and Guido Perboli. The statistical mechanics of innovation diffusion. *Annals of Operations Research*, 2025. Advance online publication.

[295] Jenna H Greene, Matthew J Gidden, Elina Brutschin, and Gregory F Nemet. Drivers of technology diffusion speed in countries. *Nature Communications*, 17, 2026.

[296] International Energy Agency. *Net Zero Roadmap: A Global Pathway to Keep the 1.5 °C Goal in Reach*. OECD Publishing, 2023.

[297] International Energy Agency. World energy outlook 2025. Technical report, IEA, Paris, 2025. Licence: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A).

[298] IRENA. Total solar capacity, 2025. processed by Our World in Data. Retrieved July 31, 2025 from https://archive.ourworldindata.org/20250721-095612/grapher/installed-solar-pv-capacity.html (archived on July 21, 2025).

[299] Diego A Comin and Bart Hobijn. The chat dataset. Working Paper 15319, National Bureau of Economic Research, September 2009. data retrieved from: https://data.nber.org/data-appendix/w15319/.

[300] Hannah Ritchie, Pablo Rosado, and Max Roser. Energy production and consumption. *Our World in Data*, 2020. https://ourworldindata.org/energy-production-

consumption.

[301] The World Bank. Mobile cellular subscriptions (per 100 people), 2023. data retrieved from World Development Indicators, https://data.worldbank.org/indicator/IT.CEL.SETS.P2.

[302] IHS Markit Ltd, 2023. Fleet register provider, acquired by S&P Global Inc. Accessed at https://ihsmarkit.com/.

[303] RehtinkX. Rethinkx technology production data, 2024.

[304] Benjamin K. Sovacool. How long will it take? conceptualizing the temporal dynamics of energy transitions. *Energy Research & Social Science*, 13:202–215, 2016.

[305] Charlie Wilson. *Meta-analysis of unit and industry level scaling dynamics in energy technologies and climate change mitigation scenarios*. JSTOR, 2009.

[306] K Usha Rao and VVN Kishore. A review of technology diffusion models with special reference to renewable energy technologies. *Renewable and sustainable energy reviews*, 14(3):1070–1078, 2010.

[307] Vijay Mahajan, Eitan Muller, and Frank M Bass. New-product diffusion models. *Handbooks in operations research and management science*, 5:349–408, 1993.

[308] Rui Baptista. Do innovations diffuse faster within geographical clusters? *International Journal of industrial organization*, 18(3):515–535, 2000.

[309] Bruce Robinson and Chet Lakhani. Dynamic price models for new-product planning. *Management science*, 21(10):1113–1122, 1975.

[310] H Peter Boswijk and Philip Hans Franses. On the econometrics of the bass diffusion model. *Journal of Business & Economic Statistics*, 23(3):255–268, 2005.

[311] Aslan Lotfi, Ali Lotfi, and William E Halal. Forecasting technology diffusion: a new generalisation of the logistic model. *Technology Analysis & Strategic Management*, 26(8):943–957, 2014.

[312] Kenneth P Burnham, David R Anderson, Kenneth P Burnham, and David R Anderson. *Practical use of the information-theoretic approach*. Springer, 1998.

[313] Leigh J Halliwell. The log-gamma distribution and non-normal error. *Variance*, 13(2):173–189, 2021.

[314] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335, 1995.

[315] Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2796–2804. PMLR, 2018.

[316] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[317] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python.

Nature methods, 17(3):261–272, 2020.

[318] Wes McKinney et al. Data structures for statistical computing in python. *scipy*, 445(1):51–56, 2010.

[319] Skipper Seabold, Josef Perktold, et al. Statsmodels: econometric and statistical modeling with python. *SciPy*, 7(1):92–96, 2010.

[320] Mylene Bedard. Optimal acceptance rates for metropolis algorithms: Moving beyond 0.234. *Stochastic Processes and their Applications*, 118(12):2198–2222, 2008.

[321] Aidan Li, Liyan Wang, Tianye Dou, and Jeffrey S Rosenthal. Exploring the generalizability of the optimal 0.234 acceptance rate in random-walk metropolis and parallel tempering algorithms. *Communications in Statistics-Simulation and Computation*, pages 1–31, 2025.

[322] Sean Talts, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*, 2018.

[323] William A Link and Mitchell J Eaton. On thinning of chains in mcmc. *Methods in ecology and evolution*, 3(1):112–115, 2012.

[324] Francis X. Diebold, Todd A. Gunther, and Anthony S. Tay. Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883, 1998.

[325] Murray Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472, 1952.

[326] Herbert A. David and Haikady N. Nagaraja. *Order Statistics*. Wiley, Hoboken, NJ, 3 edition, 2003.

[327] Charles J Clopper and Egon S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

[328] Christian P Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2nd edition, 2007.

[329] Timothy R Konold, Elizabeth A Sanders, and Kelvin Afolabi. From seinfeld to statistics: On the dangers of double dipping with bayesian inference. *Structural Equation Modeling: A Multidisciplinary Journal*, pages 1–10, 2025.

[330] Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[331] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.

[332] Hannah Ritchie and Max Roser. Technology adoption. *Our World in Data*, 2017. https://ourworldindata.org/technology-adoption.

[333] International Energy Agency. Energy statistics data browser, 2023. Provides international statistics on electricity production, consumption, and transmission & distribution (T&D) losses. Global T&D losses were approximately 6% in 2023, with historical values typically ranging 6–9% over the past decade (World Bank Indicator EG.ELC.LOSS.ZS, sourced from IEA).

[334] Leopold Janauschek. *Originum cisterciensium*, volume 1. Alfred Hoelder, 1877.

[335] United Nations Statistics Division and Organisation for Economic Co-operation

and Development. International standard industrial classification of all economic activities (isic), revision 4, 2008. Used for Industry, Infrastructure, and Transport sectors classification.

[336] International Energy Agency. World energy statistics, 2023. Used for Energy sector classification.

[337] World Health Organization. World health organization classifications, 2023. Used for Health, Waste Management, and Water sectors classification.

[338] Food and Agriculture Organization. Food and agriculture organization classifications, 2023. Used for Agriculture sector classification.

[339] International Telecommunication Union. International telecommunication union standards, 2023. Used for ICT sector classification.

[340] United States Geological Survey. Usgs mineral commodity summaries, 2023. Used for mining-related technologies consolidated into Industry sector classification.

[341] Vito MR Muggeo. Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071, 2003.

[342] International Energy Agency. Tracking clean energy progress 2023, 2025.

[343] IRENA. Sodium-ion batteries: A technology brief, 2025.

[344] International Energy Agency. Batteries and secure energy transitions, 2024.

[345] Charlotte Muller. The aluminum monopoly and the war. *Political Science Quarterly*, 60(1):14–43, 1945.

[346] Sam H. Patterson. *Bauxite reserves and potential aluminum resources of the world*. Number 1228 in Geological Survey Bulletin. U.S. Government Printing Office, Washington, D.C., 1967.

[347] O Yul Kwon. The korean financial crisis: diagnosis, remedies and prospects. *Journal of the Asia Pacific Economy*, 3(3):331–357, 1998.

[348] SY Kim. South korea's policy responses to global economic crisis. In *Proceedings of the First Middle East Conference on Global Business, Economics, Finance and Banking*, pages 1–27, 2014.

[349] International Monetary Fund. The IMF's response to the asian crisis, 1997.

[350] Seung-kyung Kim and John Finch. Living with rhetoric, living against rhetoric: Korean families and the imf economic crisis. *Korean Studies*, pages 120–139, 2002.

[351] Colin Weiss. De-dollarization? diversification? exploring central bank gold purchases and the dollar's role in international reserves. International Finance Discussion Papers 1420, Board of Governors of the Federal Reserve System, Washington, 2025.

[352] Christine T. Reilly. Food fight: Eating and identity in japan during the second world war. *International Journal of Arts and Sciences*, 3(13):410–431, 2010. Documents severe sugar scarcity during WWII in Japan, with Americans blocking shipping from Taiwan causing per capita consumption to decline sharply by 1945.

[353] L. Halliday Piel. Food rationing and children's self-reliance in japan, 1942–1952. *The Journal of the History of Childhood and Youth*, 5(3):393–418, 2012. Examines food rationing in Japan during and after WWII, including severe restrictions on sugar from 1942-1952.

[354] Bruce Foster Johnston. *Food and agriculture in Japan, 1880-1950*. Stanford University, 1953.

[355] Barak Kushner. Sweetness and empire: Sugar consumption in imperial japan. In Penelope Francks and Janet Hunter, editors, *The Historical Consumer: Consumption and Everyday Life in Japan, 1850-2000*, pages 127–151. Palgrave Macmillan, London, 2012. Discusses sugar rationing and consumption patterns in Japan during the 1930s-1940s wartime period.

[356] Yasukichi Yasuba. Japan's post-war growth in historical perspective. *Japan Forum*, 3(1):57–70, 1991.

[357] Muhammad Fauzi, Indri Kartiko-Sari, and Hemant Poudyal. Trends of dietary intakes and metabolic diseases in japanese adults: assessment of national health promotion policy and national health and nutrition survey 1995–2019. *Journal of Clinical Medicine*, 11(9):2350, 2022.

[358] Helgi Library. Sugar consumption per capita in japan. https://www.helgilibrary.com/indicators/sugar-consumption-per-capita/japan/, 2023. Data sourced from FAOSTAT. Accessed December 2025.

[359] Yasuhiro Matsumura. Nutrition trends in japan. *Asia Pacific journal of clinical nutrition*, 10:S40–S47, 2001.

[360] Kevin H. R. Rouwenhorst, Anthony S. Travis, and Leon Lefferts. 1921–2021: A century of renewable ammonia synthesis. *Sustainable Chemistry*, 3(2):149–171, 2022.

[361] James Bessen. *Learning by Doing: The Real Connection between Innovation, Wages, and Wealth*. Yale University Press, New Haven, 2015.

[362] Steven Klepper. Industry life cycles. *Industrial and Corporate Change*, 6(1):145–182, 1997.

[363] David S. Landes. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press, Cambridge, 1969.

[364] Gisela Rua. Diffusion of containerization. Finance and Economics Discussion Series 2014-88, Board of Governors of the Federal Reserve System, Washington, DC, 2014.

[365] Urs Schenker, Martin Scheringer, and Konrad Hungerbühler. Investigating the global fate of DDT: Model evaluation and estimation of future trends. *Environmental Science & Technology*, 42(4):1178–1184, 2008.

[366] National Research Council. Fostering research on the economic and social impacts of information technology. Technical report, National Academies Press, Washington, DC, 1998.

[367] Bryce Ryan and Neal C. Gross. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8(1):15–24, 1943.

[368] Harold A. Linstone and Devendra Sahal. *Technological Substitution: Forecasting Techniques and Applications*. Elsevier, New York, 1976.

[369] David A. Hounshell. *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States*. Johns Hopkins University Press, Baltimore, 1984.

[370] Suzi Kerr and Richard G. Newell. Policy-induced technology adoption: Evidence from the U.S. lead phasedown. *Journal of Industrial Economics*, 51(3):317–343, 2003.

[371] Gregory F. Nemet. Interim monitoring of cost dynamics for publicly supported energy technologies. *Energy Policy*, 37(3):825–835, 2009. Obtained via the Performance Curve Database.

[372] Didier Sornette. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Springer, 2004.

[373] Andrew Gelman, Daniel Simpson, and Michael Betancourt. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017.

[374] Stefan Van Dongen. Prior specification in bayesian statistics: three cautionary tales. *Journal of theoretical biology*, 242(1):90–100, 2006.

[375] Hannah Ritchie and Pablo Rosado. Energy mix. *Our World in Data*, 2020. Accessed at https://ourworldindata.org/energy-mix.

[376] Cesare Marchetti. Primary energy substitution models: on the interaction between energy and society. *Technological forecasting and social change*, 10(4):345–356, 1977.

[377] International Energy Agency. Energy efficiency 2021. Technical report, IEA, Paris, 2021.

[378] William Beaver. The failed promise of nuclear power. *The Independent Review*, 15(3):399–411, 2011.

[379] Richard Pfau. *No Sacrifice Too Great: The Life of Lewis Strauss*. Charlottesville: University of Virginia Press, 1985.

[380] Andrew Hopkins. Was three mile island a 'normal accident'? *Journal of contingencies and crisis management*, 9(2):65–72, 2001.

[381] Henry W Kendall. The failure of nuclear power. In *Risk, Organizations, and Society*, pages 163–218. Springer, 1991.

[382] Anne-Sophie Hacquin, Sacha Altay, Lene Aarøe, and Hugo Mercier. Disgust sensitivity and public opinion on nuclear energy. *Journal of Environmental Psychology*, 80:101749, 2022.

[383] Wolfgang Krohn and Peter Weingart. Commentary: Nuclear power as a social experiment—european political "fall out" from the chernobyl meltdown. *Science, Technology, & Human Values*, 12(2):52–58, 1987.

[384] Francois Rozon, Craig McGregor, and Michael Owen. Long-term forecasting framework for renewable energy technologies' installed capacity and costs for 2050. *Energies*, 16(19):6874, 2023.

[385] Alessandro Martulli, Kristin Brandt, Florian Allroggen, and Robert Malina. The potential scale-up of sustainable aviation fuels production capacity to meet global and eu policy targets. *Nature Communications*, 16(1):11619, 2025.

[386] IRENA. Renewable power generation costs in 2023. Technical report, International Renewable Energy Agency, Abu Dhabi, 2024.

[387] Kristina Sadovskaia, Dmitrii Bogdanov, Samuli Honkapuro, and Christian Breyer. Power transmission and distribution losses–a model based on available empirical data and future trends for all countries globally. *International Journal of Electrical Power & Energy Systems*, 107:98–109, 2019.

[388] Hannah Ritchie. Primary, secondary, final, and useful energy: Why are there different ways of measuring energy? *Our World in Data*, 2022. https://archive.ourworldindata.org/20251125-173858/energy-definitions.html.

[389] George Casella and Roger Berger. *Statistical inference*. Chapman and Hall/CRC, 2024.

[390] Lukas Wagner, Jiajia Suo, Bowen Yang, Dmitry Bogachuk, Estelle Gervais, Robert Pietzcker, Andrea Gassmann, and Jan Christoph Goldschmidt. The resource demands

of multi-terawatt-scale perovskite tandem photovoltaics. *Joule*, 8(4):1142–1160, 2024.

[391] Keiichi Komoto, Masakuza Ito, Peter van der Vleuten, David Faiman, and Kosuke Kurokawa. *Energy from the desert: very large scale photovoltaic systems: socio-economic, financial, technical and environmental aspects*. Routledge, 2007.

[392] IEA. Electricity mid-year update 2025. Technical report, International Energy Agency, Paris, 2025. Licence: CC BY 4.0.

[393] Peter E Glaser. Power from the sun: its future. *Science*, 162(3856):857–861, 1968.

[394] Robert L Olsen et al. Space solar power to meet future energy needs. In *Space Solar Power Review*, volume 1, pages 49–72, 1979.

[395] Benjamin Wagenvoort and Matthew Ives. Empirical solar and wind forecasts expose a solar-growth shortfall in AR6 mitigation scenarios. Manuscript in preparation, 2026.

[396] Jill Jäger and Tim O'Riordan. The history of climate change science and politics. In *The Politics of Climate Change*, pages 1–31. Routledge, 2019.

[397] Mike Hulme and Martin Mahony. Climate change: What do we know about the ipcc? *Progress in Physical Geography*, 34(5):705–718, 2010.

[398] R Scott. The history of the international energy agency: the first twenty years. volume 1: Origins and structures of the iea, Jan 1994.

[399] Marc Jaxa-Rozen and Evelina Trutnevyte. Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. *Nature Climate Change*, 11(3):266–273, 2021.

[400] Mengzhu Xiao, Tobias Junne, Jannik Haas, and Martin Klein. Plummeting costs of renewables-are energy scenarios lagging? *Energy Strategy Reviews*, 35:100636, 2021.

[401] Nicholas Stern. Economics: Current climate models are grossly misleading. *Nature*, 530(7591):407–409, 2016.

[402] J Doyne Farmer, Cameron Hepburn, Penny Mealy, and Alexander Teytelboym. A third wave in the economics of climate change. *Environmental and Resource Economics*, 62:329–357, 2015.

[403] Hiroto Shiraki and Masahiro Sugiyama. Back to the basic: toward improvement of technoeconomic representation in integrated assessment models. *Climatic Change*, 162(1):13–24, 2020.

[404] Keigo Akimoto, Miyuki Nagashima, Fuminori Sano, and Teruhisa Ando. Gaps between costs and potentials estimated by bottom-up assessments versus integrated assessment models. *Energy Strategy Reviews*, 55:101521, 2024.

[405] Yijing Wang, Rong Wang, Katsumasa Tanaka, Philippe Ciais, Josep Penuelas, Yves Balkanski, Jordi Sardans, Didier Hauglustaine, Junji Cao, Jianmin Chen, Lin Wang, Xu Tang, and Renhe Zhang. Global spatiotemporal optimization of photovoltaic and wind power to achieve the Paris Agreement targets. *Nature Communications*, 16(1):2127, 2025.

[406] Cameron Hepburn, Matthew C Ives, Sam Loni, Penny Mealy, Pete Barbrook-Johnson, J Doyne Farmer, Nicholas Stern, and Joseph Stiglitz. Economic models and frameworks to guide climate policy. *Oxford Review of Economic Policy*, page graf020, 2025.

[407] Robert C. Pietzcker, Falko Ueckerdt, Samuel Carrara, Harmen Sytze de Boer, Jacques Després, Shinichiro Fujimori, Nils Johnson, Alban Kitous, Yvonne Scholz, Patrick

Sullivan, and Gunnar Luderer. System integration of wind and solar power in integrated assessment models: A cross-model evaluation of new approaches. *Energy Economics*, 64:583–599, may 2017.

[408] Nathan P Gillett. Halving of the uncertainty in projected warming over the past decade. *npj Climate and Atmospheric Science*, 7(1):146, 2024.

[409] Zeke Hausfather and Glen P Peters. Heading in here running to two lines. *Nature*, 577:618–620, 2020.

[410] C Lennart Baumgärtner and J Doyne Farmer. Will national renewable costs continue declining? *Joule*, 2025.

[411] Christian Breyer, Dmitrii Bogdanov, Ashish Gulagi, Arman Aghahosseini, Larissa SNS Barbosa, Otto Koskinen, Maulidi Barasa, Upeksha Caldera, Svetlana Afanasyeva, Michael Child, et al. On the role of solar photovoltaics in global energy transition scenarios. *Progress in photovoltaics: research and applications*, 25(8):727–745, 2017.

[412] Drew Shindell and Joeri Rogelj. Preserving carbon dioxide removal to serve critical needs. *Nature Climate Change*, pages 1–6, 2025.

[413] Peiran R Liu and Adrian E Raftery. Country-based rate of emissions reductions should increase by 80% beyond nationally determined contributions to meet the 2 c target. *Communications earth & environment*, 2(1):29, 2021.

[414] Hannah Ritchie, Pablo Rosado, and Max Roser. Co2 and greenhouse gas emissions. *Our World in Data*, 2023. https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

[415] Brian C O'Neill, Elmar Kriegler, Kristie L Ebi, Eric Kemp-Benedict, Keywan Riahi, Dale S Rothman, Bas J van Ruijven, Detlef P van Vuuren, Joern Birkmann, Kasper Kok, Marc Levy, and William Solecki. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42:169–180, 2017.

[416] Keywan Riahi, Christoph Bertram, Daniel Huppmann, Joeri Rogelj, Valentina Bosetti, Anique-Marie Cabardos, Andre Deppermann, Laurent Drouet, Stefan Frank, Oliver Fricko, et al. Cost and attainability of meeting stringent climate targets without overshoot. *Nature Climate Change*, 11(12):1063–1069, 2021.

[417] Network for Greening the Financial System (NGFS). Ngfs climate scenarios for central banks and supervisors. Technical report, Network for Greening the Financial System, June 2021. Phase II (second iteration).

[418] Gunnar Luderer, Silvia Madeddu, Leon Merfort, Falko Ueckerdt, Michaja Pehl, Robert Pietzcker, Marianna Rottoli, Felix Schreyer, Nico Bauer, Lavinia Baumstark, Christoph Bertram, Alois Dirnaichner, Florian Humpenöder, Antoine Levesque, Alexander Popp, Renato Rodrigues, Jessica Strefler, and Elmar Kriegler. Impact of declining renewable energy costs on electrification in low-emission scenarios. *Nature Energy*, 7(1):32–42, 2022.

[419] Heleen L van Soest, Lara Aleluia Reis, Luiz Bernardo Baptista, Christoph Bertram, Jacques Després, Laurent Drouet, Michel den Elzen, Panagiotis Fragkos, Oliver Fricko, Shinichiro Fujimori, et al. Global roll-out of comprehensive policy measures may aid in bridging emissions gap. *Nature communications*, 12(1):6419, 2021.

[420] Bjoern Soergel, Elmar Kriegler, Isabelle Weindl, Sebastian Rauner, Alois Dirnaichner, Constantin Ruhe, Matthias Hofmann, Nico Bauer, Christoph Bertram, Benjamin Leon Bodirsky, et al. A sustainable development pathway for climate action

within the un 2030 agenda. *Nature Climate Change*, 11(8):656–664, 2021.

[421] IPCC. *Climate Change 2022 - Mitigation of Climate Change - Full Report*. Cambridge University Press, Cambridge, U.K., 2022.

[422] Glen P Peters, Alaa Al Khourdajie, Ida Sognnaes, and Benjamin M Sanderson. AR6 scenarios database: an assessment of current practices and future recommendations. *npj Climate Action*, 2(1):31, 2023.

[423] Richard Loulou and Maryse Labriet. ETSAP-TIAM: The TIMES integrated assessment model Part I: Model structure. *Computational Management Science*, 5(1-2):7–40, 2008.

[424] Hadi Vatankhah Ghadim, Jannik Haas, Christian Breyer, Hans Christian, E Grant Read, Mengzhu Xiao, and Rebecca Peer. Are we too pessimistic ? Cost projections for solar photovoltaics , wind power , and batteries are over-estimating actual costs globally. *Applied Energy*, 390(March):125856, 2025.

[425] Ember. Electricity data explorer. https://ember-energy.org/data/electricity-data-explorer/?metric=pct_share, 2025. Accessed: December 16, 2025.

[426] Stuart Jenkins, Eli Mitchell-Larson, Matthew C. Ives, Stuart Haszeldine, and Myles Allen. Upstream decarbonization through a carbon takeback obligation: An affordable backstop climate policy. *Joule*, 5(11):2777–2796, nov 2021.

[427] Falko Ueckerdt, Robert Pietzcker, Yvonne Scholz, Daniel Stetter, Anastasis Giannousakis, and Gunnar Luderer. Decarbonizing global power supply under region-specific consideration of challenges and options of integrating variable renewables in the REMIND model. *Energy Economics*, 64:665–684, 2017.

[428] Johannes Emmerling, Laurent Drouet, Kaj-Ivar van der Wijst, Detlef van Vuuren, Valentina Bosetti, and Massimo Tavoni. The role of the discount rate for emission pathways and negative emissions. *Environmental Research Letters*, 14(10):104008, oct 2019.

[429] Alexandre C Köberle. The Value of BECCS in IAMs: a Review. *Current Sustainable/Renewable Energy Reports*, 6(4):107–115, 2019.

[430] Sabine Fuss, William F Lamb, Max W Callaghan, Jérôme Hilaire, Felix Creutzig, Thorben Amann, Tim Beringer, Wagner de Oliveira Garcia, Jens Hartmann, Tarun Khanna, et al. Negative emissions—part 2: Costs, potentials and side effects. *Environmental research letters*, 13(6):063002, 2018.

[431] Neil Grant, Ajay Gambhir, Shivika Mittal, Chris Greig, and Alexandre C Köberle. Enhancing the realism of decarbonisation scenarios with practicable regional constraints on co2 storage capacity. *International Journal of Greenhouse Gas Control*, 120:103766, 2022.

[432] IAMC (Integrated Assessment Modeling Consortium). Aim-cge, 2020.

[433] S Messner and M Strubegger. User's guide for message iii, 1995.

[434] Samuel Carrara, Michela Bevione, Harmen Sytze de Boer, David Gernaat, Silvana Mima, Robert C Pietzcker, and Massimo Tavoni. Exploring pathways of solar pv learning-by-doing in integrated assessment models. In *3rd AIEE Energy Symposium on Current and future challenges to energy security-the energy transition*, pages 1–18, 2018.

[435] Potsdam Institute for Climate Impact Research (PIK). Remind model documentation. https://www.iamcdocumentation.eu/index.php/Model_Documentation_-_REMIND, 2025. Accessed: 2026-01-13.

[436] NGFS. Ngfs climate scenarios technical documentation, 2023.

[437] Andrea Bacilieri, Richard Black, and Rupert Way. Assessing the relative costs of high-CCS and low-CCS pathways to 1.5 degrees. Working Paper 23-04, Smith School of Enterprise and the Environment, University of Oxford, 2023.

[438] Sylwia Słupik. Deliverable 2.2. *University of Economics in Katowice*, 2018.

[439] Everett M Rogers, Arvind Singhal, and Margaret M Quinlan. Diffusion of innovations. In *An integrated approach to communication theory and research*, pages 432–448. Routledge, 2014.

[440] Charlie Wilson and Arnulf Grubler. Historical characteristics and scenario analysis of technological change in the energy system. *Technology and innovation for sustainable development*, 45, 2015.

[441] Arnulf Grübler. *Technology and global change*. Cambridge university press, 2003.

[442] IPCC. *Annex III: Scenarios and modelling methods*. Cambridge University Press, Cambridge, U.K., 2022.

[443] Anna Pirani, Jan S Fuglestvedt, Edward Byers, Brian O Neill, Keywan Riahi, June-yi Lee, Steven K Rose, Roberto Schaeffer, and Claudia Tebaldi. Scenarios in IPCC assessments: lessons from AR6 and opportunities for AR7. *npj Climate Action*, 2024.

[444] Wilson Ricks and Jesse D Jenkins. Pathways to national-scale adoption of enhanced geothermal power through experience-driven cost reductions. *Joule*, 9(7):101971, 2025.

[445] World Meteorological Organization. Record carbon emissions highlight urgency of global greenhouse gas watch, 2024. https://wmo.int/media/news/record-carbon-emissions-highlight-urgency-of-global-greenhouse-gas-watch#:~:text=Until%20we%20reach%20net%20zero,one%20of%20WMO's%20flagship%20initiatives. [accessed 25 August 2025].

[446] Hannah Ritchie. How much co2 can the world emit while keeping warming below 1.5°c and 2°c? *Our World in Data*, 2023. https://ourworldindata.org/how-much-co2-can-the-world-emit-while-keeping-warming-below-15c-and-2c.

[447] Robin D Lamboll, Zebedee RJ Nicholls, Christopher J Smith, Jarmo S Kikstra, Edward Byers, and Joeri Rogelj. Assessing the size and uncertainty of remaining carbon budgets. *Nature Climate Change*, 13(12):1360–1367, 2023.

[448] Jonathan M Cullen and Julian M Allwood. The efficient use of energy: Tracing the global flow of energy from fuel to service. *Energy policy*, 38(1):75–81, 2010.

[449] Simon De Stercke. Dynamics of energy systems: A useful perspective. *Interim Report IR-14-013, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria*, 2014.

[450] Brendon Tankwa, Emilien Ravigné, and J. Doyne Farmer. Climate policy reforms and the acceleration of solar and wind diffusion. INET Oxford Working Paper Series 2026-01, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Jan 2026. Working Paper.

[451] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3):917–1007, 2018.

[452] Robert E Lucas Jr. Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46. North-Holland, 1976.

[453] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.

[454] Mark A Beaumont. Approximate bayesian computation. *Annual Review of Statistics and Its Application*, 6:379–403, 2019.

Epilogue

Here I offer some concluding thoughts and reflections on the thesis, the DPhil, and my own experiences. One can so easily get lost in the details of research and forget to step back and reflect on the bigger picture. I wrote 75,000 words and made 2,500 figures—that is only 60 words per day. I consumed approximately 5,000 cups of coffee, held over 20 presentations on S-curves, was yelled at only twice, had doubts, and experienced anxiety and dread about my work.

But this is nothing compared to these statistics: I have cried of laughter 1000 times, had thousands of conversations with friends, family and loved ones supporting me, and had 1000 moments of joy and excitement when I made a new discovery or found a new insight. I got to go to a conference in China and experience what it means to be an academic conveniently travelling the world—for academic pursuit only, of course. I am grateful for all of these experiences, the good and the bad, because they have shaped me into the person I am today and made this thesis what it is.

My DPhil journey, much like the research itself, has been an S-curve: initial disruption as an incoming new DPhil, slow development but some could say rapid growth rate, sustained growth until marginal gains became harder to achieve—a natural saturation point. Looking back, I myself failed to forecast my own S-curve: I reached a higher asymptote than I ever expected. I hope that you, the reader, have found some value in this thesis, and that it can contribute to a deeper understanding of technology, energy transitions, and the future of our planet.