

Climate Econometrics: An Overview

Suggested Citation: Jennifer L. Castle and David F. Hendry (2019), "Climate Econometrics: An Overview", : Vol. xx, No. xx, pp 1–19. DOI: 10.1561/XXXXXXXXXX.

Jennifer L. Castle

Institute for New Economic Thinking at the
Oxford Martin School and
Climate Econometrics at Nuffield College,
University of Oxford
jennifer.castle@magd.ox.ac.uk

David F. Hendry

Institute for New Economic Thinking at the
Oxford Martin School and
Climate Econometrics at Nuffield College,
University of Oxford
david.hendry@nuffield.ox.ac.uk

This article may be used only for the purpose of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval.

now
the essence of knowledge

Boston — Delft

Contents

Acknowledgements	4
1 Introduction	5
2 Econometric methods for empirical climate modeling	12
2.1 Theory models and data generation processes	13
2.2 Formulating wide-sense non-stationary time series models	14
2.3 Model selection by <i>Autometrics</i>	19
2.4 Model selection facing wide-sense non-stationarity . . .	19
2.5 Understanding why model selection can work	21
2.6 Selecting models with saturation estimation	26
2.7 Summary of saturation estimation	45
2.8 Selecting simultaneous equations models	48
2.9 Forecasting in a non-stationary world	52
3 Hazards in modeling non-stationary time-series data	58
3.1 Assessing the constancy and invariance of the relationship	63
3.2 An encompassing evaluation of the relationship	66
4 A brief excursion into climate science	68
4.1 Can humanity alter the planet's atmosphere and oceans?	70
4.2 Climate change and the 'great extinctions'	76

5	The Industrial Revolution and its consequences	82
5.1	Climate does not change uniformly across the planet . .	85
6	Identifying the causal role of CO₂ in Ice Ages	87
6.1	Data series over the past 800,000 years	92
6.2	System equation modeling of the ice-age data	95
6.3	Long-run implications	100
6.4	Looking ahead	105
6.5	Conclusions on Ice-Age modeling	109
7	Econometric modeling of UK annual CO₂ emissions	111
7.1	Data definitions and sources	113
7.2	UK CO ₂ emissions and its determinants	114
7.3	Model formulation	118
7.4	Evaluating a model without saturation estimation	121
7.5	Four stages of single-equation model selection	124
7.6	Selecting indicators in the general model	125
7.7	Selecting regressors and implementing cointegration . .	126
7.8	Estimating the cointegrated formulation	128
7.9	Encompassing of linear-semilog versus linear-linear . .	130
7.10	Conditional 1-step 'forecasts' and system forecasts . . .	131
7.11	Policy implications	133
7.12	Can the UK reach its CO ₂ emissions targets for 2050? .	135
7.13	Climate-environmental Kuznets curve	141
8	Conclusions	144
	References	150

Climate Econometrics: An Overview

Jennifer L. Castle¹ and David F. Hendry²

¹*Institute for New Economic Thinking at the Oxford Martin School and Climate Econometrics at Nuffield College, University of Oxford*

²*Institute for New Economic Thinking at the Oxford Martin School and Climate Econometrics at Nuffield College, University of Oxford*

ABSTRACT

Climate econometrics is a new sub-discipline that has grown rapidly over the last few years. As greenhouse gas emissions like carbon dioxide (CO₂), nitrous oxide (N₂O) and methane (CH₄) are a major cause of climate change, and are generated by human activity, it is not surprising that the tool set designed to empirically investigate economic outcomes should be applicable to studying many empirical aspects of climate change.

Economic and climate time series exhibit many commonalities. Both data are subject to non-stationarities in the form of evolving stochastic trends and sudden distributional shifts. Consequently, the well-developed machinery for modeling economic time series can be fruitfully applied to climate data. In both disciplines, we have imperfect and incomplete knowledge of the processes actually generating the data. As we don't know that data generating process (DGP), we must search for what we hope is a close approximation to it. The data modeling approach adopted at Climate Econometrics (<http://www.climateeconometrics.org/>) is based on a model selection methodology that has excellent properties for locating an unknown DGP nested within a large set of possible explanations, including dynamics, outliers, shifts,

and non-linearities. The software we use is a variant of machine learning which implements multi-path block searches commencing from very general specifications to discover a well-specified and undominated model of the processes under analysis. To do so requires implementing indicator saturation estimators designed to match the problem faced, such as impulse indicators for outliers, step indicators for location shifts, trend indicators for trend breaks, multiplicative indicators for parameter changes, and indicators specifically designed for more complex phenomena that have a common reaction ‘shape’ like the impacts of volcanic eruptions on temperature reconstructions. We also use combinations of these, inevitably entailing settings with more candidate variables than observations.

Having described these econometric tools, we take a brief excursion into climate science to provide the background to the later applications. By noting the Earth’s available atmosphere and water resources, we establish that humanity really can alter the climate, and is doing so in myriad ways. Then we relate past climate changes to the ‘great extinctions’ seen in the geological record. Following the Industrial Revolution in the mid-18th Century, building on earlier advances in scientific, technological and medical knowledge, real income levels per capita have risen dramatically globally, many killer diseases have been tamed, and human longevity has approximately doubled. However, such beneficial developments have led to a global explosion in anthropogenic emissions of greenhouse gases. These are also subject to many relatively sudden shifts from major wars, crises, resource discoveries, technology and policy interventions. Consequently, stochastic trends, large shifts and numerous outliers must all be handled in practice to develop viable empirical models of climate phenomena. Additional advantages of our econometric methods for doing so are detecting the impacts of important policy interventions as well as improved forecasts. The econometric approach we outline can handle all these jointly, which is essential to accurately characterize non-stationary observational data. Few approaches

in either climate or economic modeling consider all such effects jointly, but a failure to do so leads to mis-specified models and hence incorrect theory evaluation and policy analyses. We discuss the hazards of modeling wide-sense non-stationary data (namely data not just with stochastic trends but also distributional shifts), which also serves to describe our notation.

The application of the methods is illustrated by two detailed modeling exercises. The first investigates the causal role of CO₂ in Ice Ages, where a simultaneous-equations system is developed to characterize land ice volume, temperature and atmospheric CO₂ levels as non-linear functions of measures of the Earth's orbital path round the Sun. The second turns to analyze the United Kingdom's highly non-stationary annual CO₂ emissions over the last 150 years, walking through all the key modeling stages. As the first country into the Industrial Revolution, the UK is one of the first countries out, with per capita annual CO₂ emissions now below 1860's levels when our data series begin, a reduction achieved with little aggregate cost. However, very large decreases in all greenhouse gas emissions are still required to meet the UK's 2050 target set by its Climate Change Act in 2008 of an 80% reduction from 1970 levels, since reduced to a net zero target by that date, as required globally to stabilize temperatures. The rapidly decreasing costs of renewable energy technologies offer hope of further rapid emission reductions in that area, illustrated by a dynamic scenario analysis.

JEL classifications: C01, C51, C87, Q54.

Keywords: Climate Econometrics; Model Selection; Policy Interventions; Outliers; Saturation Estimation; *Autometrics*; Ice Ages; CO₂ Emissions.

Acknowledgements

Financial support from the Robertson Foundation (award 9907422), the Institute for New Economic Thinking (grant 20029822) and Nuffield College is gratefully acknowledged, as are valuable contributions to the background research from Jurgen A. Doornik, Luke P. Jackson, Søren Johansen, Andrew B. Martinez, Bent Nielsen and Felix Pretis. All calculations and graphs use *PcGive* (Doornik and Hendry, 2018) and *OxMetrics* (Doornik, 2018). We thank Rob Engle, Neil Ericsson, Vivien Hendry, Luke Jackson, Jonas Kurle, Andrew Martinez, Susana Martins, Bent Nielsen, Felix Pretis, Matthais Qian, Ryan Rafaty, Moritz Schwarz and Angela Wenham for helpful comments, and Thomas Sterner for the invitation to deliver a plenary session at the 2018 World Conference in Environmental and Resource Economics in Gothenburg on which this monograph is based.

1

Introduction

Climate econometrics is a sub-discipline that has grown rapidly over the last few years, having held four annual international conferences (at Aarhus, Oxford, Rome and Milan) and with a global network.¹ A Special Issue of the *Journal of Econometrics* (<https://www.sciencedirect.com/journal/journal-of-econometrics/vol/214/issue/1>) has 14 contributions across a wide range of climate issues, and a second in *Econometrics* (https://www.mdpi.com/journal/econometrics/special_issues/econometric_climate) is in preparation. Because greenhouse gas emissions like carbon dioxide (CO₂), nitrous oxide (N₂O) and methane (CH₄) are the major cause of climate change, and are generated by human activity, it is not surprising that the tool set originally designed to empirically investigate economic outcomes should be applicable to studying many empirical aspects of climate change. Most climate-change analysis is based on physical process models embodying the many known laws of conservation and energy balance at a global level. Such results underpin the various reports from the Intergovernmental Panel on Climate Change (IPCC: <https://www.ipcc.ch/>). Climate theories can also be embedded in models of the kind familiar from macroeconomics: for example, Robert Kaufmann, Heikki Kauppi, Michael

¹ See <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=climateeconometrics>: its planned 5th Econometric Models of Climate Change Conference at the University of Victoria has had to be postponed till 2021 because of the SARS-CoV-2 pandemic.

Mann, and James Stock (2013) link physical models with statistical ones having a stochastic trend, and Felix Pretis (2019) establishes an equivalence between two-component (i.e., atmosphere and oceans) energy-balance models of the climate and a cointegrated vector autoregressive system (CVAR). Even in such a well-understood science, knowledge is not complete and immutable, and there are empirical aspects that need attention. For example, CO₂ and other greenhouse gas emissions depend on changeable human behavior; volcanic eruptions vary greatly in their climate impacts; the rate of loss of Arctic sea ice alters the Earth's albedo and such feedbacks affect warming.

Our approaches at Climate Econometrics (our research group, shown capitalized to differentiate it from the general research area) are complementary to physical process models, and use a powerful set of modeling tools developed to analyze empirical evidence on evolving processes that are also subject to abrupt shifts, called *wide-sense non-stationarity* to distinguish from the use of 'non-stationarity' purely for unit-root processes that generate stochastic trends: see Castle and Hendry (2019). A key reason is that differencing a wide-sense non-stationary time series does **not** ensure stationarity as is often incorrectly assumed in economics. Because the data are wide-sense non-stationary time series observations, the data generating process (DGP) is inevitably unknown and has to be discovered. The model selection methodology described below has excellent properties for locating an unknown DGP when it is embedded within a large set of potential explanations. Thus, we advocate commencing from a general specification that also includes variables to allow for dynamics, outliers, shifts, and non-linearities. We use a variant of machine learning called *Autometrics* that explores multi-path block searches to discover a well-specified and undominated model of the processes under analysis (see Jurgen Doornik, 2009). Hendry and Doornik (2014) analyze the properties of *Autometrics*: also see Section 2.3.² The approach is available in R by Pretis, James Reade, and Genaro Sucarrat (2018a) at <https://cran.r-project.org/web/packages/gets/index.html>, and as the Excel Add-in *XLModeler* (see <https://xlmodeler.com/>). Other model selection algorithms include the Lasso (see Robert Tibshirani, 1996) and its variants.

Our methods are designed to select models even when there are more

²For summaries, see <http://voxeu.org/article/data-mining-more-variables-observations> and <https://voxeu.org/article/improved-approach-empirical-modelling-0>.

candidate variables, N , than the number of observations, T . *Autometrics* employs a variety of saturation estimators that inevitably create $N > T$. Each is designed to match the problem faced, namely impulse-indicator saturation (denoted IIS) to tackle outliers, step-indicator saturation (SIS) for location shifts, trend-indicator saturation (TIS) for trend breaks, multiplicative-indicator saturation (MIS) for parameter changes, and designed-indicator saturation for modeling phenomena with a regular pattern, applied below to detecting the impacts on temperature of volcanic eruptions (VIS). Importantly, saturation estimators can be used in combination, and can be applied when retaining without selection a theory-model that is the objective of a study, while selecting from other potentially substantive variables. Saturation estimators, and indeed our general approaches, have seen applications across a range of disciplines including dendrochronology, volcanology, geophysics, climatology, and health management, as well as economics, other social sciences and forecasting. Although theory models are much better in many of these areas than in economics and other social sciences, modeling observational data faces most of the same problems, which is why an econometric toolkit can help.

Below, we explain our econometric methods and illustrate some of their applications to climate time series. The first illustration investigates past climate variability over the Ice Ages, where a simultaneous-equations system is developed to characterize land ice volume, Antarctic temperature and atmospheric CO₂ levels as non-linear functions of measures of the Earth's evolving orbital path round the Sun. The focus is on system modeling and how we implement that despite $N > T$, as well as the difference in how saturation estimation is applied in systems. Few economists will ever have the opportunity to consider multi-step forecasts over 100,000 years as we do here! The second illustration is a detailed study of the UK's CO₂ emission over 1860–2017 that walks through the various stages of formulation, model specification, selection while tackling outliers and location shifts, then investigating cointegration, and on to model simplification for forecasting and policy analyses. A key aim is establishing the possible impacts of past policy interventions though we also discuss possible future developments.

As Pretis (2019) remarks 'Econometric studies beyond IAMs (integrated assessment models) are split into two strands: one side empirically models the impact of climate on the economy, taking climate variation as given ... the other side models the impact of anthropogenic (e.g., economic) activity onto

the climate by taking radiative forcing—the incoming energy from emitted radiatively active gases such as CO_2 —as given... This split in the literature is a concern as each strand considers conditional models, while feedback between the economy and climate likely runs in both directions.’ Examples of approaches conditioning on climate variables such as temperature include Marshall Burke, Solomon Hsiang and Edward Miguel (2015), Pretis, Moritz Schwarz, Kevin Tang, Karsten Haustein and Myles Allen (2018b), Burke, Matthew Davis and Noah Diffenbaugh (2018), and Davis (2019). Hsiang (2016) reviews such approaches to climate econometrics. Examples from many studies modeling climate time series include Francisco Estrada, Pierre Perron and Benjamín Martínez-López (2013), Kaufmann *et al.* (2011, 2013) and Pretis and Hendry (2013). Pretis (2017) addresses the exogeneity issue in more detail. Most of the research described in this monograph concerns the second approach, although the methods are applicable both to the first and to investigating exogeneity as shown in Chapter 6. The resulting econometric tools also contrast with the methodology predominantly used in the first approach of a quasi-experimental framework using panel regressions under the assumption of strict exogeneity of climate variables.

The structure of the monograph is as follows. First, Chapter 2 describes econometric methods for empirical climate modeling that can account for wide-sense non-stationarity, namely both stochastic trends and location shifts, with possibly large outliers, as well as dynamics and non-linearities. Model selection is essential as the behavioral processes determining greenhouse gas emissions are too complicated to be known *a priori*. A basic question then concerns what is model selection trying to find? This is answered in Section 2.1 on the roles therein of theory models and DGPs by trying to find the latter, or at least a good approximation to its substantive components. Section 2.2 first discusses the formulation of models for wide-sense non-stationary time series, then Section 2.3 describes model selection by *Autometrics* and Section 2.4 explains its block multi-path selection algorithm. Next, Section 2.5 turns to understanding why automatic model selection can work well despite $N > T$. Saturation estimators are described in Section 2.6, commencing with impulse-indicator saturation (IIS) to tackle outliers. IIS is illustrated in §2.6.1, and its properties are described in §2.6.2. Then §2.6.3 considers step-indicator saturation (SIS), §2.6.4 the extension to super saturation estimation combining IIS and SIS, §2.6.5 explains a variant to handle trend saturation

estimation (TIS), followed in §2.6.6 by multiplicative-indicator saturation (MIS) which interacts SIS with regressors for detecting parameter changes. Then §2.6.7 illustrates designed-indicator saturation by formulating indicators for modeling the impacts of volcanic eruptions on temperature reconstructions (VIS). Section 2.7 summarizes the various saturation estimators. Section 2.8 considers selection, estimation and evaluation of simultaneous equations models, addressing identification in §2.8.1. Facing forecasting in a wide-sense non-stationary world, Section 2.9 discusses the consequences of not handling location shifts and describes forecasting devices that are more robust after shifts than ‘conventional’ forecasting models.

Chapter 3 considers hazards confronting empirical modeling of non-stationary time-series data using an example where a counter-intuitive finding is hard to resolve. The framework has a clear subject-matter theory, so is not mere ‘data mining’, yet the empirical result flatly contradicts the well-based theory. Section 3.1 considers whether assessing the constancy and invariance of the relationship can reveal the source of the difficulty, but does not. An encompassing evaluation of the relationship in Section 3.2 fortunately does.

Chapter 4 provides a brief excursion into climate science, mainly concerned with the composition of the Earth’s atmosphere and the role of CO₂ as a greenhouse gas. Section 4.1 considers whether humanity can alter the planet’s atmosphere and oceans, and demonstrates we can—and are. Section 4.2 discusses the consequences of changes in the composition of the atmosphere, focusing on the impacts of climate change on ‘great extinctions’ over geological time.

Chapter 5 considers the consequences, both good and bad, of the Industrial Revolution raising living standards beyond the wildest dreams of those living in the 17th century, but leading to dangerous levels of CO₂ emissions from using fossil fuels.

Against that background, we consider applications of climate econometrics. Chapter 6 illustrates the approach by modeling past climate variability over the Ice Ages. Section 6.1 describes the data series over the past 800,000 years, then Section 6.2 models ice volume, CO₂ and temperature as jointly endogenous in a 3-variable system as a function of variations in the Earth’s orbit, taking account of dynamics, non-linear interactions and outliers using full information maximum likelihood. The general model is formulated in §6.2.1, and the simultaneous system estimates are discussed in §6.2.2. Their

long-run implications are described in Section 6.3 with one hundred 1000-year 1-step and dynamic forecasts in §6.3.1. Then, §6.3.2 considers when humanity might have begun to influence climate, and discusses the potential exogeneity of CO₂ to identify its role during Ice Ages. Section 6.4 looks 100,000 years into the future using the fact that the eccentricity, obliquity and precession of Earth's orbital path is calculable far into the future, to explore the implications for the planet's temperature of atmospheric CO₂ being determined by humans at levels far above those experienced during Ice Ages. Finally, Section 6.5 summarizes the conclusions on Ice-Age modeling.

Chapter 7 models UK annual CO₂ emissions over 1860–2017 to walk through the stages of modeling empirical time series that manifest all the problems of wide-sense non-stationarity. Section 7.1 provides data definitions and sources, then Section 7.2 discusses the time-series data. Section 7.3 formulates the econometric model, then Section 7.4 highlights the inadequacy of simple model specifications. The four stages of model selection from an initial general model are described in Section 7.5, then implemented in Sections 7.6–7.8. Section 7.9 conducts an encompassing test of the linear-semilog model against a linear-linear one. Section 7.10 presents conditional 1-step 'forecasts' and multi-step forecasts from a VAR. Section 7.11 addresses the policy implications of the empirical analysis, then Section 7.12 considers whether the UK can reach its 2008 Climate Change Act (CCA) CO₂ emissions targets for 2050. Finally, Section 7.13 estimates a 'climate-environmental Kuznets curve'.

Chapter 8 concludes and summarizes a number of other empirical applications.

To emphasize the different and interacting forms of non-stationarity, Figure 1.1 records time series from climate and economic data. Panel (a) shows the varying trends in global monthly atmospheric CO₂ concentrations in ppm measured at Mauna Loa over 1958(1)–2019(6); Panel (b) records the dramatically non-stationary UK per capita CO₂ emissions, with up and down trends, outliers and shifts; Panel (c) reports the log of UK GDP, again with changing trends and large shifts; and (d) plots the log of the UK wage share, with large shifts and outliers.

The lockdowns round the world in response to SARS-CoV-2 will doubtless cause a sharp drop in global CO₂ emissions in early 2020 needing modeled. The indicator saturation estimators described in Chapter 2 are designed to

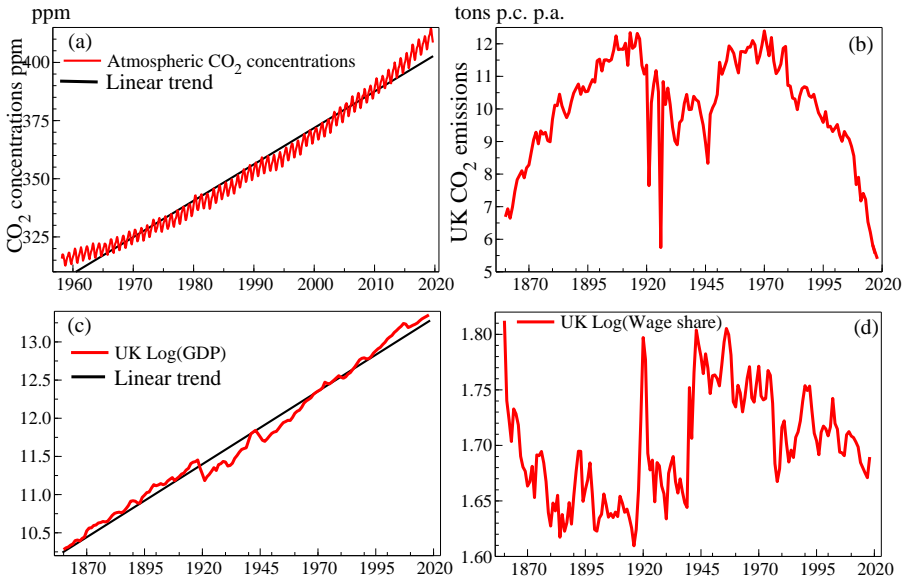


Figure 1.1: (a) global monthly atmospheric CO₂ concentrations in parts per million (ppm) measured at Mauna Loa, 1958(1)–2019(6); (b) UK CO₂ emissions in tons per capita per annum; (c) the log of UK GDP; (d) log of the UK wage share. (b)–(d) are all annual over 1860–2018.

tackle such multiple shifts of unknown magnitudes and directions at unknown dates as countries gradually bring their pandemics under sufficient control to ‘restart’ their economies.

2

Econometric methods for empirical climate modeling

In this Chapter, we describe the econometric tools that are needed for empirical climate modeling of wide-sense non-stationary data. Section 2.1 commences with a discussion of the *objective* of the study, usually a theory-based formulation, as against what should be the *target* for modeling. Often the objective is made the target, but that needs omniscience: instead the target should be the process that generated the data while retaining the object of analysis. Such an approach allows for the possibility of finding that the target and object coincide without imposing that they must. Section 2.2 describes the formulation of models for wide-sense non-stationary time series, then Section 2.3 discusses model selection by *Autometrics* and Section 2.4 explains the block multi-path selection algorithm. Section 2.5 analyzes why automatic model selection can work well despite $N > T$, building on Hendry and Doornik (2014), before Section 2.6 explains saturation estimators, summarizing the different saturation approaches in Section 2.7. The selection, estimation and evaluation of simultaneous-equations models are considered in Section 2.8 commencing from a dynamic system, with the issue of identification addressed in §2.8.1. Section 2.9 discusses forecasting in a wide-sense non-stationary world.

2.1 Theory models and data generation processes

The most basic question concerns ‘what is empirical model selection trying to find’? Given the answer to that, then one can address how best to find it. Many features of models of observational data cannot be derived theoretically, particularly facing wide-sense non-stationarity. While a theory-model provides the *object* of interest to a modeler, that theory can only be the *target* of a study if it is complete, correct and immutable, despite often being imposed as the target yet lacking those characteristics. Facing non-stationary time series, *ceteris paribus* is simply not applicable because what is excluded will not stay the same: see Marcel Boumans and Mary Morgan (2001). Viable models of non-stationarity must include everything that matters empirically if estimated models are to be constant.

To understand how any system actually functions, the appropriate target for model selection must be its *data generation process* (DGP). The DGP of the world is the joint density $D_{\mathbf{W}_T^1}(\mathbf{w}_1 \dots \mathbf{w}_T | \psi_T^1, \mathbf{W}_0)$ where \mathbf{W}_T^1 is the complete set of variables over a time period $1, \dots, T$, conditional on the past, \mathbf{W}_0 . However, $D_{\mathbf{W}_T^1}(\cdot)$ and the ‘parameters’ $\psi_T^1 \in \Psi$ of the processes may be time varying. In practice, DGPs are too high dimensional and too non-stationary to develop complete theories about, or to precisely model empirically, so local DGPs (denoted LDGP) are usually the best that can be achieved. The LDGP is the DGP for the n variables $\{\mathbf{x}_t\}$ which an investigator has chosen to model, with entailed ‘parameters’ $\theta_T^1 \in \Theta$. The theory of reduction explains how the LDGP $D_{\mathbf{X}_T^1}(\cdot)$ is derived from $D_{\mathbf{W}_T^1}(\cdot)$, the resulting transformations of the ‘parameters’ that implies, and what the properties of $D_{\mathbf{X}_T^1}(\cdot)$ will be (see e.g., Hendry, 2009).

The LDGP $D_{\mathbf{X}_T^1}(\cdot)$ can always be written by sequential factorization with a martingale difference (innovation) error, that is unpredictable from the past of the process (see Joseph Doob, 1953):

$$D_{\mathbf{X}_T^1}(\mathbf{X}_T^1 | \mathbf{X}_0, \theta_T^1) = \prod_{t=1}^T D_{\mathbf{x}_t}(\mathbf{x}_t | \mathbf{X}_{t-1}^1, \mathbf{X}_0, \theta_t). \quad (2.1)$$

Thus, the joint density can be expressed as the product of the sequential individual densities even when the ‘parameters’ are not constant. Let $\mathbb{E}_{\mathbf{X}_{t-1}}$ denote the expectation over the distribution $D_{\mathbf{X}_{t-1}}(\cdot)$, and define $\epsilon_t = \mathbf{x}_t - \mathbb{E}_{\mathbf{X}_{t-1}}[\mathbf{x}_t | \mathbf{X}_{t-1}^1]$, then $\mathbb{E}_{\mathbf{X}_{t-1}}[\epsilon_t | \mathbf{X}_{t-1}^1] = \mathbf{0}$, so $\{\epsilon_t\}$ is indeed a martin-

gale difference error process, with $\mathbb{E}_{\mathbf{x}_{t-1}} [\epsilon_t | \mathbf{E}_{t-1}^1] = \mathbf{0}$ where $\mathbf{E}_{t-1}^1 = (\epsilon_{t-1} \dots \epsilon_1)$. This provides a viable basis for laws of large numbers, central limit theorems and congruent models which are models that match the evidence, so are ‘well specified’. Note that the LDGP innovation error $\{\epsilon_t\}$ is designed, or created, by the reductions entailed in moving from the DGP to the LDGP, so is not an ‘autonomous’ process, but rather a reflection of our ignorance. A larger choice of relevant variables than $\{\mathbf{x}_t\}$ would make the LDGP a better approximation to the actual DGP, which should deliver smaller innovation errors, sustaining a progressive research strategy. Once whatever set of $\{\mathbf{x}_t\}$ has been chosen, one cannot do better than know its LDGP $D_{\mathbf{x}_T^1}(\cdot)$, which encompasses all models thereof on the same data (i.e., can explain their findings: see Christophe Bontemps and Grayham Mizon, 2008). Consequently, the LDGP is the only appropriate target for model selection.

However, LDGPs are almost always unknown in practice, so Hendry and Doornik (2014) emphasize the need to discover the LDGP from the available evidence while retaining theory information. Doing so requires nesting the LDGP in a suitably *general unrestricted model* (denoted GUM), while also embedding the theory model in that GUM, then searching for the simplest acceptable representation, stringently evaluating that selection for congruence and encompassing. Since the variables $\{\mathbf{x}_t\}$ chosen for analysis usually depend on available subject-matter theory, institutional knowledge, and previous evidence, most theory-model objects can be directly related to the target LDGP by embedding them therein.

Unfortunately, (2.1) provides cold comfort for empirical modelers: sequential factorization only delivers an innovation error when using the correct sequential distributions $D_{\mathbf{x}_t}(\cdot)$. To discover that LDGP therefore requires also finding all distributional shifts, as omitting key variables and/or shifts will adversely affect selected representations. Sections 2.2–2.8 describe modeling aimed at discovering the LDGP. Then section 2.9 considers forecasting in a non-stationary world, where surprisingly different approaches may be useful.

2.2 Formulating wide-sense non-stationary time series models

Empirical modeling of observational time series inevitably involves uncertainty about which ‘explanatory’ variables are relevant and which are irrelevant, their functional forms, lag lengths, unit roots and cointegration, possible

outliers, structural breaks and location shifts, as well as the constancy and invariance of parameters. Moreover, it is essential for valid inference to check the exogeneity status of contemporaneous conditioning variables, whether the resulting residuals satisfy the error properties assumed in deriving the distributions of parameter estimates, so that selected models are congruent, and whether the selected model can also encompass alternative explanations of the same dependent variables. Thus, there is a set of tests for evaluating empirical modeling outcomes, and consequently, satisfying these should also constrain the selection process.

As it is rare not to have some theoretical basis to guide an empirical analysis, algorithms should retain without selection all the variables in a theory-model when selecting other features. Doing so enables much tighter than conventional significance levels to be used during selection to reduce adventitious retention of irrelevant candidates without jeopardizing the retention of any theory-relevant variables that are the object of the analysis. Hendry and Søren Johansen (2015) propose an approach in which those other variables are orthogonalized with respect to the theory variables, so the distributions of estimators of the parameters of the object are unaffected by selection: for additional analyses and an empirical illustration, see Hendry (2018).

To formalize our notation, denote the variables to be modeled by \mathbf{y}_t , the theory-model contemporaneous ‘explanatory’ variables by the $n_1 \times 1$ vector \mathbf{z}_t , and the other current-dated candidate variables by the $n_2 \times 1$ vector \mathbf{v}_t (possibly after orthogonalization with respect to \mathbf{z}_t). Let $\mathbf{w}'_t = (\mathbf{z}'_t, \mathbf{v}'_t)$ which is $n \times 1$ where $n = n_1 + n_2$ (although \mathbf{w}_t is the same symbol as in Section 2.1, here it denotes a tiny subset of all possible variables). §2.2.1 considers specifying lag length; §2.2.2 functional forms; §2.2.3 the formulation of the resulting general unrestricted model; and §2.2.4 model evaluation. Saturation estimators are not discussed in detail till Section 2.6.

2.2.1 Lag length specification

Almost all empirical econometric models of time series are dynamic as events take time to work through economies. The Earth’s climate also adjusts relatively slowly to changes in greenhouse gas emissions because of the absorption and temperature interactions between the oceans and the atmosphere, so such models have to be dynamic as well. The first step is to create s lags,

$\mathbf{w}_t \dots \mathbf{w}_{t-s}$, to implement the sequential factorization in equation (2.1) for the GUM. This is to avoid residuals from the general model being autocorrelated. The sequential densities in (2.1) should include all lags of the data, but in practice, lags are truncated at s , where a sufficient number of lags are retained to ensure there is no loss of dynamic information, with residuals that are white noise. Nevertheless, as noted above, distributional shifts will also need to be modeled to achieve that, a topic left till Section 2.6, though included in (2.2).

2.2.2 Functional forms

Many economic models are non-linear in that the variables are transforms of the original measurements (not simply logs) as with the quadratic relationship between real wages and the unemployment rate in Castle and Hendry (2014b). Because of the potential for climate tipping points, such as when an ice-free Arctic Sea led to large-scale methane release from the permafrost melting in the tundra causing rapid climate warming (see Anton Vaks *et al.*, 2020), non-linear relationships cannot be neglected. More positively, sensitive intervention points in the post-carbon transition could induce leverage in policy actions inducing non-linearities in models (see Doyme Farmer *et al.*, 2019).

A class of functional-form transformations for non-linearity tests (denoted F_{nl} below) was proposed by Castle and Hendry (2010) based on $(u_{i,t}^2; u_{i,t}^3; u_{i,t}e^{-|u_{i,t}|})$ obtained from the principal components of the \mathbf{w}_t , given by $\mathbf{u}_t = \hat{\mathbf{H}}'(\mathbf{w}_t - \bar{\mathbf{w}})$, where $\hat{\mathbf{\Omega}} = T^{-1} \sum_{t=1}^T (\mathbf{w}_t - \bar{\mathbf{w}})(\mathbf{w}_t - \bar{\mathbf{w}})' = \hat{\mathbf{H}}\hat{\mathbf{\Lambda}}\hat{\mathbf{H}}'$. When $\hat{\mathbf{\Omega}}$ is non-diagonal, each $u_{i,t}$ is a linear combination of many $w_{j,t}$, so (e.g.) $u_{i,t}^2$ involves squares and cross-products of almost every $w_{j,t}$ etc. This approach can also be used to automatically generate a set of non-linear transforms of the variables for including in the GUM. The formulation is low dimensional compared to a general cubic, with no collinearity between the $\{u_{i,t}\}$, yet includes many departures from linearity (see Castle and Hendry, 2014b, for an illustration). However, such non-linear transforms should be restricted to $l(0)$ variables as Castle *et al.* (2019a) show that calculating principal components on $l(1)$ variables with stochastic trends, where some may be irrelevant, can distort analyses of the cointegrating combinations. In §7.3.1 below, we consider log versus linear transformations in the UK CO₂ model, as well as applying F_{nl} .

2.2.3 Formulation of single-equation general unrestricted model

Since climate time series are wide-sense non-stationary, models thereof must include features to tackle that. As well as many candidate explanatory variables, dynamics and non-linearities, we include T indicators for impulses denoted $1_{\{t\}}$ equal to zero except for unity at t , and $T - 2$ step shifts $S_{\{i \leq t\}} = \sum_{j=1}^t 1_{\{j\}}$ (which Neil Ericsson, 2012, calls super-saturation: see §2.6.1 and §2.6.3). This results in $N = 4n(s + 1) + s + 2(T - 1)$ candidate regressors including indicators, so $N \gg T$. Denoting by $[]$ the set of variables that will be retained without selection (which may include lags but we avoid that notational complication), the resulting general unrestricted model is given by:

$$\begin{aligned}
 y_t = & \left[\sum_{i=1}^{n_1} \theta_i z_{i,t} \right] + \sum_{i=1}^{n_2} \phi_i v_{i,t} + \sum_{i=1}^n \sum_{j=1}^s \beta_{i,j} w_{i,t-j} + \sum_{i=1}^n \sum_{j=0}^s \lambda_{i,j} u_{i,t-j}^2 \\
 & + \sum_{i=1}^n \sum_{j=0}^s \gamma_{i,j} u_{i,t-j}^3 + \sum_{i=1}^n \sum_{j=0}^s \kappa_{i,j} u_{i,t-j} e^{-|u_{i,t-j}|} + \sum_{j=1}^s \rho_j y_{t-j} \\
 & + \sum_{i=1}^T \delta_i 1_{\{t\}} + \sum_{i=1}^{T-1} \eta_i S_{\{i \leq t\}} + \epsilon_t.
 \end{aligned} \tag{2.2}$$

It is also assumed that $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$, denoting an independent, Normal distribution with constant mean zero and constant variance σ_ϵ^2 . Since outliers and shifts are modeled, Normality seems reasonable, and the sequential decomposition entailed by including $s \geq 1$ lags on all variables makes ‘independence’ (in the sense of a martingale difference sequence) a viable assumption. The three more critical assumptions that need addressed later are (a) the constancy and (b) invariance of the parameters, and (c) the super exogeneity of the contemporaneous variables for the parameters in (2.2): see Robert Engle, Hendry and Jean-François Richard (1983). The first is discussed below in the context of multiplicative-indicator saturation after an initial selection; the second by testing for invariance of the parameters of the selected model to interventions affecting the subset of the \mathbf{w}_t retained (described in Section 3.1), and the third indirectly by that invariance test when the data are wide-sense non-stationary, and by an appropriate estimator allowing for potential endogeneity. Generalizations to systems are discussed in Section 2.8.

The general concept underlying (2.2) is that of designing the model to characterize the evidence as discussed in Section 2.1. This notion applies to

the saturation estimators as well, in that indicators have formulations to detect specific departures from the rest of the model: see Section 2.6. Thus, impulse indicators detect outliers not accounted for by other variables, step indicators are designed to detect location shifts etc., and below we consider indicators with break shapes that are explicitly designed to detect regular shift patterns.

2.2.4 Model evaluation

Reductions from the DGP via the LDGP to a general model have testable implications discussed in Hendry (1995). There are 7 testable null hypotheses about the congruence of the initial feasible GUM, and also the final model, albeit there are many alternatives to each null. First are Normal, homoskedastic, innovation errors, $\{\epsilon_t\}$. Below, these null hypotheses are tested by $\chi^2_{nd}(2)$ for non-Normality (see Doornik and Henrik Hansen, 2008), F_{AR} for residual autocorrelation (see Lesley Godfrey, 1978), F_{ARCH} tests for autoregressive conditional heteroskedasticity (see Engle, 1982), and F_{Het} for residual heteroskedasticity (see Halbert White, 1980).¹

Next, conditioning variables w_t should be weakly exogenous for the complete set of parameters, denoted ψ , which should be constant and invariant. As just noted, we will test that joint hypothesis through super exogeneity (see Engle and Hendry, 1993), partly by saturation estimators applied to the conditioning variables (see Hendry and Carlos Santos, 2010) and partly by forecast evaluation discussed in Section 2.9. Also, the conditioning relationship should be linear in the transforms used to define the w_t , tested below by the RESET test F_{Reset} (see James Ramsey, 1969), and F_{nl} described in §2.2.2.

Given a formulation like (2.2) where $N \gg T$, how can a model of the underlying LDGP be selected from the GUM? A powerful machine-learning tool like automatic model selection is needed for such high-dimensional problems that humans cannot tackle: there are many available algorithms to do so, we use *Autometrics* (Doornik, 2009).

¹Most of these tests were developed for known models fitted to stationary data, and usually need simulation calibration outside that context. Bent Nielsen (2006) extends the theory for residual autocorrelation tests to non-stationary autoregressive-distributed lag models (ADL) with polynomial trends. Carlos Caceres (2007) shows that residual-based tests for mis-specification in models of unit-root processes have asymptotic sizes that remain close to the nominal significance level under the null. Vanessa Berenguer-Rico and Ines Wilms (2020) develop analyses of heteroscedasticity testing after outlier removal.

2.3 Model selection by *Autometrics*

Autometrics is implemented in a likelihood framework, separating the four key components of the selection approach: the model class; the search algorithm; model evaluation; and the decision on when a final selection has been found. Separating these roles allows considerable flexibility in what models and data types can be analyzed, how they are estimated, selected and evaluated, how theory information is incorporated and how the final model choice is made.

The model class includes simultaneous systems through to conditional single equations, their associated data types from discrete, time series, cross sections, panels etc., and their corresponding estimation criteria, from least squares, through instrumental variables to maximum likelihood. The search algorithm is a multiple-path tree search that ‘learns’ what variables are relevant as it proceeds, with many potential settings for the thoroughness of search, and the significance level needed to retain variables. Next, evaluation checks are carried out at every stage to try and ensure congruent selections, starting with mis-specification testing of the first feasible GUM and applying the same diagnostic tests to intermediate models. These are used in the selection process to limit the loss of information from the attempted reduction, not to evaluate the model. Thus, we call them mis-specification tests when they are first applied, and diagnostic tools at later stages. As shown in Hendry and Hans-Martin Krolzig (2005), the distributions of the tests are not affected by their role as diagnostic tools. Fourth, the termination decision is based on the initial nominal significance level for false null retention set by the user (called the gauge: see §2.5.1), and is also checked by parsimonious encompassing of the GUM to ensure that the cumulation of model simplifications did not lead to a poor model choice: see Doornik (2008). The program records all terminal model selections that are undominated, and if several are found, uses an information criterion like AIC (see Hirotugu Akaike, 1973), BIC or SC (see Gideon Schwarz, 1978) or HQ (see Edward Hannan and B.G. Quinn, 1979).

2.4 Model selection facing wide-sense non-stationarity

Once $N > T$, selection is essential as the GUM cannot be estimated as it stands. More generally, every decision about (a) how a theory is formulated, (b) how it is implemented, (c) its evidential base, (d) its empirical specifica-

tion, and (e) its evaluation, involves selection, although these are not usually reported as such. Consider a simple theory model that derives the equation $y = f(\mathbf{x})$. Then the formulation concerns the choice of $f(\cdot)$, and any transformations (such as logs) of the variables. How it is then implemented depends on many decisions about the dynamics linking y and \mathbf{x} , hence on the decision time frame of the agents involved (e.g., daily, weekly, monthly etc.), and whether the components of \mathbf{x} are given or need to be instrumented. The evidential base may be at a different data frequency and aggregation level across agents, space, commodities, etc., than the theory, and for a selected time period, perhaps chosen at the maximum available, but sometimes excluding turbulent periods like wars or crises. It will be interesting to see how researchers handle the massive disruptions of the pandemic. The empirical specification may lead to a GUM like (2.2), where there are additional features not included in the theory-model, so may include more candidate variables as ‘robustness’ checks, as well as data-based indicators. This is often the stage to which ‘selection’ is confined in reporting empirical results, though sometimes the reasons behind the earlier choices are noted. Finally, there are many possible evaluation tests, only a subset of which is often selected for reporting. When $N > T$ is very large, there is also a computational problem of being able to analyze large numbers of candidate variables, an issue not addressed here (but see e.g., Doornik and Hendry, 2015, and Castle *et al.*, 2019a). Castle *et al.* (2019c) describe the selection algorithm in detail.

In the absence of an omniscient researcher, selection is inevitable when modeling observational data: the issue is not whether to select, but how best to do so. Our viewpoint is that an investigator must *discover* what actually matters empirically, in the context of retaining theory insights and institutional knowledge, while encompassing previous findings and alternative explanations: see Hendry and Doornik (2014). The software implementation of that approach is *Autometrics* in Doornik and Hendry (2018), in essence a machine learning algorithm within each study, building knowledge of the main determinants and non-stationarities. To explain how and why model selection can be successful, we first consider its application to linear regression equations with orthogonal variables when $T \gg N$. Although this may not seem a likely setting for investigating non-stationary data modeled by $N > T$, the key principles can be explained, then extended to more variables than observations while retaining theory-models and tackling wide-sense non-stationarity.

2.5 Understanding why model selection can work

Consider a perfect setting: a well specified linear model with constant parameters β_j , an independent Normal error $\{\epsilon_t\}$ that is also independent of the N mutually orthogonal regressors $\{z_{j,t}\}$:

$$y_t = \sum_{j=1}^N \beta_j z_{j,t} + \epsilon_t, \text{ where } \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2], \quad (2.3)$$

with $\sum_{t=1}^T z_{i,t} z_{j,t} = 0 \forall i \neq j$ and $\sum_{t=1}^T z_{i,t}^2 > 0$, estimated by least-squares from $T \gg N$ correctly measured observations. The first n variables are relevant with $\beta_i \neq 0$, whereas the last $N - n$ are irrelevant with $\beta_j = 0$, but this is not known. Nevertheless, (2.3) nests the DGP.

Estimate (2.3), then order the N sample t^2 -statistics, denoted τ_j^2 , testing $H_0: \beta_j = 0$ (using squares to obviate the need to consider signs), as:

$$\tau_{(N)}^2 \geq \dots \geq \tau_{(m)}^2 \geq c_\alpha^2 > \tau_{(m-1)}^2 \geq \dots \geq \tau_{(1)}^2 \quad (2.4)$$

where the cut-off $\tau_{(m)}^2$ between retaining and excluding variables uses the critical value c_α^2 for significance level α . Variables with the $r = N - m + 1$ largest τ_j^2 values greater than c_α^2 are retained whereas the remaining $m - 1$ others are eliminated. Only one decision is needed (which we call 1-cut) however large N is: there is no ‘repeated testing’, and ‘goodness of fit’ is never explicitly considered, though will be implicitly determined by the magnitude of c_α^2 and the fit of the DGP equation. Ordering all the test statistics as in (2.4) requires considerable computational power for large N , and ‘looks like’ repeated testing, but is not in fact necessary, as each t^2 can just be compared to c_α^2 given orthogonality.

The key issue is what determines how close r is to n , when some or all of the n relevant variables are not retained without selection? There are two main components: the probability of eliminating the $N - n$ irrelevant variables, and the probability of retaining the n relevant. Setting $\alpha = 1$ ensures that all relevant variables will be retained, but so will all irrelevant; and setting $\alpha = 0$, no irrelevant variables will be retained, but no relevant either. Intermediate values of α will retain differing combinations, so we first address the probability of eliminating irrelevant variables when $n = 0$ in §2.5.1 (called the *gauge*), then consider the probability of retaining relevant variables when $n > 0$ in §2.5.2 (called the *potency*).

2.5.1 Probability of eliminating irrelevant variables

The average number of false null retentions can be kept at a maximum of k variables on average by setting $\alpha \leq k/N$. We call α the nominal significance level, and the resulting outcome, g , the theoretical, or expected, gauge. In the context of a 1-off similar test under the null, size is used to describe the null rejection frequency. However, there is no guarantee of selection distributions being independent of nuisance effects, and tests here are being used for selection, hence the need for a different term like gauge to describe false null retentions: see Johansen and Nielsen (2016) for an analysis.

When $n = 0$, so all N variables are irrelevant, $\alpha \approx g$ as shown in Table 2.1, which records the probabilities of all 2^N null rejection outcomes in t-testing at critical value c_α in (2.3). The first column records the events that can happen from all t-tests being insignificant through to all being significant. The second column shows the probability of each such event when the tests have independent t-distributions, as would be the case for (2.3), and the third reports the number of rejection outcomes that would result. The fourth column, denoted $p_{0.005}$, illustrates the numerical probabilities for column 2 when $\alpha = 0.005$ (i.e., 0.5%) and $N = 100$ when all variables are irrelevant.

event	probability	reject	$p_{0.005}$
$P(t_i < c_\alpha, \forall i = 1, \dots, N)$	$(1 - \alpha)^N$	0	0.61
$P(t_i \geq c_\alpha \mid t_j < c_\alpha, \forall j \neq i)$	$N\alpha(1 - \alpha)^{N-1}$	1	0.30
$P(t_i , t_k \geq c_\alpha \mid t_j < c_\alpha, \forall j \neq i, k)$	$\frac{1}{2}N(N-1)\alpha^2(1 - \alpha)^{N-2}$	2	0.08
\vdots	\vdots	\vdots	\vdots
$P(t_i < c_\alpha \mid t_j \geq c_\alpha, \forall i \neq j)$	$N\alpha^{(N-1)}(1 - \alpha)$	$N - 1$	0
$P(t_i \geq c_\alpha, \forall i = 1, \dots, N)$	α^N	N	0

Table 2.1: Rejection probabilities under the null.

The average number of null variables retained is given by:

$$k = \sum_{i=0}^N i \frac{N!}{i!(N-i)!} \alpha^i (1 - \alpha)^{N-i} = N\alpha \quad (2.5)$$

When $N = 100$, there are more than 10^{30} events (given 2^{100} possible test outcomes for the events in column 1), nevertheless $k = 0.5$ at $\alpha = 0.005$ so with $g = k/N$, then $g = \alpha$ and one irrelevant variable will be significant by

chance every second time such a decision rule is used. In practice, empirical values of g have a sampling distribution with a variance that depends on α , so is small when α is (see e.g., Johansen and Nielsen, 2009, 2016).

Formally the empirical gauge is the null retention frequency of selection statistics across M replications of a given DGP. Correspondingly, we define the empirical potency as the average non-null retention frequency, and will consider that shortly. For each variable, j , denote the retention rate by \tilde{r}_j , then when $1_{(|t_{\beta_{j,i}}| \geq c_\alpha)}$ denotes an indicator equal to unity (zero) when its argument is true (false), then for (2.3):

$$\begin{aligned} \text{retention rate: } \tilde{r}_j &= \frac{1}{M} \sum_{i=1}^M 1_{(|t_{\beta_{j,i}}| \geq c_\alpha)}, \quad j = 1, \dots, N, \\ \text{gauge: } g &= \frac{1}{N - n} \sum_{j=n+1}^N \tilde{r}_j, \\ \text{potency: } p &= \frac{1}{n} \sum_{j=1}^n \tilde{r}_j \end{aligned} \tag{2.6}$$

Although there will be considerable apparent ‘model uncertainty’ in (say) a Monte Carlo experiment with $N = 100$, $n = 0$ and $\alpha = 0.005$, because with a large number of replications, many hundreds of different models will be selected, such variation is essentially inconsequential since most will have just one or two irrelevant variables, and it does not matter which particular irrelevant variable(s) are adventitiously retained. Moreover, while it is not reflected in Table 2.1, as their null distributions will be Normal around zero, most chance significant selections will have $|t|$ -values close to c_α .

It may be thought that small values of α (so large values of c_α) will seriously reduce the chances of retaining relevant variables. Table 2.2 addresses this issue for a Normal distribution when $N = 500$, under the null that no variables matter for t-testing, using the Normal as the relevant baseline after taking account of outliers.

As can be seen, critical values increase slowly as α decreases and are just 2.85 at $\alpha = 0.005$. Using a ‘conventional’ 5% would lead to 25 variables being retained on average. However, doing so at such a loose significance level in any sample will lead to an underestimate of σ_ϵ^2 in (2.3), which can lead to more irrelevant variables being retained and hence serious overfitting. Consequently, we recommend setting $\alpha \leq \min[0.01, 1/N]$, which will increase c_α to at least

α	0.05	0.01	0.005	0.0025	0.001	0.0001
$k = N\alpha$	25	5	2.5	1.25	0.50	0.05
c_α	1.98	2.61	2.85	3.08	3.35	4.00
c_α^2	3.92	6.81	8.12	9.56	11.2	16.00

Table 2.2: Significance levels, retained variables, & critical values under the null for $N = 500$.

2.61, and here for $N = 500$ to $\alpha = 0.002$ so c_α to just over 3.

The main application of very tight α is when selecting with indicator saturation, in which case we set $\alpha \leq \min[0.01, 1/(N + qT)]$ for q indicator sets. For example, with super-saturation, $q = 2$, so $qT = 200$ at $T = 100$ (see e.g., Jonas Kurle, 2019, and §2.6.4) and if $N = 500$ then α could be rounded to 0.001. This formula for setting α also entails that it declines towards zero as $T \rightarrow \infty$ to ensure a consistent selection procedure for a GUM which nests a DGP that remains finite with constant parameters. In practice, with saturation estimation when $N \ll T$, we suggest retaining all N variables without selection when selecting indicators at a very tight α , and once the indicators are found, select over the $(N - n_1)$ variables plus those indicators at $\alpha = 1/(N - n_1)$ when n_1 theory variables are retained without selection. Chapter 6 illustrates this procedure. However, when selecting a forecasting device as discussed in Section 2.9, a much looser target α can be appropriate.

2.5.2 Probability of retaining relevant variables

The trade-off of using a very tight α like 0.005 (tight relative to 0.05) is not retaining relevant variables with estimated t-values between 1.98 to 3, so we now turn to the second key component, the potency, which is the probability of retaining the n relevant variables. Above we noted how to mitigate some of the costs of tight α s by retaining theory-entailed variables without selection. As gauge is not test size, potency is not test power given the selection context but can be evaluated against power to measure any costs from not knowing the DGP. The potency in (2.3) depends on the non-centralities, denoted ψ , of their t-tests relative to c_α . When $\psi_i = c_\alpha$ for a Normal distribution, there is a 50% chance that $t_i \geq c_\alpha$ in which case the i^{th} variable will be retained. For example, when $c_\alpha \approx 2$ there is just a 50-50 chance that a variable with $\psi = 2$ will be retained, as its t-test will be approximately Normally distributed

around a mean of 2. Similarly at $c_\alpha \approx 2.6$, a value of $\psi = 2.6$ will lead to retention half the time. It must be stressed that this example is not a problem of selection potency, but of test power, albeit such a conflation is often forgotten. If a variable would not be significant at the chosen c_α on a single t-test when the DGP is known, it cannot be expected to be selected at that critical value when the DGP is unknown. The approximate t-test powers when a false null hypothesis is tested once at different non-centralities are shown in Table 2.3, as well the potency when five such tests are conducted at the same non-centrality (based on a standard Normal distribution).

ψ	α	$\Pr(t \geq c_\alpha)$	$(\Pr(t \geq c_\alpha))^5$
2	0.05	0.50	0.031
2	0.01	0.27	0.001
3	0.01	0.65	0.116
4	0.01	0.92	0.651
4	0.005	0.87	0.513
5	0.01	0.992	0.959
6	0.001	0.996	0.980

Table 2.3: Approximate t-test powers at different non-centralities and critical values.

There is a 50–50 chance of retaining one variable with $\psi = 2$ for $c_\alpha = 2$, but only a 3% chance of finding 5 such variables significant in any trial. This is a probability statement, not a problem due to ‘repeated testing’ or selection. As Table 2.3 shows, the power rises non-linearly with ψ , and importantly, when it is very high, as in the bottom two rows, it remains high even for 5 variables.

The potency \mathfrak{p} is then the average rejection frequency of a false null hypothesis when selecting at a given significance level. As shown in a number of simulation studies, the potency is close to the average power for models like (2.3): see e.g., Castle *et al.* (2011) and Hendry and Doornik (2014). Since problems confronting selection potency, such as non-orthogonality, also confront power, potency remains close to average power more generally.

Combining these two key components, with $g \approx \alpha$ and potency close to power, we can see that commencing selection at the same significance level from the DGP (i.e., with no irrelevant variables) or from a GUM that nests that DGP, a similar model will be selected, with on average $g(N - n)$ additional

irrelevant variables in the latter. Of course this is no guarantee that k will equal n , as non-centralities of ‘relevant variables’, defined by having $\beta_i \neq 0$, may be too small to lead to ‘significant’ outcomes even if the DGP is estimated.

The principles behind the approach just described apply to non-orthogonal models, although 1-cut is not a sensible strategy in such a setting. Instead, the software *Autometrics* uses multi-path block searches to order the underlying values of τ_j^2 taking account of their intercorrelations (see Doornik, 2009). *Autometrics* has two additional features to maintain its effectiveness as noted earlier. First, given that the feasible estimated GUM is congruent on the desired mis-specification tests, the same tests are applied during path searches as diagnostic checks so that paths are not followed if deleting variables makes any of those tests significant. This is a logical requirement since the LDGP must be a congruent representation of itself, and consequently a non-congruent model cannot be the LDGP. Similarly, when $N \ll T$, *Autometrics* checks that the current selection encompasses the GUM so no salient information is lost (see Doornik, 2008, and Bontemps and Mizon, 2008). These two checks on potential information losses play an important role in helping *Autometrics* select appropriate models.

However, while all retained variables must be significant by design at c_α in 1-cut, that will not necessarily occur with automated *Gets*. In *Autometrics*, variables may also be retained because of the roles played by:

- (a) diagnostic checks—when a variable is insignificant, but its deletion makes a diagnostic test significant;
- (b) encompassing tests—a variable can be individually insignificant, but not jointly with all variables deleted so far.

Thus, the gauge may be larger than α if retained variables are irrelevant, but equally the potency can also exceed average power. When $N > T$, both expanding and contracting block searches are needed, and as this inevitably arises for indicator saturation estimators, we now consider that setting.

2.6 Selecting models with saturation estimation

The approach to tackling outliers that we use, called impulse-indicator saturation (denoted IIS), was accidentally discovered by Hendry (1999), then developed by Hendry *et al.* (2008) for a special case, and analyzed more generally by Johansen and Nielsen (2009, 2016). IIS is an integral part of the

general model selection algorithm that not only selects over the relevance of explanatory variables, lag lengths and non-linearities, but also over observations during selection, all jointly—even though that creates more candidate variables in total (here denoted N) than observations T , so $N > T$. By selecting over observations, IIS is a robust-statistical device, as shown by Johansen and Nielsen (2016). One aspect of commencing from (2.2) is to achieve robustness against a range of potential mis-specifications, an issue discussed in Castle *et al.* (2019c), since the costs of selection are small compared to those of mis-specification. Variables deemed relevant from prior reasoning can be retained without selection while selecting over other aspects at a tight significance level as in Hendry and Johansen (2015).

The simplest explanation of how model selection works for IIS (and hence in general when $N > T$), is to consider an independent, identically-distributed Normal random variable with constant mean μ and variance σ^2 denoted $y_t \sim \text{IN}[\mu, \sigma^2]$ for $t = 1, \dots, T$. There are no genuine outliers, but random samples will still on average deliver observations that lie outside $\pm 2\sigma$ with the appropriate probabilities (e.g., 5% for that range, and 1% for $\pm 2.6\sigma$). Consider an investigator who seeks to check for potential outliers in the data process, as in Hendry *et al.* (2008). Create a complete set of impulse indicators $\{1_{\{t\}}\}, t = 1, \dots, T$, so there is one for every observation, which is why the approach is called saturation. Set a nominal significance level α , such that $\alpha T \leq 1$, and its associated critical value c_α . Then, under the null hypothesis of no outliers, $\alpha T \leq 1$ impulse indicators should be significant on average by chance, entailing that number of observations will be removed adventitiously.

Of course, the complete set of impulse indicators cannot sensibly be included because a perfect fit would always result, so nothing would be learned. Although *Autometrics* considers many splits and does multiple searches across different splits, consider implementing a simple ‘split-half’ version of the IIS procedure, namely add half the indicators, select then repeat for the other half of the data, which is what inadvertently occurred in Hendry (1999). Taking T as even for convenience, regress y_t on a constant and the first half of the indicators, $\{1_{\{t\}}\}, t = 1, \dots, T/2$. Doing so ‘dummies out’ the first half of the data set, so is in fact just a standard regression on an intercept using only the second half of the sample, which delivers unbiased estimates of μ and σ^2 . However, it also provides estimates of the impulse indicators in the first half, and any with significant t-statistics are recorded. Each individual t-test is a

standard statistical test under the null hypothesis, with $\Pr(|t_{1\{t\}}| \geq c_\alpha) = \alpha$. Impulse indicators are mutually orthogonal, so their estimated coefficients are unaffected by eliminating insignificant indicators.

Drop all the first half indicators, and now regress y_t on the intercept and the second half of the indicators. Again, this delivers unbiased estimates of μ and σ^2 , now from the first half, and reveals any significant indicators in the second half. Finally, regress y_t on the intercept and all the sub-block significant indicators using the full sample, and select only those indicators that remain significant at c_α . Overall the algorithm solves a variant of a multiple testing problem for more candidate regressors than observations, where we control the average number of ‘false outlier’ discoveries, or gauge as defined above and analyzed by Johansen and Nielsen (2016).

This split-half algorithm is merely an expository device, and as shown under the null by Hendry *et al.* (2008) the same outcome will be found using more blocks which can be of different sizes. The main reason for considering split-half analyses is that they can be undertaken analytically to establish the feasibility of saturation estimation, and as shown in Castle *et al.* (2019b), reveal the structure of different specifications of indicators. No ‘repeated testing’ is involved by exploring many choices of blocks, as an impulse indicator will only be significant at the preset α if there is a large residual at that observation, with the *caveat* we noted above that $\alpha \leq 1/T$. Loose significance levels can lead to overfitting as selecting more irrelevant indicators than one or two will result in underestimating σ^2 , leading to more indicators being adventitiously selected, unless $\hat{\sigma}^2$ is bias corrected as in Johansen and Nielsen (2009).

As described in Doornik (2009), the operational approach in *Autometrics* uses a multi-path block search. The whole sample is divided into blocks usually smaller than about $T/4$, selection at α is applied within blocks, and any significant indicators found in each block are added to the current selection until: (i) the resulting estimated equation does not fail any of the diagnostic tests; and (ii) none of the as-yet-unselected indicators are significant if added, which thereby creates a terminal model. Then different block mixes are created and a new multi-path block search is commenced from these different partitions until there are no new directions to explore. In the more general case of regressors, the same algorithm applies, and if several terminal models are found, their variables are combined, and selection recommenced till an undominated model is chosen. If several terminal models remain, an information

criterion can choose between terminal models that are mutually undominated (i.e., mutually encompassing as in Bontemps and Mizon, 2008).

Although creating a set of candidate variables that exceeds the number of available observations with impulse indicators for every observation may seem unlikely to be a successful approach to model selection, we must stress that many well-known and widely-used statistical procedures are in fact variants of IIS: rather like we all speak prose even if we don't know it. Examples are: (a) recursive estimation, which is equivalent to including impulse indicators for the 'future' observations, then sequentially dropping those indicators as the algorithm moves through the sample;

(b) also, moving windows essentially uses IIS on 'pre' and 'post' data;

(c) and 'holding back data' (e.g., to control size) is equivalent to IIS applied to the excluded data points;

(d) as is the less acceptable process of prior sample selection by excluding data like war years, or ending a sample early to avoid including shifts, etc., which also implicitly uses IIS on the omitted observations.

(a) and (b) can detect outliers or shifts when the model is otherwise known, but even so, ignore the information on the unused data. Importantly, IIS can be included as part of the general selection process when the model specification has to be discovered from the evidence. Moreover, as shown by David Salkever (1976), the Gregory Chow (1960) test can be implemented by including an appropriate set of impulse indicators, so is essentially IIS restricted to a specific subset of data for a given model.

2.6.1 Illustrating impulse-indicator saturation

We now illustrate the application of split-half selection to impulse-indicator saturation when there are no outliers, then when there is one, using 12 observations on the simple time series:

$$y_t = 5 + \epsilon_t, \quad \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2], \quad (2.7)$$

where $\sigma_\epsilon^2 = 1$, with an outlier of magnitude $\delta = -2$ at $t = 8$ in the second setting. To implement split-half selection, we create all 12 impulse indicators, then add the first 6 to a regression of y_t on a constant to see if any have $|t|$ -statistics exceeding the critical value c_α . If so, these are recorded and the first 6 are replaced by the second 6, and the process repeated. Finally, all

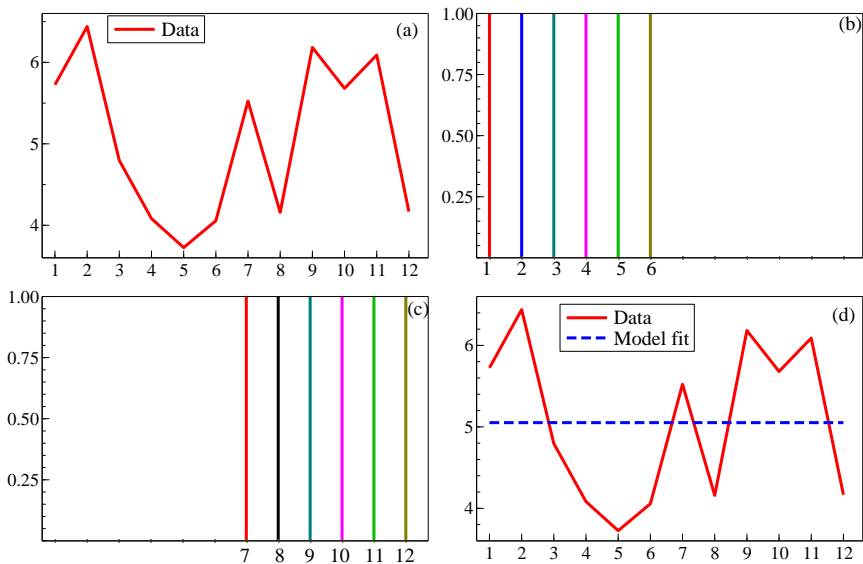


Figure 2.1: (a) Time series. (b) First 6 of the impulse indicators. None significant at $\alpha = 0.05$. (c) Second 6 indicators: again none significant. (d) Final outcome, retaining just the intercept.

selected indicators are included and their $|t|$ -values checked, and only those still greater than c_α are retained.

Figure 2.1 records the split-half sequence. We chose $\alpha = 0.05$ as the theoretical gauge so $\alpha T = 0.6$, suggesting less than one impulse will be retained on average by chance. However, the empirical gauge has an approximate standard deviation of 0.063 at $T = 12$, as shown in Johansen and Nielsen (2016), so could be a bit larger or somewhat smaller. With $T = 12$, no null indicators are likely to be selected at $\alpha = 0.01$, so although that would still be a feasible choice, we felt it might be thought to bias the procedure in favor of finding no irrelevant indicators.

We actually compared four closely related approaches, namely the split half, checking which indicators are significant; *Autometrics* applied to each half with their indicators included; commencing with all 12 indicators in the candidate information set and selecting by *Autometrics* general algorithm; and *Autometrics* pre-programmed IIS algorithm. Panel (a) shows the original (uncontaminated) data; (b) the first six impulse indicators, none of which is

significant at $\alpha = 0.05$ when added, and none is retained when applying *Autometrics* to that specification, where the constant is not selected over; (c) the second six, where again none is significant—and none retained by *Autometrics*; and (d) the model fit, which is just the mean of y .

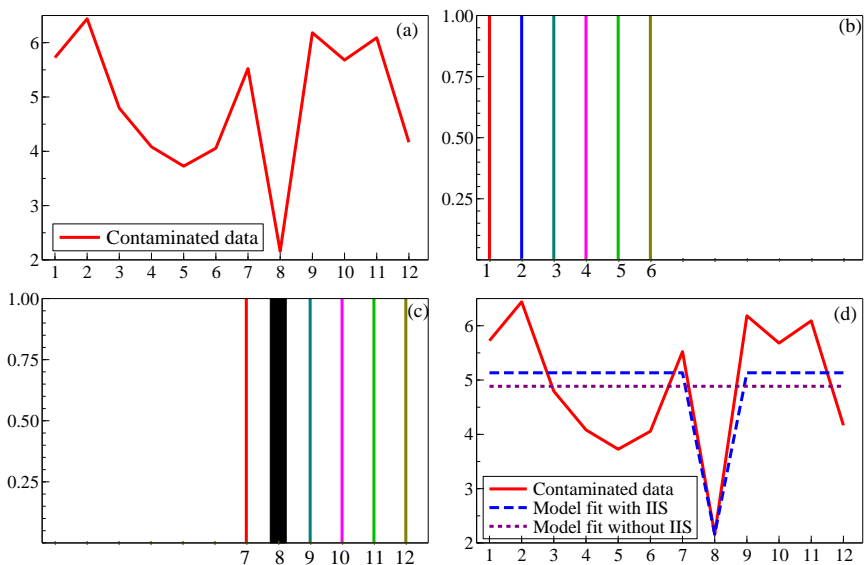


Figure 2.2: (a) Time series with an outlier of $\delta = -2$ at $t = 8$. (b) First 6 of the impulse indicators: none significant at $\alpha = 0.05$. (c) Second 6 indicators: now one is retained, shown as wide. (d) Outcome with and without that indicator.

The third approach is to add all 12 indicators to make an (infeasible) GUM and apply *Autometrics* to that set of candidate variables, which here also delivers the null model, as does the fourth, using the pre-programmed IIS algorithm in *Autometrics*, which commences from a model like (2.7) and automatically creates the impulse indicators if IIS is chosen as the estimator. All four ways are applied below to other saturation estimators, with the health warning that the first two are just a convenient form for analysis and need not work for more complicated DGPs: the third and fourth deliver the same outcomes at the same α , but with the benefit that the pre-programmed IIS saves creating the indicators.

Figure 2.2(a) illustrates a DGP when there is an outlier denoted $\delta = -2$

at $t = 8$. Figure 2.2(b) shows the first 6 indicators, and as before none is significant and none retained by *Autometrics* applied to that specification. But when the second 6 are added as regressors, $1_{\{8\}}$ is significant, corresponding to the smallest observed data value, and is also retained by *Autometrics* for that specification with $t = -2.87$, $p = 0.016$. The intercept is then estimated as 5.13 ($SE = 0.30$) as opposed to 4.89 ($SE = 0.37$) if the outlier is ignored. When the IIS option in *Autometrics* is used for a model of y_t on a constant, the same indicator, $1_{\{8\}}$, is selected. As $\alpha = 0.05$, the potency given by $\Pr(|t_\delta| \geq c_\alpha)$ is approximately 0.5, so $1_{\{8\}}$ will be significant about half the time, even if the correct indicator was added at that observation. At $\alpha = 0.01$, an outlier of $\delta = 2$ would be found on average about a $1/4$ of the time even if its precise location was known.

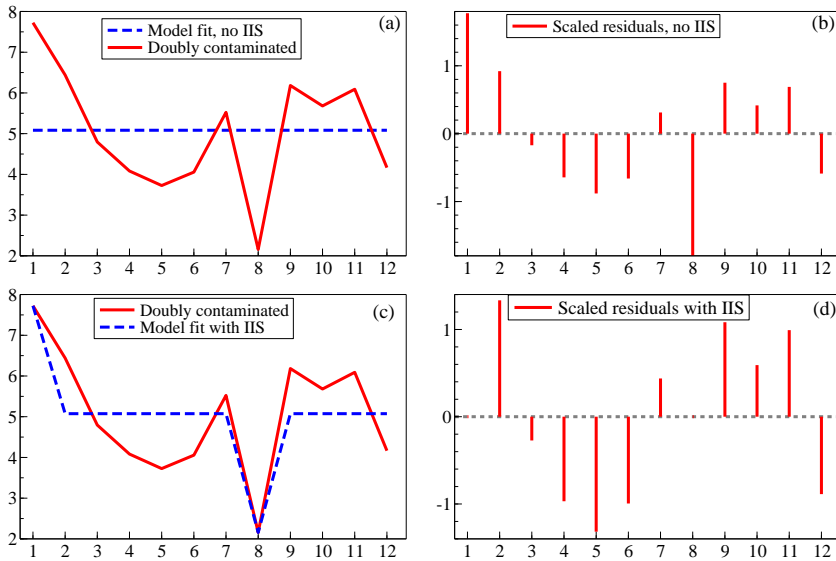


Figure 2.3: (a) Time series with two outliers at $t = 1$ and $t = 8$ and the model fit without IIS. (b) Residuals scaled by $\hat{\sigma}_\epsilon$. (c) Time series and the model fit with IIS. (d) Residuals scaled by $\tilde{\sigma}_\epsilon$ (after IIS).

So far there is just one outlier and it is relatively small. Having two outliers in a sample of $T = 12$ would be fairly bad contamination ($> 15\%$), and their detectability depends on their magnitudes, signs and the particular values

of the observations at which they happen. Setting $\delta = 2$ at $t = 1$, so the first observation is incorrectly recorded, creates a doubly contaminated data set. Although it is not obvious that there are 2 outliers from the data plot or the scaled residuals in Figure 2.3(b), with one affecting the first observation (which would be excluded by some approaches), IIS selects both $1_{\{1\}}$ and $1_{\{8\}}$. However, the split-half approach selects neither. This occurs because one of the outliers lies in each half, so contaminates each baseline half in turn. Such a problem applies even more forcibly to 1-step searches, and emphasises the benefits of multi-path block searches: see e.g., the discussion between Edward Gamber and Jeffrey Liebner (2017) and Ericsson (2017b).

IIS can also be beneficial when the underlying error distribution is fat-tailed, as shown in Castle *et al.* (2012) where the errors have a t_3 distribution. IIS ‘removes’ many of what would be classified as outliers relative to a Normal and so can reduce the mean square error (MSE) of the resulting model parameter estimates around their population values. Also Hendry and Santos (2010) apply IIS to testing for parameter invariance under changes in the distributions of the unmodeled (weakly exogenous) variables, and show that impulse indicators may also detect heteroskedasticity. Finally, IIS can be useful for detecting forecast biases as in Ericsson (2017a).

2.6.2 Properties of impulse-indicator saturation

When any additional regressor variables are included in the model of y_t , and are retained without selection, Johansen and Nielsen (2009) show that the usual rates of estimator convergence to their asymptotic distribution (namely \sqrt{T} under stationarity) are unaffected by IIS (despite selecting from more candidate variables than observations) with a loss of estimation efficiency dependent on the error distribution and the choice of α . When the error $\{\epsilon_t\}$ is distributed symmetrically with no outliers, as with a Normal distribution, applying IIS to a regression with n variables \mathbf{x}_t (retained without selection), constant parameter vector β and data second moment matrix Σ , with $\mathbb{E}[\mathbf{x}'_t \epsilon_t] = \mathbf{0}$ and $t = 1, \dots, T$ for the model:

$$y_t = \beta' \mathbf{x}_t + \epsilon_t, \text{ where } \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2], \quad (2.8)$$

the limiting distribution of the IIS estimator $\tilde{\beta}$ converges to β at the usual rate of \sqrt{T} , where orthogonality is not required. That limiting distribution is

Normal, with a variance that is somewhat larger than $\sigma_\epsilon^2 \Sigma^{-1}$ so:

$$\sqrt{T} (\tilde{\beta} - \beta) \xrightarrow{D} N_n \left[\mathbf{0}, \sigma_\epsilon^2 \Sigma^{-1} \Omega_\alpha \right]. \quad (2.9)$$

The efficiency of $\tilde{\beta}$ with respect to the usual (and here valid) ordinary least-squares (OLS) estimator $\hat{\beta}$, as measured by Ω_α , depends on c_α and the error distribution, but is close to $(1 - \alpha)^{-1} \mathbf{I}_n$ for small α : see Johansen and Nielsen (2009). Consequently, despite adding T irrelevant impulse indicators to the candidates for selection, there is a very small cost: for $\alpha = 1/T$, on average from (2.5) just one observation will be ‘lost’ by being dummied out.

Conversely, there is the potential for major gains under the alternatives of data contamination and/or breaks, and as noted IIS can be done jointly with all other selections. Under the alternative that there are one or more outliers, IIS locates the ones with t-statistics exceeding c_α in absolute value. Estimates of impulse indicators cannot be consistent for outliers as there is only ever a single observation from which to estimate them, but if embedded in a process where the error variance tended to zero (small sigma asymptotics), would be selected with probability approaching unity. In practice, what matters is ensuring that the estimates of the parameters of interest in the modeling exercise are robust to the presence of a moderate number of relatively large outliers, and simulations suggest this occurs except where the outliers are ‘evenly’ spread through the data so most sub-samples ‘look alike’.

The simplest setting of an impulse indicator model can be useful when teaching test power, relating back to §2.5.2. Let:

$$y_t = \delta 1_{\{\tau\}} + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2] \quad (2.10)$$

where $1 \leq \tau \leq T$ is known, and $\delta = 2\sigma_\epsilon$. Then $\hat{\delta} = \delta + \epsilon_\tau < 2\sigma_\epsilon$ whenever $\epsilon_\tau < 0$ which has a probability of $\frac{1}{2}$, so will not be significant at $c_\alpha = 2$. Moreover, even for $\delta = 3\sigma_\epsilon$, then $\hat{\delta}$ will not be significant when $\epsilon_\tau < -\sigma_\epsilon$ which has a probability of about 16%; and so on.

In the context of testing for parameter invariance, Hendry and Santos (2010) derive the powers of individual impulse-indicator tests, which depend on the magnitude of the outlier they detect relative to σ_ϵ . Hendry and Mizon (2011) demonstrate the dangers of not tackling outliers in empirical models, as a failure to handle large outliers can lead to the rejection of even a sound theory—in that instance because a positive price elasticity for the demand

for food results when estimating without IIS. Castle *et al.* (2019c) show that IIS can shed insight into the long-standing problem noted by Thomas Hettmansperger and Simon Sheather (1992) when using ‘conventional’ robust-statistical estimators like least median squares (LMS: see Peter Rousseeuw, 1984), and least trimmed squares (LTS: see Rousseeuw, 1984, and Jan Věšek, 1999), which transpire not to be robust to a small accidental measurement error created when inputting the data. In the general context of selecting variables, lags and non-linear functions along with IIS, Hendry and Doornik (2014) show in simulations that *Autometrics* has an empirical gauge close to α for small α under the null hypothesis of no additional relevant variables, and also describe its ability to select relevant variables under the alternative.

2.6.3 Step-indicator saturation

Step-indicator saturation (SIS) uses steps rather than impulses to capture permanent shifts in the location of a relationship: see Castle *et al.* (2015b). Step indicators are cumulations of impulse indicators up to the given date, so terminate with the last observation for the date shown: e.g., $S_{\{1960\}}$ is unity till 1960 then zero thereafter. Consequently, changes in steps, such as $\Delta S_{\{1960\}} = S_{\{1960\}} - S_{\{1961\}}$, correspond to $-1_{\{1961\}}$. Castle, Hendry, and Andrew Martinez (2017) apply SIS to invariance testing, discussed below.

Much of the analysis of SIS is similar to that for IIS above, even though the steps are not orthogonal. Nielsen and Matthais Qian (2018) provide an analysis of the asymptotic properties of the empirical gauge of SIS. While showing that its distribution is more complicated than that of IIS in Johansen and Nielsen (2016) described above, it converges to the nominal significance level in a range of models with stationary, random walk and deterministically trending regressors. They conclude the gauge of SIS can be reliably used at a tight significance level in applications. While there is no formal analysis as yet, a result similar to that in Johansen and Nielsen (2009) seems likely to hold for the distribution of estimated retained regression parameters when SIS is applied at a suitably stringent level under the null of no step shifts. Under the null, SIS has a somewhat larger empirical gauge than IIS as it retains two successive roughly equal magnitude, opposite-signed step indicators to capture an outlier, although that could be adjusted manually. Like IIS, coefficients of step indicators will not be consistently estimated when new steps keep

occurring, but would be in a scenario of a fixed number of steps that are all proportional to $T \rightarrow \infty$.

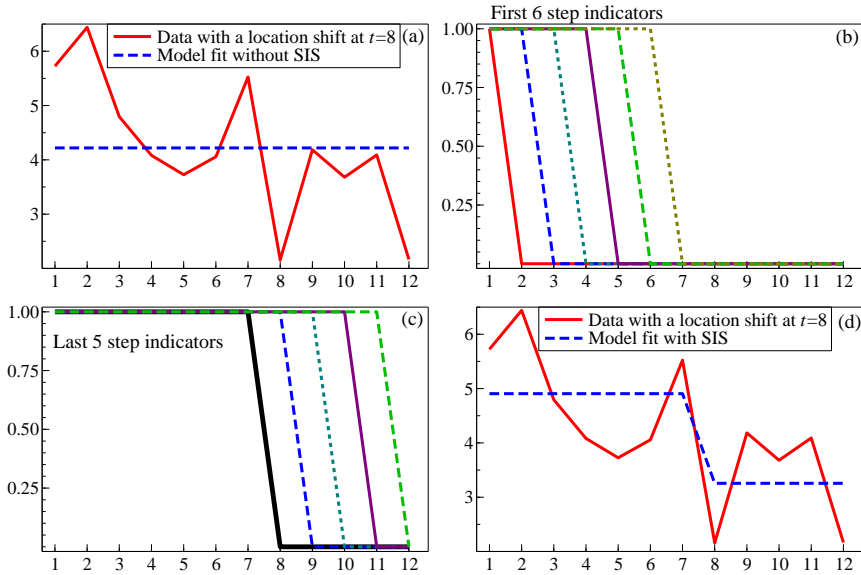


Figure 2.4: (a) Time series with a location shift. (b) Add the first 6 of the step indicators: none significant, nor selected by *Autometrics* at $\alpha = 0.05$. (c) Add the other 5 steps: $S_{\{7\}}$ (thick solid line) significant and also retained by *Autometrics*. (d) Outcome with SIS.

We now illustrate SIS for handling shifts of distributions when there is a location shift in (2.7) of $\delta = -2$ starting at $t = 8$. Only $T - 1$ step indicators are needed as the T^{th} step coincides with the constant term. Figure 2.4(a) shows the resulting data, together with the fit that would be obtained if SIS was not used; (b) the first 6 step indicators, none of which is retained; (c) the remaining 5, where $S_{\{7\}}$ is significant on the split half, and also retained by *Autometrics* at $\alpha = 0.05$ applied to that specification (as well as by SIS at $\alpha = 0.05$); and the model's fitted values. The equation standard error $\hat{\sigma}_\epsilon$ is 1.28 without SIS and 1.02 with. Figure 2.5 shows the resulting 1-step ahead forecasts with and without SIS: the former is both more accurate and more precise. Although this outcome certainly favors SIS as expected because the DGP does have a location shift, the illustration did not use knowledge of that, nor of the timing, sign and magnitude of the shift.

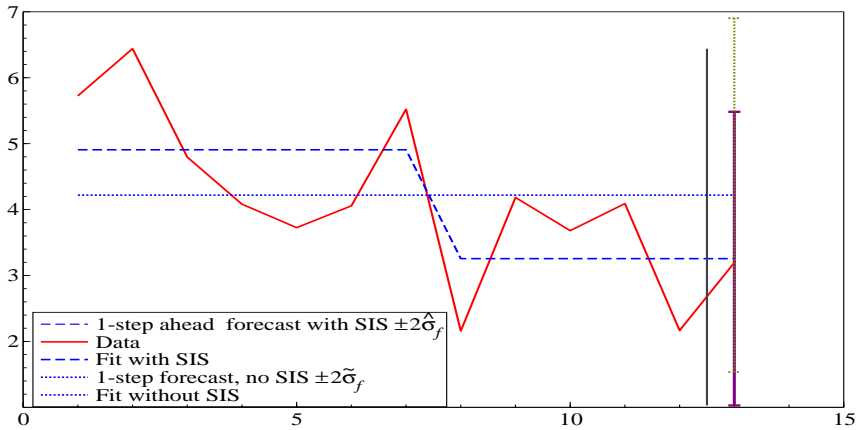


Figure 2.5: 1-step ahead forecast $\pm 2\hat{\sigma}_f$ with SIS (solid error bar) and without $\pm 2\tilde{\sigma}_f$ (dotted).

2.6.4 Super saturation estimation

Combining IIS and SIS leads to super saturation, named by Ericsson (2012), which can be helpful when there is a mix of outliers and location shifts. Kurlle (2019) undertook a detailed simulation study of super saturation and found that the gauge was proportional to the total number of indicators (roughly $2(T - 1)$) so the target α needed to be tight and about half that for either alone to achieve an appropriate gauge and avoid ‘overfitting’ by selecting at too loose a significance level. For example, at $\alpha = 0.005$ for $T = 100$, the gauge was around 0.01, both under the null of an IID process, and for retaining irrelevant indicators when relevant ones were found. We use super saturation when modeling the UK’s CO₂ emissions since 1860 as both outliers (e.g., from the General Strike of 1926) and location shifts (e.g., from the introduction of natural gas in place of coal gas starting in 1969) occur.

2.6.5 Trend saturation estimation

Deterministic linear trends are found in a number of empirical settings, but are unlikely to maintain the same growth rate over long periods. For example, annual UK productivity since 1860 (defined as output per worker per year at constant prices), shown in Figure 2.6(a) on a log scale, exemplifies such trend

shifts. An overall trend line completely fails to characterize the evidence: early periods of below average growth are above the line; the boom during World War I (WWI) is below and the dramatic increase in the growth rate after World War II (WWII) is far below. The figure also shows the fit of six separate trends to roughly 25-year subperiods: these obviously provide a far better description, but are both arbitrary and occasionally do not match changes in trend growth. This is most obvious at the end of the period, shown by the ellipse, where measured productivity has flat-lined since the 2008 ‘Great Recession’.

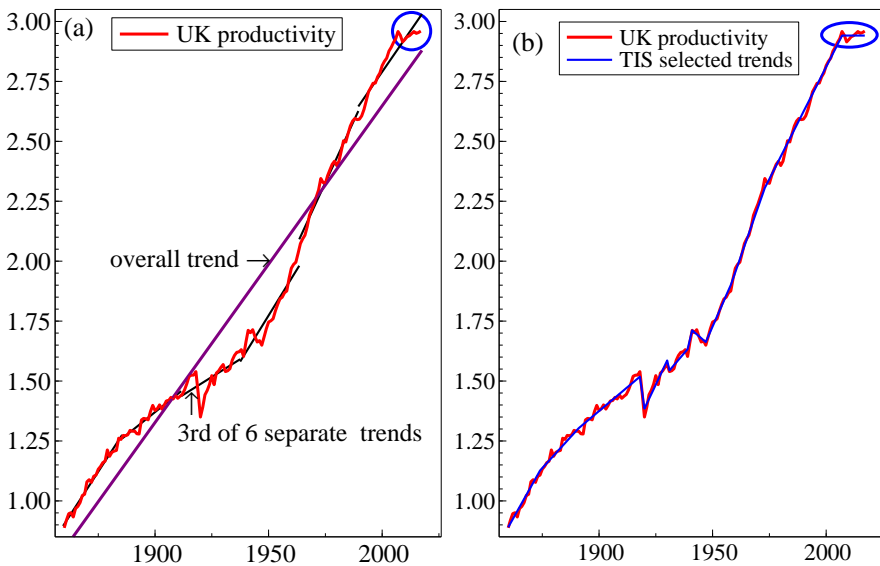


Figure 2.6: (a) UK productivity since 1860 with overall and six trend lines at roughly 25-year subperiods; (b) UK productivity since 1860 with TIS selected trend lines at $\alpha = 0.0001$.

Trend indicator saturation (TIS: see Castle *et al.*, 2019d, for an analysis), was applied at $\alpha = 0.0001$ to compensate for larger critical values in non-stationary processes and for non-IID errors. Figure 2.6(b) shows the outcome after manual deletion of trend indicators that were selected to avoid failures of congruence: because *Autometrics* implements a range of mis-specification tests, when a model is under-specified, as here, indicators can be retained that do not satisfy the target significance level. These were deleted by the authors

given the aim of the present illustration. Even so, 12 trend indicators remained (although an overall trend was not retained). As with super-saturation, TIS could be combined with IIS or SIS, appropriately adjusting α .

We noted above that two successive step indicators can capture an outlier, and as can be seen in Figure 2.6(b), several trend indicators combined can act to correct outliers and location shifts rather than trend breaks as such. It is well-known that second differencing removes linear trends, changes breaks in trends to impulses and converts location shifts to ‘blips’. Combining the near equal magnitude, opposite-sign trend indicators found for 1917–1919, 1929–1930 and 1938–1940, left 9 trend breaks, namely at the earlier of each of those three dates and at 1873, 1890, 1946, 1957, 1972, and 2006, noting that these times are when the trend indicators end. Many of these are salient dates: in historical order, 1873 was the start of the UK’s ‘Great Depression’ which ended gradually by 1896; 1919 was just after the end of WWI (where 1917 had been the peak output year), and also the start of the worldwide flu’ pandemic; 1930 followed the US stock market crash of 1929, and the onset of the Great Depression; 1940 was the start of mass production for WWII and 1946 the end of war production; 1972 saw the end of the ‘post-war recovery’ growth in productivity, as 1973 was hit by the first Oil Crisis leading to the UK’s high inflation and industrial troubles; and 2006 was the end of a long period of growth, when the flat-lining of productivity starts with the Great Recession of 2008. However, we are not aware of conspicuous events in 1890 or 1957, although these had the least significant estimated coefficients.

To interpret what TIS finds, imagine you are in 1874 with data on productivity over 1860–1873 and fit a linear trend. Then that will be the same as the one selected by TIS from the whole sample if the trend rate of growth then changed in 1874. Now move forward to 1891 with data from 1860–1890 and fit that first trend line jointly with one over 1860–1890: the first is replicated as the sum of the trend coefficients and the second reveals a shift. Thus, TIS replicates what the historical record would have shown at the time on the available subsamples of data. That contrasts with fitting an overall trend to the full sample, which is not what would have been found at the time, and hence distorts the historical record—as we have just discussed.

A perhaps surprising application of TIS is to health care, studying the rates of adoption of Desogesterol to replace Cerazette (a synthetic progestogen) at all 213 UK National Health Service (NHS) Clinical Commissioning Groups

(CCGs). Alex Walker, Pretis, Anna Powell-Smith and Ben Goldacre (2019) applied TIS to analyze the heterogeneity in the extent, time, and speed of diffusion of innovations as measured by the prescribing behaviour of the CCGs, and the resulting additional costs to the healthcare system of slow adoption of the replacement pharmaceutical (note the favorable editorial).

2.6.6 Multiplicative-indicator saturation for parameter changes

Multiplicative-indicator saturation (MIS) focuses on changes in the parameters of variables in models. In MIS, every variable is multiplied by a complete set of step indicators: see e.g., Oleg Kitov and Morten Tabor (2015) and Castle *et al.* (2017) for simulations and applications, and Castle *et al.* (2019b) for an analysis of its properties. Thus, SIS can be interpreted as MIS for the constant term, but would have little potency facing changes in β when that was the coefficient of (say) $(x_t - \bar{x})$ where \bar{x} denotes the mean of x . Indeed, Michael Clements and Hendry (1998) show that zero-mean changes have little impact on forecasts, and that forecast failure is primarily caused by direct or induced shifts in the long-run mean (an issue discussed further in Section 2.9). Nevertheless, MIS could detect such parameter changes. To see why, consider knowing that T_1 was the shift date: fitting separate models up to and after T_1 will naturally deliver estimates of the different subsample parameters. Using MIS to select that split should find the correct indicator, or one close to it, despite the large number of interactive indicators, albeit with added variability. Searching through many initial candidate variables efficiently is essential here as the number of regressors increases rapidly if testing for non-constancy in the parameters of many variables as each variable is multiplied by $T - 1$ step indicators.

To illustrate MIS, we consider a generalization of (2.7) adding a regressor $\{x_t\}$ with a mean of zero, first when its coefficient is constant with $\beta_1 = 10$, and then when β_1 shifts to 5 at observation t , which date is revealed below, with $\beta_0 = 5$ throughout, so under the null:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad \epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]. \quad (2.11)$$

To create the GUM to implement MIS, we add to 1 and x_t the candidate set of

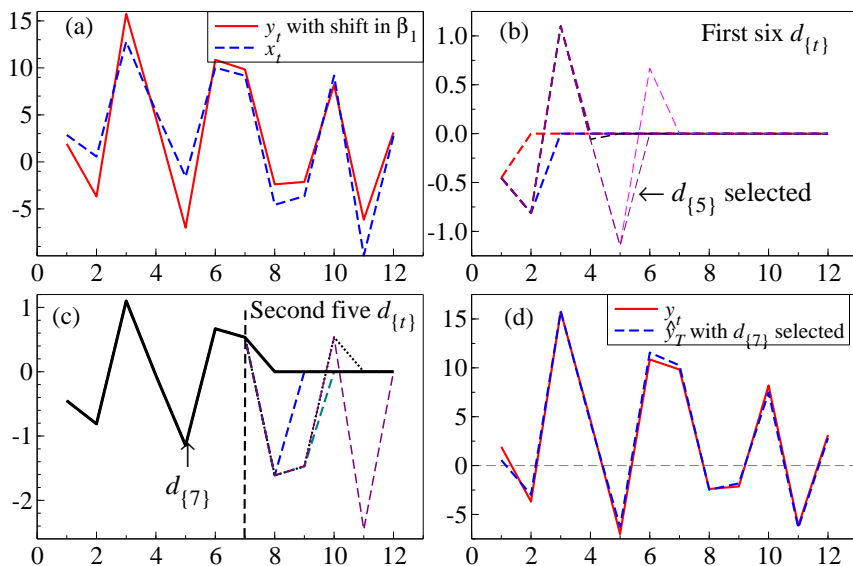


Figure 2.7: (a) Time series with a change in parameter β_1 of $\delta = -5$ hitting at $t = 7$; (b) Six multiplicative indicators: $d_{\{5\}}$ is significant at $\alpha = 0.01$; (c) Next five multiplicative indicators: $d_{\{7\}}$ is selected. (d) Add $d_{\{5\}}$ & $d_{\{7\}}$ with the outcome that only $d_{\{7\}}$ is selected.

11 variables $S_{\{j \leq t\}} \times x_t = d_{\{j\}}$, $j = 1, \dots, T - 1$ noting that $S_T = 1$:

$$y_t = \beta_0 + \beta_{1,0}x_t + \sum_{j=1}^{T-1} \beta_{1,j}d_{\{j\}} + e_t, \quad e_t \sim \text{IN}[0, \sigma_e^2], \quad (2.12)$$

and apply the usual split-half approach, where no $d_{\{j\}}$ are significant in either half, nor are any when applying the general search procedure in *Autometrics* to (2.12) at $\alpha = 0.01$, in both cases retaining 1 and x_t without selection.

Next, Figure 2.7 records the situation when β_1 shifts to $(\beta_1 - 5)$ using the ‘split-half’ approach. Panel (a) shows y_t and x_t where the plot scales the data such that both y_t and x_t have the same mean and range, which here is equivalent to their regression. Panel (b) reports the first six multiplicative indicators $d_{\{1\}} \dots d_{\{6\}}$ of which $d_{\{5\}}$ is significant. It can be proved that the indicator closest to the shift is the most likely to be retained when the shift lies in the other half. Panel (c) records the next five multiplicative indicators $d_{\{7\}} \dots d_{\{11\}}$ where $d_{\{7\}}$ is significant. Including the two significant indicators,

only $d_{\{7\}}$, is selected leading to the improved outcome in Panel (d), which is precisely what *Autometrics* finds.

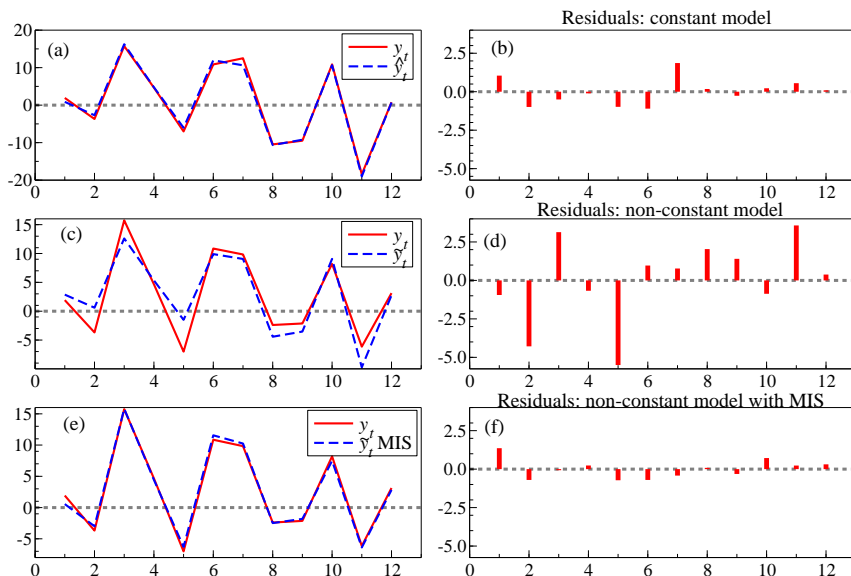


Figure 2.8: (a) Time series with model fit when constant parameters; (b) Residuals from (a); (c) Time series and model fit when non-constant parameters; (d) Residuals from (c); (e) Time series and model fit with MIS when non-constant parameters; (f) Residuals from (e): all residuals on same scale.

Figure 2.8(a) shows the data with the model's fitted values for constant parameters in (2.11) and Panel (b) records the residuals, which are almost the same as the DGP errors. The non-constant parameter graphs shown in Figures 2.8(c) and (d) are where β was halved to 5 at a point in the sample. It is not immediately obvious from ocular econometrics (looking at the data) that the poorer fit in (c) is due to a major shift in the parameter of the regressor, or precisely where that occurred: $t = 5$ looks suspect but is not the switch date. Finally, Figures 2.8(e) and (f) provide the corresponding graphs after MIS, where $d_{\{7\}}$ was correctly selected. Fitting separately to $t = 1, \dots, 6$ and $t = 7, \dots, 12$ delivers almost the same parameter estimates as MIS.²

²Here, recursive graphs do in fact reveal the problem of parameter non-constancy.

Applying MIS delivers, noting that $S_{\{12\}}x_t = x_t$:

$$\hat{y}_t = \underset{(0.58)}{4.89} S_{\{6\}}x_t + \underset{(0.31)}{4.89} S_{\{12\}}x_t + \underset{(0.31)}{5.34} \quad (2.13)$$

revealing the shift was at $t = 7$ and was from $\beta_1 = 10$ to $\beta_1 = 5$. If MIS+SIS is undertaken for $T = 12$, despite having 24 variables, the same equation results, since the intercept is constant.

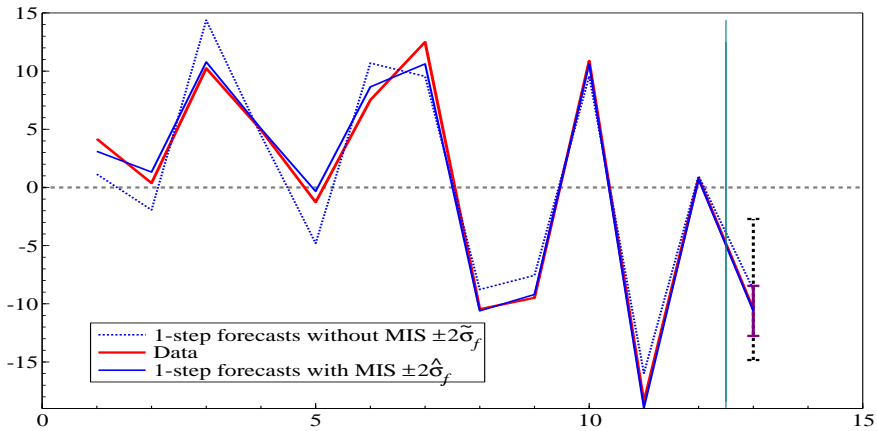


Figure 2.9: 1-step ahead forecast $\pm 2\hat{\sigma}_f$ with MIS (solid error bar) and $\pm 2\tilde{\sigma}_f$ without (dotted error bar).

The fit and residuals from (2.13) are essentially the same as Figure 2.8(a) & (b) despite the non-constancy, because the shift in the parameter is correctly ‘picked up’ by MIS. Since x_t has a mean of zero and the intercept is constant, there is little difference between the 1-step ahead forecasts with and without MIS, but a large difference in the interval forecasts as Figure 2.9 records.

2.6.7 Designed-indicator saturation

Indicators for saturation approaches can be designed to match known properties of a physical or social process under analysis, denoted VIS for volcano-indicator saturation as applied to volcanic impacts in Pretis, Lea Schneider, Jason Smerdon, and Hendry (2016). Their indicators have the form $\mathbf{d}'_t = (0, \dots, 0, 1, 0.5, 0.25, 0.125, 0, \dots, 0)$ for $t = 1, \dots, T$, so the first commences with $(1, 0.5, 0.25, 0.125, 0, \dots)$ and so on. A saturating set of such

indicators were selected over to model the impacts of volcanic eruptions on dendrochronological temperature reconstructions as follows.

Mauna Loa Observatory Atmospheric Transmission

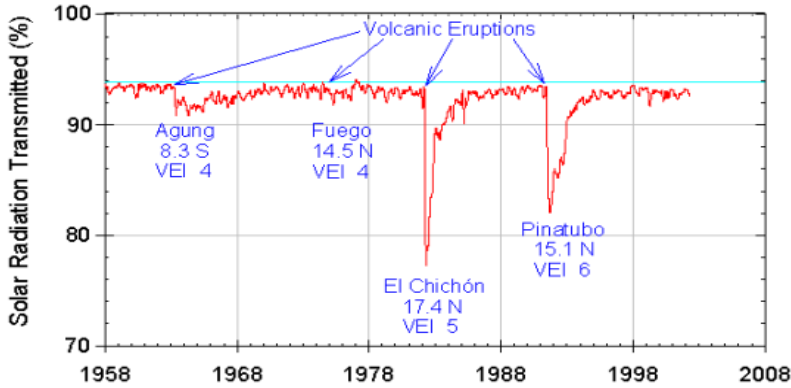


Figure 2.10: Solar radiation patterns of some recent volcanic eruptions.

Volcanoes erupt both gasses and material, and if sufficiently large amounts are ejected high into the atmosphere, their emissions can block solar radiation thereby reducing temperatures, sometimes on a global scale. Because their ejected material gets gradually removed from the atmosphere, the temperature ‘shape’ caused by an eruption is relatively similar across different volcanoes, as illustrated by Figure 2.10. The abrupt initial fall in temperature creates ‘outliers’ in temperature reconstructions, such as that based on dendrochronology used here. To locate statistically significant drops in temperature of a form likely to be from a volcanic eruption, Pretis *et al.* (2016) ‘designed’ a saturating set of indicators from the physical-theory shape of ν to match the temperature response using d_t above, an illustrative subsample of which is shown in Figure 2.11.

The principle of selecting significant indicators from the VIS saturating set is just like that for IIS. Although there is considerable uncertainty about the timings and magnitudes of some eruptions, VIS saturation estimation can help correct dendrochronological temperature records (see Schneider *et al.*, 2017, for a new archive of large volcanic events over the past millennium based on VIS). Figure 2.12 top panel shows a temperature reconstruction since 1200bp.

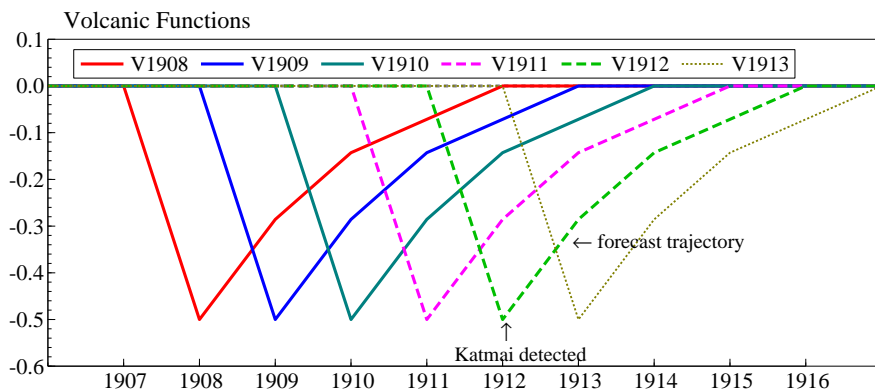


Figure 2.11: Illustrative subsample of volcanic designed indicators.

The issue is whether the large drops coincide with volcanic eruptions. Having detected significant indicators, many of these can be checked against dates of known eruptions such as Tambora in 1816 (the year without a summer when it snowed in New Haven in July, *Frankenstein* was written by Mary Shelley when stuck for months in a Swiss Chateau from almost ceaseless rain and J.M. Turner painted his remarkable skies) or Krakatoa in 1883. Figure 2.12 reports the outcome and the names of the volcanoes detected applying VIS in a first-order autoregressive model.

2.7 Summary of saturation estimation

This section summarizes saturation estimators under the null as delivered by split half for the DGP $y_t \sim \text{IN}[0, \sigma_\epsilon^2]$, $t = 1, \dots, T$ when the model fitted to the first-half is given by $y_t = \sum_{j=1}^{T/2} \gamma_j d_{\{j\}} + \epsilon_t$ so no intercept is included in the model. Here, $d_{\{j\}}$ denotes the appropriate indicator. For IIS, $d_{\{j\}} = 1_{\{j=t\}}$, for SIS, $d_{\{j\}} = 1_{\{j \leq t\}}$, for TIS $d_{\{j\}} = 1_{\{j \leq t\}} t$ and for MIS, $d_{\{j\}} = 1_{\{j \leq t\}} x_t$. Finally, the indicator form for the one example of VIS is $\mathbf{d}_{\{t\}}$ in the previous subsection. The formulae for the t-tests, t_t , on the t^{th} indicator under the null, as well as their non-centralities, ψ_1 , for a single alternative, for all these saturation estimators are derived in Castle *et al.* (2019b), who also record the corresponding non-centrality, ψ_k , when applying those t-tests for a known alternative.

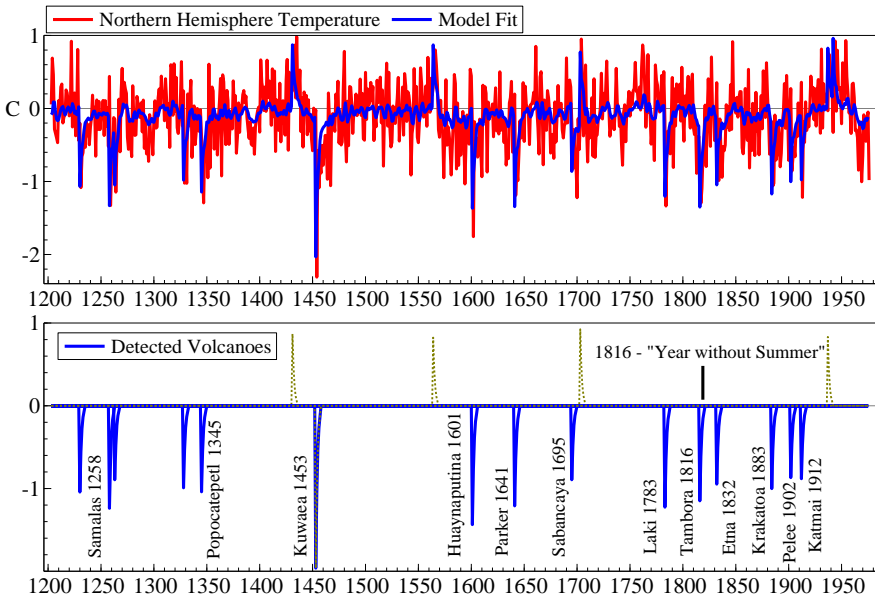


Figure 2.12: Top panel: temperature data and the fit of the model; bottom panel: detected volcanoes, 1200–2000.

Their analysis reveals the basic structure of the various indicator saturation estimators and t_t tests in a split-half analysis. All the saturation methods transpire to estimate a weighted combination of the current and next error $(w_{j,t}\epsilon_t - w_{j,t+1}\epsilon_{t+1})$. For example, the IIS estimate of the coefficient $\hat{\gamma}_t$ of $1_{\{t\}}$ can be written as $(\epsilon_t - 0\epsilon_{t+1})$ so $w_{1,t} = 1$ and $w_{1,t+1} = 0$ as each impulse exactly removes the corresponding error. For SIS, $w_{2,t} = 1 = w_{2,t+1}$; for TIS, $w_{3,t} = t^{-1}$ and $w_{3,t+1} = (t+1)^{-1}$; and for MIS with a regressor z_t , $w_{4,t} = z_t^{-1}$ and $w_{4,t+1} = z_{t+1}^{-1}$ so the estimate can be erratic for near-zero z_t ; and for VIS, $w_{5,t} = 1$ and $w_{5,t+1} = 0.5$ (ignoring smaller weights on ϵ_{t+i}). Thus SIS is MIS for the constant, and TIS is MIS for the trend, two important special cases. The resulting t_t tests are then a scaled combination of the weighted current and next error, albeit with different weights. The ‘future’ only appears to be included because the steps are defined as cumulating the impulses up to the given time, so induce forward differencing: using the isomorphic reverse formulation would lead to backward differencing.

Because impulses are mutually orthogonal, $\hat{\gamma}_t$ for IIS will be the same under the null as when a single impulse indicator is included at t . However,

for SIS, when only a single step is included at T_1 , $\hat{\gamma}_{T_1}$ becomes $(\bar{\epsilon}_{(1)} - \bar{\epsilon}_{(2)})$ where $\bar{\epsilon}_{(1)} = T_1^{-1} \sum_{t=1}^{T_1} \epsilon_t$ and $\bar{\epsilon}_{(2)} = (T - T_1)^{-1} \sum_{t=T_1+1}^T \epsilon_t$, so has the same form, but would have a much smaller variance. Similar changes occur for the other cases when a single indicator is used, although under the null such formulae are not overly insightful.

More importantly are the non-centralities, ψ_1 when split-half saturation estimation is applied and the break occurs in the first half of the sample at $T_1 \leq T/2$. The comparable non-centrality, ψ_k , is that of a t-test for the known form, timing and length of break, which would also apply when the saturation estimator found precisely the correct outcome. As Hendry and Santos (2005) note, for the IIS t-test, $\psi_1 = \psi_k$ so both are inconsistent. For the other saturation estimators, $\psi_1 \leq \psi_k$, so selection is essential to eliminate irrelevant but highly collinear indicators and improve potency over split half. For SIS, the non-centrality is increased by $\sqrt{2T_1}$ from correct selection, with similar large improvements for the others. In practice, breaks are often detected at dates close to the actual occurrence rather than the precise date since errors with the opposite signs to a shift can delay detection or bring forward the apparent end, whereas with the same sign can lead to an earlier start or later end. The analysis around (2.10) explained why. Consequently, when simulating saturation estimators, we usually calculate the potency for the actual date \pm a few periods.

The resulting rejection frequencies depend on the choice of α and correctly retaining all other relevant variables. When only IIS is used, but there is a location shift, many similar magnitude, same-sign indicators may be retained, and these can be costlessly combined at their average value—this will be almost the same outcome as applying SIS: see Hendry and Santos (2005).

Conversely, using SIS when there is a single outlier requires two indicators to characterize it. As noted above, such indicators can be combined manually to recreate an impulse indicator; alternatively, super saturation can help avoid that, but at the cost of a tighter α and hence potential loss of potency for smaller breaks. All these aspects are illustrated in Chapter 7.

When IIS and SIS are applied in a system, as in Chapter 6, indicators are retained by a likelihood-ratio test, so depend on their significance in the system as a whole not in any individual equation therein. However, an unrestricted VAR can be estimated an equation at a time, so saturation estimation can also

be applied an equation at a time and compared to the system selection to see if very different shifts or outliers occur in different equations. All selection decisions in both settings are based on likelihood-ratio tests as described in Section 2.3, albeit these often coincide with conventional t-tests, which allows a seamless transition between classes of models.

2.8 Selecting simultaneous equations models

A simultaneous equations representation is a model of a system of n variables, \mathbf{y}_t , that are to be modeled as endogenous by m other variables, \mathbf{z}_t , that are non-modeled. The properties of such systems were first analyzed by Trygve Haavelmo (1943), and are a potential representation when modeling the local data generation process (LDGP: see Section 2.1, & Hendry, 2009, 2018). To validly condition on \mathbf{z}_t requires that those variables are known to be at least weakly exogenous (see Tjalling Koopmans, 1950b, and Engle *et al.*, 1983). In essence, the weak exogeneity of \mathbf{z}_t requires that the DGP of \mathbf{z}_t does not depend on the parameters of the DGP of \mathbf{y}_t conditional on \mathbf{z}_t . When the status of \mathbf{z}_t is not certain, as with CO₂ in Chapter 6, \mathbf{z}_t should be treated, at least initially, as a component of \mathbf{y}_t . The strong exogeneity of \mathbf{z}_t , as applies to the orbital drivers in Chapter 6, requires that their DGP also does not depend on lagged values of \mathbf{y}_t , in which case non-linear functions of the \mathbf{z}_t can also be included as ‘conventional’ conditioning determinants of \mathbf{y}_t .

A dynamic representation of the system $(\mathbf{y}_t, \mathbf{z}_t)$ can be formulated as a vector autoregression (VAR) conditional on the \mathbf{z}_t (often denoted by VARX), lags of all the variables, and deterministic terms such as intercepts and any indicator variables, denoted \mathbf{d}_t :

$$\mathbf{y}_t = \Psi_0 \mathbf{z}_t + \sum_{j=1}^s \Psi_j \mathbf{z}_{t-j} + \sum_{i=1}^s \Gamma_i \mathbf{y}_{t-i} + \mathbf{A} \mathbf{d}_t + \mathbf{u}_t \quad (2.14)$$

where $\mathbf{u}_t \sim \text{IN}_n[\mathbf{0}, \Omega_u]$. The assumptions on the error process require that s is sufficiently large to create a martingale difference process, and the variables included in \mathbf{d}_t remove any outliers, location shifts and parameter changes not captured by the other regressors so that homoskedasticity and constant parameters are viable. Then given that \mathbf{z}_t is weakly exogenous, the error process will also be uncorrelated with the regressors.

To check the specification of (2.14), the system should be tested for congruence: once the initial system is congruent, all later reductions of it should be congruent as well to avoid relevant information being lost. Next, a parsimonious version of the system in (2.14) can be selected, while ensuring that congruence is maintained (denoted PVARX): Hendry and Mizon (1993) propose evaluating such dynamic models by their ability to encompass the VAR. Since the initial system is identified, all later non-simultaneous reductions from eliminating insignificant variables must be as well. At this selection stage, the system should also have been reduced to a non-integrated ($I(0)$) representation so that conventional critical values can be used: if the data are $I(1)$, cointegration and differencing can do so: see e.g., Johansen (1995), and See Doornik and Katarina Juselius (2018).

A system like (2.14), also called the ‘reduced form’ in the economics literature, is always identified, so that multivariate least-squares estimators of its parameters are unique, and under these assumptions will deliver consistent estimates. A simultaneous equations representation is a model of the system derived by reduction from (2.14). However, in economics there is often a pre-specified theory of that representation from which the ‘reduced form’ is derived (hence the terminology), inverting the correct order of the relationship between the system and a model thereof.

Written in a concise notation, with the $N \times 1$ vector \mathbf{w}_t denoting all the right-hand side variables, the system in (2.14) is:

$$\mathbf{y}_t = \mathbf{\Pi} \mathbf{w}_t + \mathbf{u}_t \text{ where } \mathbf{u}_t \sim \text{IN}_n[\mathbf{0}, \mathbf{\Omega}_u]. \quad (2.15)$$

Then a simultaneous-equations model of (2.15) is a reduction to:

$$\mathbf{B} \mathbf{y}_t = \mathbf{C} \mathbf{w}_t + \boldsymbol{\epsilon}_t \text{ where } \boldsymbol{\epsilon}_t \sim \text{IN}_n[\mathbf{0}, \mathbf{\Sigma}_\epsilon], \quad (2.16)$$

with:

$$\mathbf{B} \mathbf{\Pi} = \mathbf{C} \text{ and } \mathbf{\Sigma}_\epsilon = \mathbf{B} \mathbf{\Omega}_u \mathbf{B}'. \quad (2.17)$$

A necessary condition for (2.17) to be solvable is that there are no more non-zero parameters in \mathbf{B} and \mathbf{C} than the $n \times N$ in $\mathbf{\Pi}$, which is called the order condition. In addition, when the rank condition discussed in §2.8.1 is satisfied, \mathbf{B} and \mathbf{C} have a unique relation to $\mathbf{\Pi}$ and (2.16) is fully identified. We use ‘structure’ (in quotation marks) to denote an equation with more than one endogenous variable as in (2.16), without any claim that it is a structural

equation in the sense of being invariant to extensions of the information set for new variables, over time, and across regimes. A simultaneous-equations model like (2.16) then needs to be estimated appropriately, because including the i^{th} endogenous variable in the equation for the j^{th} will induce a correlation with its equation error. An infinite number of possible estimation methods exists, characterized by the estimator generating equation in Hendry (1976). Here we use full information maximum likelihood (FIML) first proposed in Koopmans (1950a). The general formulation and estimation procedures underlying FIML are described in Hendry, Adrian Neale and Frank Srba (1988). Since a simultaneous-equations model is a reduction from the system, automatic model selection is applicable as discussed in Hendry and Krolzig (2005); Doornik and Hendry (2017) propose an algorithm for doing so based on the multi-path search procedure of *Autometrics*, a variant of which is applied in Chapter 6.

2.8.1 Identification

Identification, in the sense of uniqueness of \mathbf{B} , \mathbf{C} , in systems like (2.16) given $\mathbf{\Pi}$ has been extensively explored in the econometrics literature: see e.g. Koopmans (1949), Koopmans and Olaf Reiersøl (1950), Frank Fisher (1966) and Thomas Rothenberg (1973) *inter alia*. The rank condition for identification determines the extent to which each equation is or is not identified. In that literature, identification is usually a synonym for uniqueness, although usage also entails connotations of ‘interpretable in the light of subject matter theory’ and ‘corresponding to reality’ (as in ‘identify a demand curve’, as opposed to a supply relation, or a mix). Whether or not \mathbf{B} , \mathbf{C} in (2.16) can be recovered uniquely from $\mathbf{\Pi}$ in (2.15) requires the exclusion of some different variables in every equation and the inclusion of some others, otherwise linear combinations of equations cannot be distinguished.³ Given the appropriate exclusions and inclusions corresponding to particular elements of \mathbf{B} and \mathbf{C} being zero, the rank condition is then satisfied so (2.16) is fully identified. Consequently, \mathbf{B} and \mathbf{C} are uniquely related to $\mathbf{\Pi}$, which entails restrictions on the $\mathbf{\Pi}$ matrix in (2.15). The system for the three ice-age variables in Chapter 6 is highly

³Other forms of restriction than exclusions could identify ‘structural’ parameters that do not directly satisfy the rank condition, such as a diagonal error covariance matrix, or cross-equation links, but these are not considered here.

overidentified, so all the $y_{i,t}, i \neq j$ can be included in equations for $y_{j,t}, j \neq i$.

To ensure a unique relationship and hence avoid ‘spurious identification’ of a simultaneous representation, all the right-hand side variables need to be significant at a reasonable level both in the system and in their associated equations. Otherwise, claiming identification by excluding insignificant regressors from any equation based on their apparent presence in other equations, when in fact they are also insignificant there, will be misleading when such variables are actually irrelevant to the system as a whole. Throughout selection of a simultaneous representation, the rank condition for identification can be imposed as a constraint, both to ensure that essentially the ‘same equation’, but with different normalizations, is not included twice, and that at every stage, the current form is identified (see Hendry *et al.*, 1988).

There are three possibilities of lack of identification, just identification, and over identification (subsets of parameters could be identified or not when others are the converse, in which case the following comments apply to the appropriate set). When \mathbf{B} , \mathbf{C} are not identified, then (2.15) is the least restricted but still fully identified, representation. Any just-identified simultaneous representation with a form like (2.16) will also be minimally identified, so there is an equivalence class of such specifications with equal likelihood (see e.g., Rothenberg, 1971), although in such a setting, reductions may be possible by eliminating irrelevant regressor variables from the entire system.

When \mathbf{B} , \mathbf{C} are over identified by the rank condition, then (2.16) is a unique representation for the given restrictions. However, Hendry, Maozu Lu, and Mizon (2009) show there may exist different sets of restrictions embodied in matrices \mathbf{B}^* , \mathbf{C}^* which are not linear transforms of \mathbf{B} , \mathbf{C} (precluded by their identifiability), but under which (2.16) is equally over identified. Thus, again an equivalence class of such specifications with equal likelihood can result: a given degree of over identification by itself does not ensure a unique model even when there is a unique DGP. The validity of any set of over-identified restrictions can be checked through parsimonious encompassing of the system by the ‘structure’. When \mathcal{L} is the log-likelihood of the system (2.15), and \mathcal{L}_0 that of the ‘structural’ form (2.16), in stationary DGPs, the test is $2(\mathcal{L} - \mathcal{L}_0) \sim \chi^2_{\text{OR}}(s)$ for s over-identifying restrictions (see Koopmans, 1950a, and Hendry and Mizon, 1993).⁴

⁴Hence the earlier advice to obtain an $l(0)$ and constant representation of the system.

2.9 Forecasting in a non-stationary world

We will undertake forecasts for both major illustrations below, so need to address two key aspects of wide-sense non-stationarity. First, it affects forecasting directly though the different approaches needed to select models for such data as discussed in the preceding sections; and secondly, because the observations to be forecast will also be non-stationary, different forecasting devices may be required. Specifically, location shifts at or near the forecast origin can lead to forecast failure as emphasized by Clements and Hendry (1998), as of course can shifts that occur after forecasts are made. Systematic forecast failure, defined as when forecasts are significantly different from the later outcomes compared to their *ex ante* forecast intervals, is mainly caused by direct or induced shifts in the long-run means of the variables being forecast. We have suggested saturation estimation during model selection as a complement to cointegration to jointly handle non-stationarity in-sample. In this section, we describe both the consequences for forecasts of not handling location shifts near the forecast origin, and consider forecasting devices that are more robust than ‘conventional’ methods after such shifts.

Almost all econometric model formulations are implicitly or explicitly equilibrium correction: this huge class includes regressions, autoregressions, VARs, cointegrated systems, dynamic stochastic general equilibrium models (DSGEs), autoregressive conditional heteroskedasticity error processes (ARCH) and generalizations thereof like GARCH. For example, a stationary scalar first-order autoregression of the form:

$$y_t = \rho_0 + \rho_1 y_{t-1} + \epsilon_t = \mu + \rho_1 (y_{t-1} - \mu) + \epsilon_t, \quad (2.18)$$

where $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$ with $|\rho_1| < 1$, and $\mu = \rho_0/(1 - \rho_1)$ is the long-run mean, so $\mathbb{E}[y_t] = \mu$, can be rewritten as:

$$\Delta y_t = (\rho_1 - 1)(y_{t-1} - \mu) + \epsilon_t. \quad (2.19)$$

Since $|\rho_1| < 1$, when $y_{t-1} > \mu$ then $\Delta y_t < 0$ and the process is brought back towards μ , and similarly when $y_{t-1} < \mu$. In that way, (2.19) ‘error corrects’, as such mechanisms are often called. Unfortunately, if μ changes to μ^* say, (2.19) will still equilibrium correct towards μ and will continue to do so until revised to replace μ by μ^* , so does not error correct to the new long-run mean.

Should such a shift occur at the forecast origin y_T , where ρ_1 changes to ρ_1^* so $\mu^* = \rho_0/(1 - \rho_1^*)$ then the next observation will actually be:

$$\Delta y_{T+1} = (\rho_1^* - 1)(y_T - \mu^*) + \epsilon_{T+1}, \quad (2.20)$$

whereas from (2.19), the 1-step ahead forecast will have been (ignoring parameter estimation uncertainty for simplicity as second order compared to the shift):

$$\widehat{\Delta y}_{T+1|T} = (\rho_1 - 1)(y_T - \mu), \quad (2.21)$$

leading to the forecast error $\widehat{\epsilon}_{T+1|T} = \Delta y_{T+1} - \widehat{\Delta y}_{T+1|T}$:

$$\begin{aligned} \widehat{\epsilon}_{T+1|T} &= (\rho_1^* - 1)(y_T - \mu^*) - (\rho_1 - 1)(y_T - \mu) + \epsilon_{T+1} \\ &= (1 - \rho_1^*)(\mu^* - \mu) + (\rho_1^* - \rho_1)(y_T - \mu) + \epsilon_{T+1}, \end{aligned} \quad (2.22)$$

and since $\mathbb{E}[y_T - \mu] = 0$:

$$\mathbb{E}[\widehat{\epsilon}_{T+1|T}] = (1 - \rho_1^*)(\mu^* - \mu) \neq 0 \text{ if } \mu^* \neq \mu. \quad (2.23)$$

Thus, (2.21) fails to correct the error induced by the location shift.

An error like (2.23) will persist in future periods if the in-sample model is used unchanged as:

$$\widehat{\Delta y}_{T+2|T+1} = (\rho_1 - 1)(y_{T+1} - \mu), \quad (2.24)$$

whereas:

$$\Delta y_{T+2} = (\rho_1^* - 1)(y_{T+1} - \mu^*) + \epsilon_{T+2},$$

so that letting $\nabla \rho_1 = \rho_1^* - \rho_1$ and $\nabla \mu = \mu^* - \mu$:

$$\widehat{\epsilon}_{T+2|T+1} = (1 - \rho_1^*)\nabla \mu + \nabla \rho_1 (y_{T+1} - \mu) + \epsilon_{T+2}, \quad (2.25)$$

with:

$$\mathbb{E}[\widehat{\epsilon}_{T+2|T+1}] = (1 - \rho_1^*)(1 + \nabla \rho_1)\nabla \mu,$$

because:

$$\mathbb{E}[y_{T+1} - \mu] = \mathbb{E}[(\rho_1^* - 1)(y_T - \mu^*)] = (1 - \rho_1^*)\nabla \mu.$$

Similar mistakes of not error correcting after a shift in the long-run mean will affect all members of the equilibrium-correction class.

A surprising feature of these second period 1-step ahead forecasts in (2.24) is that if (say) $\nabla\mu < 0$ so there has been a **downward** shift in the long-run mean, then:

$$\mathbb{E} [\hat{y}_{T+2|T+1} - y_{T+1}] = -(1 - \rho_1)(1 - \rho_1^*) \nabla\mu \geq 0, \quad (2.26)$$

so that on average, $\hat{y}_{T+2|T+1} \geq y_{T+1}$, and the next forecast is usually **above** the previous outcome, and conversely when $\nabla\mu < 0$. This creates a ‘hedgehog’ effect in the graph of forecasts around outcomes, and is caused by (2.19) correcting to the old equilibrium μ and hence in the opposite direction to μ^* .

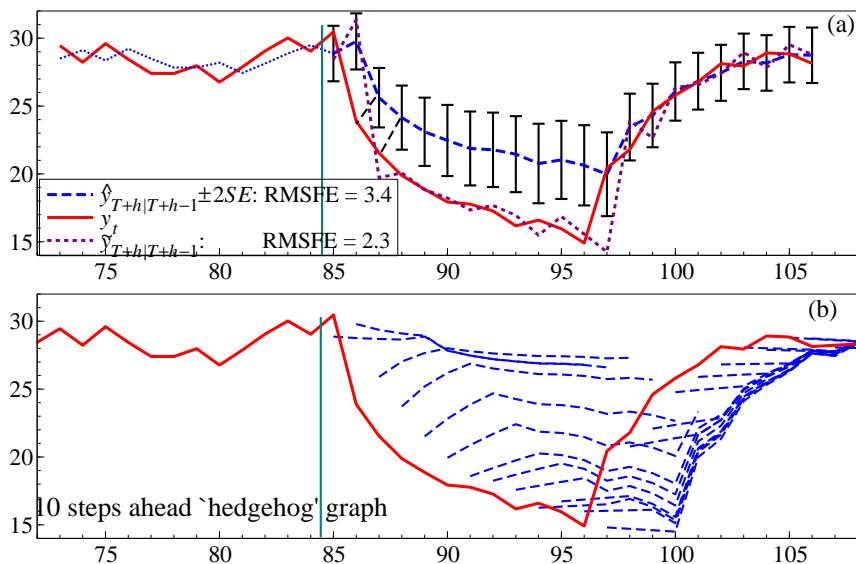


Figure 2.13: (a) successive 1-step ahead forecasts after a location shift at $t = 86$ and back at $t = 96$ for both ‘conventional’, $\hat{y}_{T+h|T+h-1}$, and ‘robust’, $\tilde{y}_{T+h|T+h-1}$, forecasts; and (b) 10-steps ahead forecasts, $\hat{y}_{T+h|T+h-10}$, to highlight the hedgehog effect.

Figure 2.13(a) illustrates with computer generated data on a model matching (2.18) where $\rho_0 = 10$ and $\rho_1 = 0.65$, but here ρ_0 is changed to $\rho_0^* = 6$ which again shifts μ , now from 28.6 to 16.1 at observation $t = 86$ and back to its original value at $t = 96$ to create a ‘recession’ like pattern. The forecasts are based on an in-sample model matching the DGP with estimated values of $\hat{\rho}_0 = 10.6$ and $\hat{\rho}_1 = 0.63$, so are close to the DGP parameter values.

The forecast failure from $\hat{y}_{T+h|T+h-1}$ is marked as all the outcomes from $t = 86, \dots, 96$ lie outside the 95% error bars, and the forecasts come back to the data only after the shift ends. To illustrate the ‘hedgehog’ effect, lines are drawn from the outcomes at $t = 86$ and $t = 87$ to the corresponding forecasts for the next periods, both of which lie well above the previous observed values.

Figure 2.13(a) also records the forecasts from a robust device defined for $h = 2, \dots, H$ by (see Hendry, 2006):

$$\begin{aligned}\tilde{y}_{T+h|T+h-1} &= y_{T+h-1} + \hat{\rho}_1 \Delta y_{T+h-1} \\ &= y_{T+h-1} + \hat{\rho}_1 (y_{T+h-1} - y_{T+h-2}),\end{aligned}\quad (2.27)$$

which are dramatically better over the shift period with a 33% smaller root mean-square forecast error (RMSFE) overall. So why does a mis-specified ‘model’ like (2.27) forecast better than the estimated in-sample DGP? An explanation follows from comparing the second expression in (2.18) with the second expression in (2.27), which reveals that the first long-run mean μ in the former is replaced by the ‘instantaneous’ estimator y_{T+h-1} and the second by y_{T+h-2} . These are very noisy but unbiased estimators when there is no shift in μ , and highly adaptive estimators of μ^* after a shift. Simplifying by ignoring parameter estimation, and taking expected values, when μ shifts because of a constant ρ_1 but changed ρ_0 , then forecasting from $T + 1$ to $T + 2$:

$$\mathbb{E} [\tilde{y}_{T+2|T+1}] = \mathbb{E} [y_{T+1}] + \rho_1 \mathbb{E} [\Delta y_{T+1}] = \mu^* - \rho_1^2 \nabla \mu, \quad (2.28)$$

since:

$$\mathbb{E} [y_{T+1}] = \mu^* - \rho_1 \nabla \mu \text{ and } \mathbb{E} [\Delta y_{T+1}] = (1 - \rho_1) \nabla \mu,$$

then for $\tilde{\epsilon}_{T+2|T+1} = y_{T+2} - \tilde{y}_{T+2|T+1}$ as $\mathbb{E} [y_{T+2}] = \mu^* - \rho_1^2 \nabla \mu$:

$$\mathbb{E} [\tilde{\epsilon}_{T+2|T+1}] = \mu^* - \rho_1^2 \nabla \mu - (\mu^* - \rho_1^2 \nabla \mu) = 0,$$

as against:

$$\mathbb{E} [\hat{\epsilon}_{T+2|T+1}] = (1 - \rho_1) \nabla \mu.$$

In Chapter 7, we forecast by both the equivalent of (2.21) and (2.27). The robustness depends on forecasting later than the shift, and does not improve the forecast from T as Figure 2.13(a) shows. Although the algebra does not simplify neatly when ρ_1 changes, the principle is the same and much of the

error cancels unless the dynamics change greatly. However, the robust device in (2.27) is noisy and over-shoots when a shift ends, and an improved device is proposed by Martinez *et al.* (2019).

Castle *et al.* (2015a) re-interpret (2.27) as:

$$\tilde{y}_{T+h|T+h-1} = \tilde{\mu}_a + \hat{\rho}_1(y_{T+h-1} - \tilde{\mu}_b), \quad (2.29)$$

where $\tilde{\mu}_a$ and $\tilde{\mu}_b$ can be estimated by averages of past data rather than just a single data point. Then (2.29) defines a class of forecasting devices where $\hat{\mu}$ is the full sample average based on T observations, through to (2.27) which is based on one data point. Equally, combinations of several members of a robust class could be used, as using longer averages in estimating μ entails slower adjustment, but less volatility, as does forecast combination in general.

Figure 2.13(b) records 10-steps ahead forecasts to highlight the hedgehog effect: well above during the drop in the variable to be forecast, and well below during the rise (an upside-down hedgehog). The former is similar to the substantive over-forecasts of productivity by the UK Office of Budget Responsibility since 2008, not adjusting to the ‘flat-lining’ seen in §2.6.5 above, much improved by the device in Martinez *et al.* (2019).

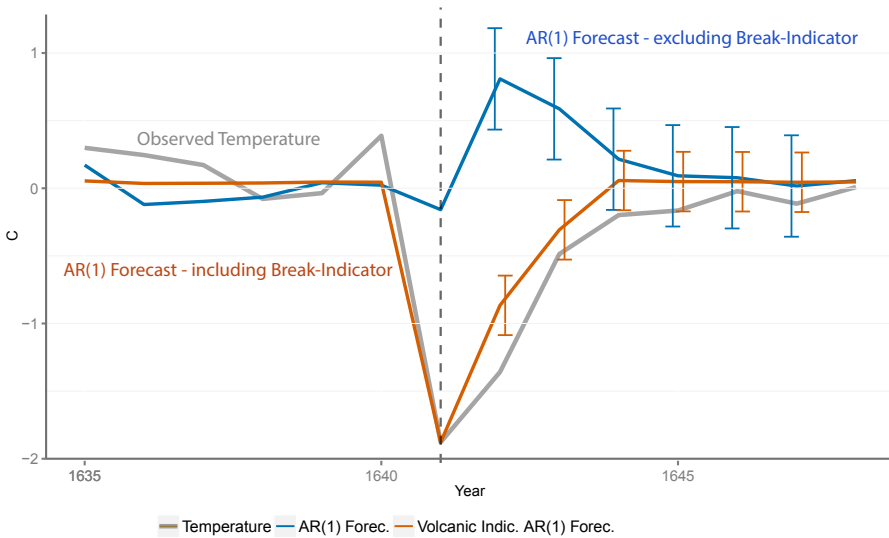


Figure 2.14: Forecasting through the example eruption in 1641 of volcano Parker in the Philippines.

Another example of forecasting through breaks is provided in Figure 2.14 for the eruption in 1641 of volcano Parker in the Philippines. The model using VIS to detect volcanic eruptions was described in §2.6.7, and the Figure shows the forecasts from the first-order autoregression (denoted AR(1)) without saturation estimation, and that model with the detected indicator selected from the single observation at the eruption date. As can be seen, the AR(1) without the break indicator forecasts a rise as derived from the theory above, whereas that with the VIS form estimated just from the initial fall does a reasonable job of tracking the temperature recovery despite forecasting six periods ahead from an indicator fitted to a single data point. Because the break pattern is very different from a location shift, the robust device described in (2.27) also performs poorly, although its forecasts are not shown.

3

Some hazards in empirical modeling of non-stationary time-series data

We describe some of the hazards that can be encountered when empirically modeling non-stationary time-series data, with potential implications for analyzing observations on climate variables. Most importantly, the degree of integration of time-series data need not be constant, as integrability is a derived, not an intrinsic, property of a time-series process, a key issue addressed in this section. Other hazards include unmodeled shifts in relationships, or more generally, incorrectly modeled relations omitting substantively important explanations; data measurement errors; and aggregation bias. Any of these can seriously distort empirical statistical studies, leading to mistaken inferences and hence fallacious conclusions, and possibly false causal attribution. We use an example from societal behavior related to climate change (vehicle distances traveled) to illustrate how false implications can arise, even when an analysis is undertaken in a substantive and well understood framework, not ‘mere data mining’, and to explore the possibility of revealing such problems. To highlight as many hazards as possible in a simple example, the section is written like a ‘detective’ story where a potential culprit is only revealed towards the end.

Our analysis of what can go wrong with an empirical statistical study builds on the critique by Pretis and Hendry (2013) of the claimed absence of links

between temperature and radiative forcing of greenhouse gases in Michael Beenstock, Yaniv Reingewertz, and Nathan Paldor (2012). Those authors used the measure of the changes in the radiative forcing of CO₂ shown in Figure 3.1, merging ice-core based data with Charles Keeling’s direct atmospheric readings (discussed below). The time-series properties of the subperiods are clearly different: the data up to 1958 seem to be $I(0)$ around a level, and after are trending up with a much larger variance. Combining the two sub-samples suggests that the overall sample is $I(2)$, and as the temperature time series is $I(1)$ Beenstock *et al.* claim it cannot be caused by radiative forcing despite the well-established theory to the contrary.

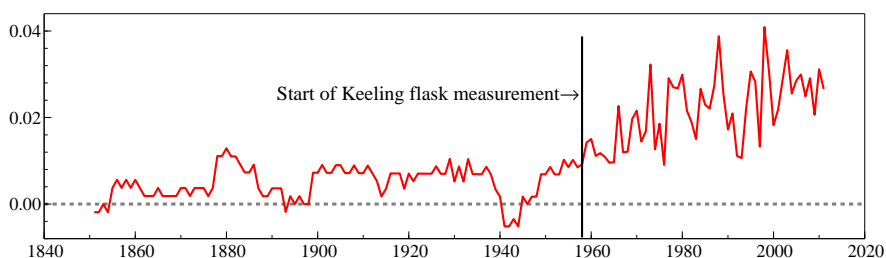


Figure 3.1: Change in the radiative forcing measure of CO₂.

Pretis and Hendry (2013) highlighted the hazards that can be encountered in statistical analyses by an example that implies the absurd conclusion that moving vehicles do not cause human road fatalities, to match the equally absurd ‘proof’ that greenhouse gases don’t cause climate change. The examples of Venus, boiled by an excess of greenhouse gases, and Mars, frozen by a lack thereof, are notable extremes. Here, we address the problems that can be encountered in empirical modeling using their road fatalities example as one where the link to moving vehicles is not to our knowledge disputed.

For millennia, from horse-drawn carriages to engine-driven vehicles, collisions with people have injured and killed them, vastly more so as cars initially proliferated. Figure 3.2 updates the UK data in Pretis and Hendry (2013) for total vehicle distances driven in billions of kilometers p.a. (denoted D_t) and road fatalities (F_t) to 2017, with 6 new observations since their earlier analysis for an out-of-sample evaluation. The four panels labeled a, b, c, d respectively show F_t , $\Delta F_t = F_t - F_{t-1}$, D_t and ΔD_t .

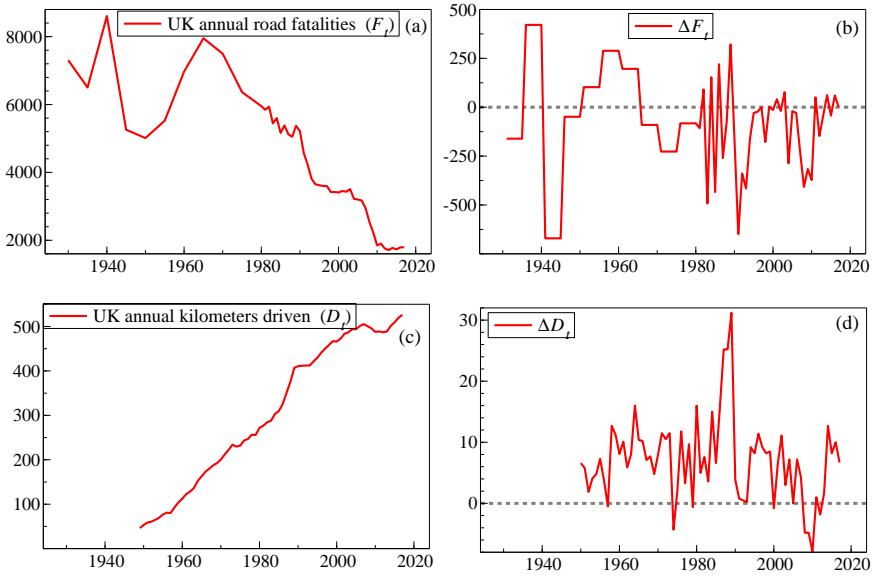


Figure 3.2: (a) Road fatalities p.a. in the UK (F_t); (b) its first differences (ΔF_t); (c) vehicle kilometers driven p.a. in billions (D_t); and (d) its first differences (ΔD_t).

Data on UK road fatalities are only available continuously from 1979 onwards and previously were interpolated from intermittent data, as the graph of ΔF_t reveals. The ‘blocks’ and large jumps suggest major data measurement errors, changing its time-series properties, although only observations from 1949 can be used in the regressions below given the shorter sample on D_t . It is also manifest from their graphs that F_t and D_t are highly non-stationary with changing means and variances, and have strong opposite trends. If an empirical analysis is undertaken of these two variables, absent subject-matter knowledge, then it ‘proves’ that greater distances driven by vehicles each year lead to fewer road deaths.

We establish that finding by a statistical analysis of these two variables that includes testing for cointegration, checking the constancy and invariance of the relationship, and conditional forecasting of the six new observations on fatalities given the traffic data. The empirical model passes all the required tests, providing a near congruent explanation that even satisfies a test for

invariance to large shifts in D_t . This example also serves to establish our notation. The aim is to stress that however sophisticated a statistical analysis may appear to be, the implications have to be understood in a substantive context: the claims that road fatalities are not caused by moving vehicles, or could be reduced by vehicles driving greater distances, are both absurd.

Our illustration is a first-order autoregressive-distributed lag model (ADL: see Hendry, 1995, Ch. 7), relating F_t to its first lag, F_{t-1} , a constant, D_t and D_{t-1} estimated by least squares over 1951–2011:

$$\begin{aligned}
 \hat{F}_t &= 728 + 0.912 F_{t-1} + 12.3 D_t - 13.7 D_{t-1} \\
 &\quad (169) \quad (0.022) \quad (3.16) \quad (3.2) \\
 \hat{\sigma} &= 163, \quad R^2 = 0.99, \quad F_{AR}(2, 55) = 3.53^* \\
 \chi_{nd}^2(2) &= 3.66, \quad F_{Het}(6, 54) = 1.17 \\
 t_{ur} &= -4.08^{**}, \quad F_{ARCH}(1, 59) = 0.02, \quad F_{Reset}(2, 55) = 2.83
 \end{aligned} \tag{3.1}$$

In (3.1), estimated coefficient standard errors are in parentheses below estimated coefficients, $\hat{\sigma}$ is the residual standard deviation, and R^2 is the coefficient of multiple correlation: see §2.2.4 for model evaluation test statistics.

All the estimated coefficients are highly significant in (3.1), there is a near perfect fit, only one mis-specification test is significant at even 5%, and the *PcGive* unit-root t-test, denoted t_{ur} , rejects the null hypothesis of no cointegration at 1% (see Ericsson and James MacKinnon, 2002). Moreover, the equation is constant over the new observations with $F_{Chow}(6, 57) = 0.06$ where F_{Chow} is a parameter constancy forecast test over 2012–2017 with a RMSFE of 42.1, which is a 1/4 of $\hat{\sigma}$ in (3.1), so the model fits the data that has arrived since Pretis and Hendry (2013) far better than the previous sample. The long-run derived solution from (3.1) is:

$$\tilde{F} = 8257 - 15.7D \tag{3.2}$$

(636) (2.4)

which has a **negative** effect from D . Thus, there is a long-run stationary relation between the non-stationary series of road fatalities and vehicle kilometers driven, such that road fatalities *decrease* with vehicle kilometers driven.

Figure 3.3 records the fitted and actual values and 1-step ahead conditional forecasts $\hat{F}_{T+h|T+h-1}|D_{T+h}$ for $h = 1, \dots, 6$, the residuals from (3.1) scaled by $\hat{\sigma}$, with their histogram, density and correlogram. There is residual autocorrelation, as well as visual evidence of some residual non-normality and

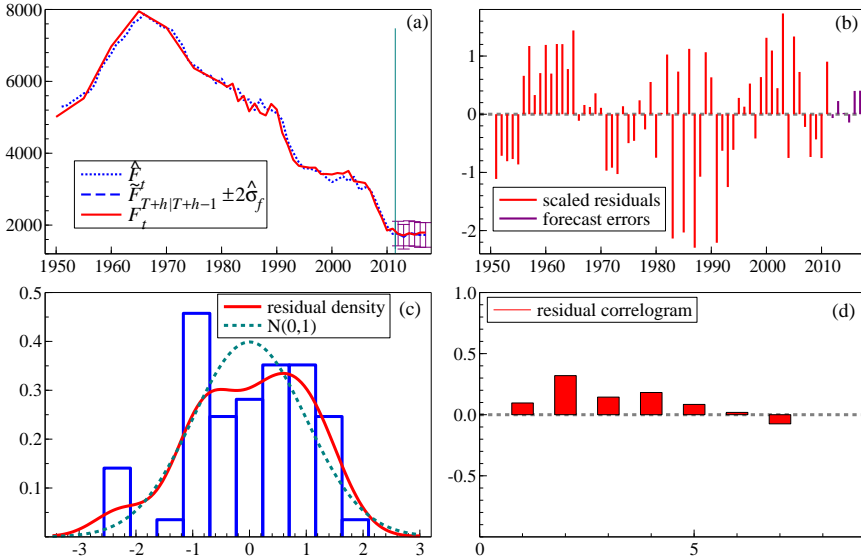


Figure 3.3: (a) Fitted, actual values and conditional forecasts of road fatalities p.a. in the UK; (b) residuals scaled by $\hat{\sigma}$; (c) scaled residual histogram and density with $N[0,1]$; and (d) residual correlogram.

heteroskedasticity. Applying IIS and SIS at 1% (see §2.6), one outlier was selected for 1987, denoted $1_{\{1987\}}$, with step shifts that ended in 1955 and 1995, denoted $S_{\{1955\}}$ and $S_{\{1995\}}$, respectively, the first of these applying to a very short initial sample that may not have been accurate. The outcome was a larger negative impact of D on F , with no significant diagnostic tests.

To assess short-run effects, Pretis and Hendry (2013) estimated an equilibrium-correction model using the derived long-run (cointegrating) solution in (3.2), modeling the changes in road fatalities by changes of vehicle kilometers driven and deviations from the long-run equilibrium:

$$\widehat{\Delta F}_t = \underset{(2.48)}{12.3} \Delta D_t - \underset{(0.012)}{0.088} \tilde{Q}_{t-1} \quad (3.3)$$

where $\tilde{Q} = F - \tilde{F}$, with a residual standard deviation of $\hat{\sigma} = 160$. Equation (3.3) shows a short-run increase in deaths as vehicle kilometers driven increase, with the long-run decrease embodied in \tilde{Q} .

3.1 Assessing the constancy and invariance of the relationship

The first check is on the constancy of the relationship in (3.1). In one interpretation, the step shifts $S_{\{1955\}}$ and $S_{\{1995\}}$ found above are evidence against that hypothesis, although the former only applies to 4 years. Conditional on their inclusion, the resulting recursive estimates of the other four coefficients are remarkably constant as shown in Figure 3.4.

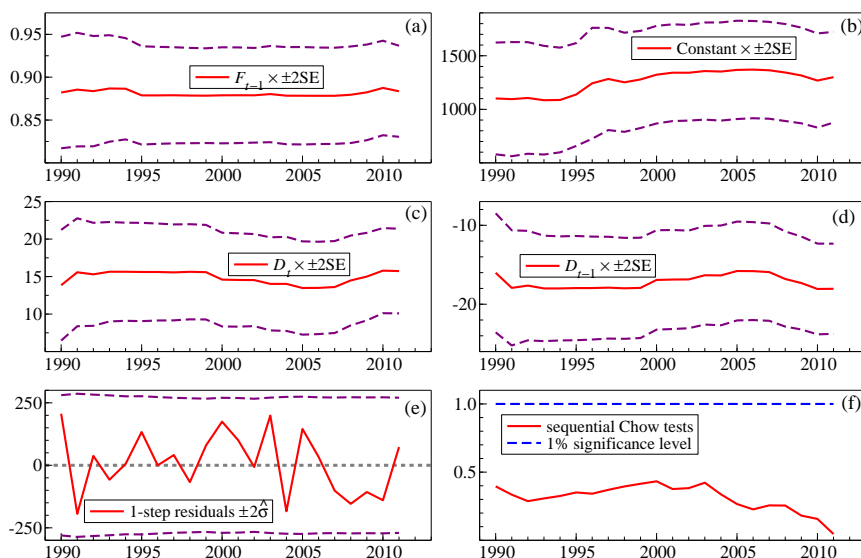


Figure 3.4: Recursive estimates of the coefficients of: (a) F_{t-1} ; (b) constant term; (c) D_t ; and (d) D_{t-1} , with (e) 1-step residuals and recursive estimates of σ ; and (f) recursively computed Chow tests.

To examine the possible ‘causal’ nature of (3.1), we next assessed its invariance to large changes in D_t (see e.g., Engle and Hendry, 1993, and Castle *et al.*, 2017). Applying SIS to a first-order autoregression in D_t over the whole period, four highly significant step indicators were retained ending in 1985, 1989, 2007, and 2013 ($p < 0.0002$). Adding these to (3.1) yielded an $F_{\text{exclude}}(4, 57) = 1.79$ which is insignificant. Thus, a powerful test for invariance of (3.1) does not reject that hypothesis. It would appear that F_t and D_t cointegrate in a congruent constant relation that remains so outside

the original data sample, and is invariant to the large shifts in D_t . How can such powerful statistical evidence fly in the face of the obvious falsity of the proposition that moving vehicles decrease road fatalities?

The third check is whether the assumed degree of integration is constant (here, $I(1)$). Merging data of ostensibly the same variable from different measurement systems can alter the apparent degree of integration, and that was a key problem with the Beenstock et al. (2012) study. We use an augmented Dickey–Fuller test (ADF: see David Dickey and Wayne Fuller, 1981) with a constant but no trend, roughly splitting the sample pre and post 1975. That for F_t over 1933–1975 yielded $t_{\text{adf}} = -2.74$ which is close to the 5% significance level of -2.93 despite the small sample, where the first lagged difference was highly significant with $t = 8.13$. In the second period, $t_{\text{adf}} = -1.31$ and no lagged differences were significant, which is a marked change. Referring back to Figure 3.2, the data behavior certainly changes noticeably after interpolation ends. The converse happens with D_t , albeit the samples are even smaller. Over 1951–1983, $t_{\text{adf}} = 0.06$ whereas over 1984–2017, $t_{\text{adf}} = -3.96^{**}$ with no lagged differences included, or $t_{\text{adf}} = -2.80$ with a significant first lagged difference. At best, the assumption of a constant degree of integration is dubious. Including a trend in these tests alters the outcomes such that the only near significant outcome is now for F in the second period with $t_{\text{adf}} = -3.25$ where the critical value is -3.52 .

However, that check highlights a possible issue: by failing to include a trend in the cointegration analysis, the formulation did not ensure the test was similar (see Nielsen and Anders Rahbek, 2000). In the present conditional single equation specification, adding a linear trend to an unrestricted version of (3.3) leads to:

$$\begin{aligned}
 \widehat{\Delta F}_t &= \underset{(638)}{1834} - \underset{(0.028)}{0.154} F_{t-1} - \underset{(1.16)}{1.16} D_{t-1} - \underset{(10.5)}{9.55} t \\
 &\quad + \underset{(2.94)}{11.8} \Delta D_t - \underset{(118)}{438} S_{\{1955\}} \\
 R^2 &= 0.55, \quad \hat{\sigma} = 151.9, \quad F_{\text{AR}}(2, 54) = 1.97, \quad F_{\text{ARCH}}(1, 60) = 0.65 \\
 \chi_{\text{nd}}^2(2) &= 2.82, \quad F_{\text{Het}}(9, 52) = 1.67, \quad F_{\text{Reset}}(2, 54) = 1.90
 \end{aligned} \tag{3.4}$$

with $F_{\text{Chow}}(6, 56) = 0.09$, where the remaining significant step indicator has also been retained. While both the trend and D_{t-1} are insignificant, if D_{t-1} is

deleted, the result is:

$$\begin{aligned}
 \widehat{\Delta F}_t &= \begin{matrix} 2337 \\ (389) \end{matrix} - \begin{matrix} 0.155 \\ (0.028) \end{matrix} F_{t-1} - \begin{matrix} 19.6 \\ (2.94) \end{matrix} t + \begin{matrix} 11.6 \\ (2.94) \end{matrix} \Delta D_t \\
 &\quad - \begin{matrix} 463 \\ (115) \end{matrix} S_{\{1955\}} \\
 R^2 &= 0.54, \quad \hat{\sigma} = 151.6, \quad F_{AR}(2, 55) = 2.22, \quad F_{ARCH}(1, 60) = 1.44 \\
 \chi^2_{nd}(2) &= 3.06, \quad F_{Het}(7, 54) = 1.66, \quad F_{Reset}(2, 55) = 2.18
 \end{aligned} \tag{3.5}$$

with $F_{Chow}(6, , 57) = 0.85$. This is at last interpretable: increases in road traffic lead to increased fatalities, but the trend reduction in deaths is due to cumulative improvements in many aspects of road safety.

There are many potentially relevant explanatory variables omitted from a model simply relating road fatalities to distances driven. A partial list affecting deaths from vehicle accidents would include improved driving standards after 1935 from more stringent driving tests, increasingly tough this century; safer cars with better brakes such as discs from the mid 1950s, then anti-lock braking systems (abs) in the 1980s, and improved crash impact designs; reduced fatalities from retractable front seat belts, compulsory from 1983 (see the analysis in Andrew Harvey and James Durbin, 1986) and in the later 1980s from air bags; increasing separation of opposite direction traffic flows on motorways from 1959 onwards; cameras at traffic lights as well as speed cameras; reductions in drunk driving from electronic breathalyzer tests after 1967 and lower acceptable alcohol limits; possibly heavier penalties for traffic violations; and so on.

In terms of reducing UK pedestrian fatalities, we note ‘zebra crossings with Belisha beacons’ dating back to 1934, and pedestrian controlled traffic lights in 1969 (often called pelican crossings in the UK); lower speed limits in urban areas; better road safety training, especially for children, etc. Converse effects come from faster cars; driver overconfidence; driving after taking drugs; and recently driving, and even walking, while using mobile phones etc. Also less has been done in the UK to protect cyclists: although fatalities have decreased slightly this century, the number of serious injuries has risen. Modeling total fatalities involves changing aggregation biases from different sub-populations being killed (pedestrians, pedal and motor cyclists, cars and other drivers) as a consequence of the differential effects of the above changes.

Overall, the reduction from almost 8000 deaths p.a. in 1967 to under 1800 in 2017 reflects these many factors cumulatively, although a constant linear trend is obviously a crude approximate description (but see the analysis of trends in Hendry, 1995, Ch. 15), and the pandemic will create large location shifts in both ΔF_t and ΔD_t .

3.2 An encompassing evaluation of the relationship

To discriminate between the two non-nested explanations by trend or D_{t-1} , we use an encompassing test between (3.5) and (3.3) over 1950–2011, both with the three indicators. The formal analysis of encompassing originates with Mizon and Richard (1986) who also relate it to non-nested hypothesis tests: see Bontemps and Mizon (2008) for a recent overview. The two hypotheses are denoted M_1 which relates ΔF_t to $1, F_{t-1}, t, \Delta D_t, S_{\{1955\}}, 1_{\{1987\}}, S_{\{1995\}}$ with $\hat{\sigma}[M_1] = 136.2$ from (3.5), and M_2 which relates ΔF_t to $\Delta D_t, \hat{Q}_{t-1}, S_{\{1955\}}, 1_{\{1987\}}, S_{\{1995\}}$ with $\hat{\sigma}[M_2] = 139.5$ from (3.3) augmented by the indicators. The instruments are the regressors of the joint nesting model: $1, F_{t-1}, t, \Delta D_t, S_{\{1955\}}, 1_{\{1987\}}, S_{\{1995\}}, \hat{Q}_{t-1}$ with $\hat{\sigma}[\text{Joint}] = 134.0$. The resulting test statistics are shown in Table 3.1. The results reject M_2 in favour of M_1 , though not decisively. Nevertheless, the combined evidence is consistent with a long-run trend decrease that is an approximation to many safety improvements, despite increased distances driven causing more fatalities.

Test	Form	M_1 vs. M_2	Form	M_2 vs. M_1
David Cox (1962)	N[0,1]	−1.96*	N[0,1]	−3.06**
Ericsson (1983) IV	N[0,1]	1.70	N[0,1]	2.58**
Denis Sargan (1964)	$\chi^2(1)$	2.68	$\chi^2(3)$	7.13
Joint model	F(1,54)	2.77	F(3,54)	2.57

Table 3.1: Encompassing test statistics.

Overall, the illustration highlights some of the difficulties that can arise when not commencing from a sufficiently general model to embed the local data generation process, albeit that formalizing the effects of the many technological, social, educational and legal changes affecting road safety would be hard. In wide-sense non-stationary processes, even well-established tests like

those for a unit root may implicitly make untenable assumptions, such as a non-changing degree of integration, leading to misinterpretations.

We now turn to the reasons for seeking to apply our tools to climate modeling.

4

A brief excursion into climate science

The Earth's climate depends on the balance between the sun's incoming radiation and the heat loss back to space. Short-wave radiation from the Sun enters the Earth, warms the planet's land and sea surfaces, then heat is radiated back through the atmosphere to space. However, greenhouse gases like CO₂ absorb some of that long-wave radiation en route, and that is then re-radiated, with some being directed back towards the planet's surface. Consequently, higher concentrations of such greenhouse gases will increase the extent of re-radiation, raising temperatures.¹ The sun has itself warmed since its formation, increasing the radiation reaching the Earth over geological time, but has been relatively stable for the epoch of human (i.e., homo sapiens) existence.

The composition of the Earth's atmosphere has also changed greatly over geological time, altering radiation balances. That atmosphere currently comprises the four major components of water vapor, nitrogen (almost 80% of dry air), oxygen (about 20%) and carbon dioxide. There are also smaller volumes of related greenhouse gases including nitrous oxide, N₂O, which is becoming an increasing component of greenhouse gases, methane, CH₄,

¹Gunnar Myhre, Arne Myhre, and Frode Stordal (2001) review the radiative forcing of different greenhouse gases, and e.g., Kaufmann *et al.* (2013) show that the stochastic trend in global temperature is driven by the stochastic trends in anthropogenic forcing series.

and various chlorofluorocarbons, CFCs, halons and other halocarbons.² These atmospheric components differ greatly in their roles in retaining heat in our planet. Water vapor is crucial, created by evaporation and providing rain, but also retaining heat: as the climate warms, more evaporation will lead to greater heat retention and more cloud cover, but that in turn will reflect back some of the incoming radiation.

The roles in atmospheric heat retention of water vapor, carbon dioxide, and dry air (mainly nitrogen and oxygen) were elegantly demonstrated by Eunice Foote in 1856, who filled separate glass jars with them to compare how they heated when placed in sunlight. She showed that the flask containing water vapor heated more than one with dry air, but that carbon dioxide heated considerably more and took far longer to cool. Her simple experiment could be demonstrated to school children to explain why CO₂ emissions are causing climate change, leading to the worrying trend in global temperatures. Foote's research predated that of the confirming and more exact experiments by John Tyndall.³ The physics of greenhouse gases was established by Svante Arrhenius (1896), who argued that the atmospheric change in temperature was proportional to the logarithmic change in CO₂ concentrations: Spencer Weart (2010) provides a history of the discovery of global warming.

Next, nitrous oxide emissions have doubled in the last 50 years (see e.g., U.S. Energy Information Administration, 2009) and are about 300 times more potent per molecule than carbon dioxide as a greenhouse gas. Catalytic converters for car exhaust emissions oxidize and reduce nitrogen oxides and carbon monoxide to CO₂, nitrogen and water, but can produce nitrous oxide when the exhaust system is cold or malfunctioning. Excess fertilizer on fields that runs off into rivers and lakes also releases nitrous oxide. The Black Sea is an indication of how fast such problems can happen due to excess nitrogen and phosphates from run-off, with the anoxic layer reaching the surface and killing its fish (see Laurence Mee, 2006), fortunately now tackled.

Dieter Lüthil, Martine Le Floch, Bernhard Bereiter and Thomas Blunier

²Despite being non-reactive, CFCs gained notoriety for destroying the ozone layer by breaking down from ultraviolet radiation in the upper atmosphere. Although the Montreal Protocol has led to a major reduction, they remain powerful greenhouse gases by absorbing infrared radiation, as are substitutes like HCFCs. Nitrous oxide also now poses a serious problem for the ozone layer: see A.R. Ravishankara, John Daniel and Robert Portmann (2009).

³<http://www.climatechangenews.com/2016/09/02/the-woman-who-identified-the-greenhouse-effect-years-before-tyndall/>.

et al. (2008) establish that methane concentrations in the atmosphere are now double the levels seen over the past 800,000 years, and that ‘strong correlations of methane and CO₂ with temperature reconstructions are consistent’ over that period. As noted in Hendry (2011), go to any lake in northern Siberia, drill a hole in the ice and hold a flame over it—but jump back quickly to avoid being burned by the methane explosively catching fire. Melting the permafrost in Siberia’s tundra could lead to a marked increase in global temperatures (see Vaks *et al.*, 2020). Methane is about 20 times as powerful as CO₂ as a greenhouse gas, with a half-life in the upper atmosphere of around 15 years, as it gradually gets converted to CO₂, so has a second effect on climate. Current estimates of the world’s methane hydrates are over 6 trillion tonnes which is roughly twice the carbon content of all fossil fuels.⁴

The proportions of all these components of the atmosphere have been greatly altered over geological time by many natural processes, as well as more recently by humanity. These natural processes include the evolution of photosynthesis converting CO₂ into energy and releasing oxygen, thereby cooling the planet once iron oxidation was completed; massive volcanism, releasing vast volumes of greenhouse gases and shorter-term cooling particulates; and tectonic plate movements altering the locations and depths of the oceans, which also play a key role through both heat and CO₂ absorption and release to maintain a temperature balance with the Earth’s atmosphere. The consequences of these changes are the subject of §4.2. First we digress to consider the possibility that despite the magnitude of the planet and its apparently vast oceans, human behavior has become a geological force, reflected in the suggestion of renaming the current epoch from Holocene to Anthropocene.

4.1 Can humanity alter the planet’s atmosphere and oceans?

To answer the first part of the question in the title, we just need to look at a satellite photograph of the atmosphere as in Figure 4.1. This shows that our atmosphere is but a thin blue line atop the Earth, relatively as thick as a sheet of paper round a soccer ball. Almost everyone knows that the peak of Mt Everest at just over 29,000 feet (8,48m) is above a height where there is sufficient oxygen to sustain life, yet most people seem to act as if the

⁴Methane hydrates are crystalline solids in which a gas molecule is surrounded by a cage of water molecules that act as ‘cement’ at low temperatures and high pressure.

atmosphere is almost unbounded. Given how little air there is, it should not come as a surprise that human economic activity can alter the composition of the atmosphere to influence the Earth's climate, as our emissions of a variety of greenhouse gases like CO_2 , N_2O , CH_4 , and various CFCs have done—and are increasingly doing so. Indeed, sulphur hexafluoride (SF_6), widely used in the electricity industry as an insulator to prevent short circuits and fires, is a very long-lived gas estimated to be more than 20,000 times worse over a century as a greenhouse gas than CO_2 , and is already leaking into the atmosphere at almost double the rate of 20 years ago: see Matthew Rigby, Jens Mühle, B. R. Miller and Ronald Prinn *et al.* (2010).



Figure 4.1: Satellite photograph of the Earth's atmosphere available from NASA.

4.1.1 Anthropogenic increases in greenhouse gases

Adding greenhouse gases to the atmosphere increases global temperatures, a growing concern reflected in the Paris Accord agreement at COP21 to seek to limit temperature increases to less than 2°C , and “to pursue efforts to limit it to 1.5°C ”. The recent Special Report by the Intergovernmental Panel on Climate Change (IPCC: <https://www.ipcc.ch/sr15/>) emphasises that the latter

is still just achievable, but that rapid action is required if it is to be achieved. A comparison of current with estimated atmospheric CO₂ levels over the last 800,000 years of ice ages is informative as shown in Figure 4.2. Over the long period shown, atmospheric CO₂ varied over the range of roughly 175 parts per million (ppm) to 300ppm, for reasons addressed in Chapter 6.

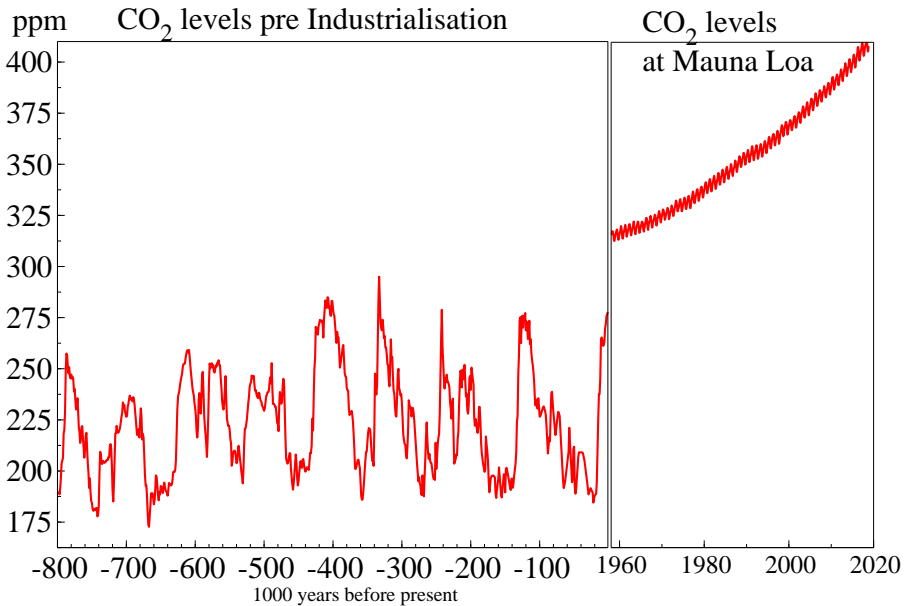


Figure 4.2: Atmospheric CO₂ levels pre-Industrialization and recent Mauna Loa recordings.

However, the recent CO₂ records collected at Mauna Loa in Hawaii by Charles Keeling from 1958 (see Keeling, Robert Bacastow, Arnold Brainbridge and Carl Ekdahl *et al.* 1976, and Eric Sundquist and Keeling, 2009), show a strong upward trend from more than 300ppm to now exceed 400ppm: see CO₂ Program Scripps (2010). These increases in atmospheric levels of CO₂ are clearly anthropogenic in origin, as shown by the different isotopic ratios of CO₂ from using fossil fuels compared to its release from photosynthesis by plants. This matches the attribution of CO₂ emissions to human activity in (e.g.) Hendry and Pretis (2013), as the trend dominates the marked seasonal variation. Plants absorb more CO₂ in their growing seasons and release more as they die back in winter. The marked seasonality which that creates is due to the greater proportion of the planet’s land being in the northern hemisphere.

The climate change resulting from higher CO₂ concentrations has potentially dangerous implications, as highlighted by various IPCC reports,⁵ and many authors including Nicholas Stern (2006), leading to the agreement in Paris at COP21. Malte Meinshausen, Nicolai Meinshausen, William Hare and Sarah Raper *et al.* (2009) analyze the difficulties of even achieving 2°C as annual changes are increasing, so Paris COP21 has not yet even slowed the growth of CO₂ emissions.

How can we be sure human activity is responsible? Here is how. There have been trend increases in our use of fossil fuels and in deforestation. Since Hans Suess (1953) it has been known that radioactive isotope carbon-14 is created by cosmic rays in the upper atmosphere hitting CO₂ molecules, after which the radioactivity gradually decays. Since coal and oil deposits were laid down hundreds of millions of years ago, their radioactivity has dissipated, so carbon dioxide released by their burning lacks this radioactive isotope. The changing ratio of the isotopes of carbon detected in the atmosphere would point directly at anthropogenic sources. Unfortunately, atmospheric nuclear explosions have radically altered that ratio, making it inapplicable as an indicator of human fossil fuel consumption. However, the ratio of another heavier isotope, carbon-13, relative to carbon-12 in atmospheric CO₂ is also larger than its ratio in fossil fuels, and is not affected by nuclear tests. Consequently, if additional CO₂ output is due to burning fossil fuels, the ratio of carbon-13 to carbon-12 should be decreasing—as is occurring. Moreover oxygen is needed for combustion, and matching the increases in CO₂, atmospheric levels of oxygen have been falling slowly, albeit from a substantial level.

4.1.2 The Earth's available water

Although the oceans seem vast—and seen from space, Earth is a blue sphere—actually collecting all the Earth's water makes but a 'puddle', as Figure 4.3 shows. The spheres shown represent respectively:

- (1) All water on Earth, in the largest sphere over western USA, a mere 860 miles in diameter.
- (2) All fresh liquid water in the ground, lakes, swamps, and rivers in the sphere over Kentucky, just 169.5 miles in diameter.
- (3) Fresh water in lakes and rivers in the tiny sphere over Georgia, only 34.9

⁵See e.g., <https://www.ipcc.ch/report/ar5/wg2/>.

miles in diameter:

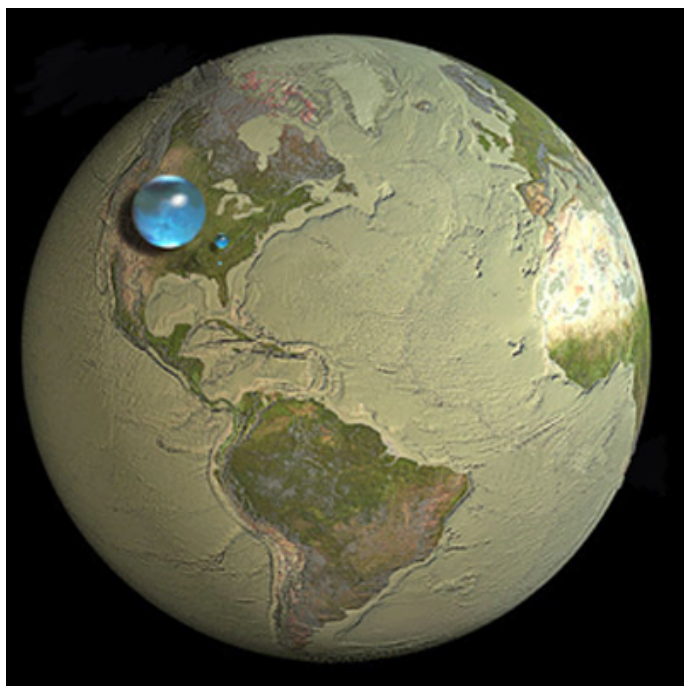


Figure 4.3: Earth's water resources. Credit: Howard Perlman, USGS; globe illustration by Jack Cook, Woods Hole Oceanographic Institution © Adam Nieman.

Imagine heating up these small spheres, or adding millions of tons of plastic or other polluting substances to them, or make them absorb gigatonnes of CO_2 . It is easy to see how we can affect oceans, lakes and rivers. These tiny spheres may seem to conflict with pictures of apparently endless oceans shown on programs like *Blue Planet*, and other films. But oceans are relatively shallow: the Atlantic is around 2.25 miles deep on average. The Pacific is wider and deeper at about 2.65 miles—at its deepest in the Challenger Deep of the Mariana Trench, roughly 6.8 miles down—and holds 170 million cubic miles of water in total, just over half the 330 million cubic miles of water in the largest sphere shown in Figure 4.3.

From the narrower viewpoint of climate change, ocean warming, sea-level rises and creating a weak carbonic acid are the three key issues. The warming climate leads to thermal expansion and to loss of glaciers and ice sheets over land. Based on satellite altimetry, the global mean sea level has been rising

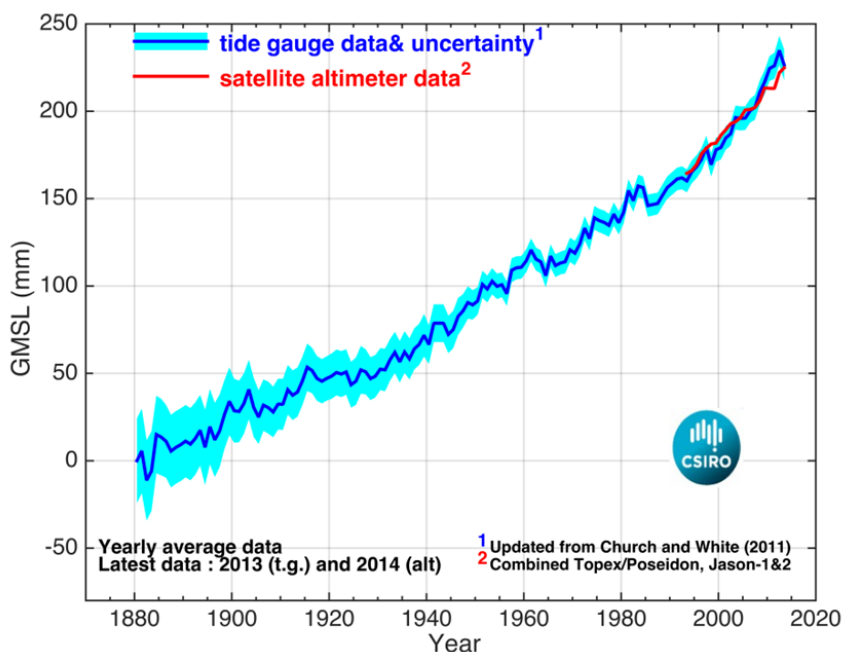


Figure 4.4: Global mean sea level (GMSL) has risen more than 20cm since 1880 and is now rising at 3.4mm p.a. versus 1.3mm p.a. over 1850-1992.

since 1880 as seen in Figure 4.4. Although there is uncertainty around early measures based on tide gauges, these are dwarfed by the trend; and the later satellite altimetry matches the trend for the overlapping period. Not only has that rise accelerated this century to 3.4mm per annum, future rises till about 2050 are inexorable from the oceans' gradual response to atmospheric temperatures already achieved, though later rises depend on how much more greenhouse gas is emitted.

To put that annual rise of 3.4mm in context, dumping into the world's oceans a volume of earth, pebbles and stones equivalent to the top five inches from the 9.1 million square miles of the land area of the United States would raise sea levels by about 3.3 mm—**once**.⁶

Such rises have dangerous implications for populations living near coasts

⁶Testimony by Philip Duffy, President of Woods Hole Research Center before the USA House Committee on Science, Space, and Technology, as reported by *The Washington Post*, 17 May, 2018.

from increases in the heights and frequencies of extreme sea levels from combinations of high tides, wind-driven waves and storm surges (see *inter alia*, Michalis Vousdoukas, Lorenzo Mentaschi, Evangelos Voukouvalas and Martin Verlaan *et al.*, 2018). What were rare flooding events historically will occur more frequently after sea-level rises. Failing adequate advance preparations, the risks to natural and human systems from sea-level rise and extreme sea levels include flooded cities, beach and cliff erosion, biodiversity loss, territorial losses (e.g., small island nations), displacement of people, harm to physical and psychological health and well-being, and stress upon energy, transport and communication systems (see e.g., Jackson and Hendry, 2018).⁷

Confirming that part of the reason for the rising sea level is that Earth's oceans are warming, recent measurements reveal that the ocean heat content is higher than previously estimated and goes to a greater depth: see Figure 4.6. Laure Zanna, Samar Khatiwala, Jonathan Gregory and Jonathan Ison *et al.* (2019) and Lijing Cheng, John Abraham, Zeke Hausfather and Kevin Trenberth (2019). Not only will thermal expansion be greater, warmer oceans threaten sea life, from coral reefs (whose destruction is further exacerbated by increasing carbonic acid) through a rise in the chemocline between oxygenated water above and anoxic water below (see e.g., Anthony Riccardi, Lee Kump, Michael Arthur and Steven D'Hondt, 2007).

4.2 Climate change and the 'great extinctions'

A number of major extinction events have occurred over geological time when many of the world's life-forms ceased to exist, defined as their permanent disappearance from the fossil record. That record is incomplete, as the recent spate of discoveries of fossils of new dinosaur species in China attests; and disappearance is not final, as revealed by the curious tale of the rediscovery of a living coelacanth, previously thought extinct for about 70 million years (see Keith Thomson, 1991, for that exciting story). Although dating is difficult for very distant events, and species go extinct intermittently in the absence of major events, the vast numbers of marine and land species vanishing over several relatively short geological time intervals is convincing evidence of

⁷Worse, it seems satellite measurements have overestimated coastal heights so an extra 300 million people will be affected by sea-level rises by 2050: see <https://sealevel.climatecentral.org/news/new-study-triples-global-estimates-of-population-threatened-by-sea-level-rise/>

'great extinctions'.

'Mass extinctions' occurred even before land life evolved, including in the pre-Cambrian era, before about 600 mya (million years ago). That was so severe that almost all micro-organisms vanished, possibly from large scale cooling and global glaciation (called 'snowball Earth': see Paul Hoffman and Daniel Schrag, 2000). The Cambrian period then appears to have suffered four more major marine extinctions, possibly also from global sea cooling.

These early disasters were followed by five others over the next 500 million years, shown by the labelled peaks in Figure 4.5. Figure 4.5 shows one estimate of the percentage of species vanishing from the fossil record, a clear demonstration of the fragility of life forms to the major climate changes that occurred at the boundaries of the Ordovician, Devonian, Permian, Triassic and Cretaceous periods.

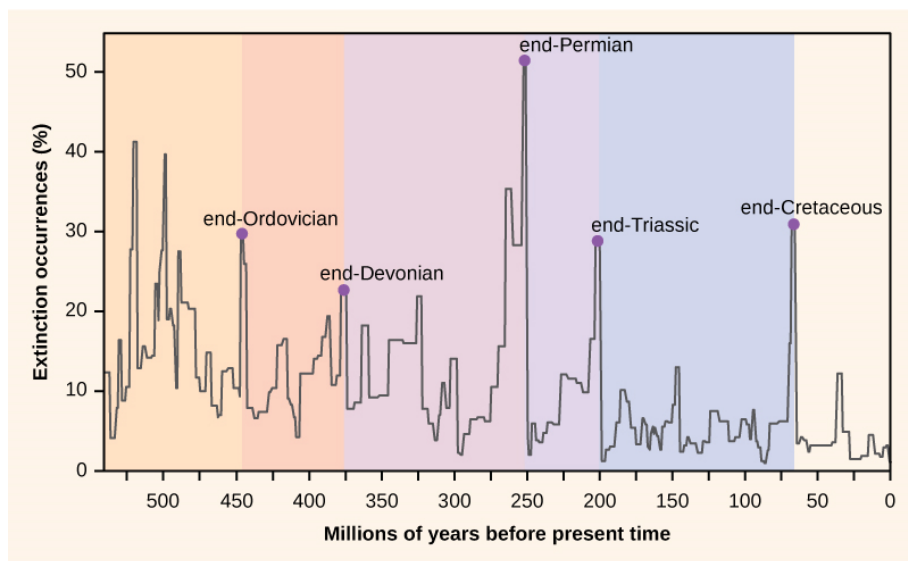


Figure 4.5: Fossil record disappearances showing an 'extinctions timeline' at period endings. Source: <https://courses.lumenlearning.com/wm-biology2/chapter/mass-extinctions/>

The first of the 5 mass extinctions came at the end of the Ordovician period, approximately 440 mya (dates rounded for simplicity), again probably from global cooling, possibly followed by warming. The next, about 375 mya, occurred toward the close of the Devonian period, probably from the rapid

spread of plant life on land reducing atmospheric CO₂ by photosynthesis.

The third mass extinction at the Permian–Triassic (P/Tr) boundary around 250 mya was the worst, leading to major losses of both ocean life and plants, animals, and insects on land (see *inter alia*, Douglas Erwin, 1996, 2006).⁸ Explanations include the formation of massive flood basalts from extensive and prolonged volcanic eruptions, called large igneous provinces (LIPs).⁹ The LIP in Siberia forming at that time covered in excess of 2 million square kilometers with global temperatures about 6°C higher than now. Methane hydrates released from relatively shallow continental shelves by the formation of the LIP are a possible cause (see e.g., Ezat Heydari, Nasser Arzani and Jamshid Hassanzadeh, 2008). Magma pouring into seas may also have disturbed deep ocean levels (see Peter Ward, 2006). In particular, underwater volcanism can induce oxygen deficiency when LIPs disrupt the ocean conveyor belts (see Timothy Bralower, 2008). Extinctions from oceanic warming can become drastic if the chemocline between oxygenated water above and anoxic water below reaches the surface (see e.g., Riccardi *et al.* 2007). Then archaea and anaerobic bacteria, such as green-sulfur bacteria, proliferate and can generate vast quantities of hydrogen sulfide (H₂S), which is almost as toxic as hydrogen cyanide. As with CO₂, hydrogen sulfide is heavier than air, so can accumulate on the surface.¹⁰ There is evidence that the chemocline rose during the end-Permian extinction, with a large increase in phototrophic sulfur bacteria replacing algae and cyanobacteria, consistent with huge loss of ocean life.¹¹ Ocean circulation may also have slowed, or even stopped, from a lack of ice at the poles. While that initially affects marine life, CO₂ dissolves more readily in cold water, and is released when water warms (as from an open glass of sparkling water): a massive overturn of cold oceanic water can release large

⁸Michael Rampino and Shu-Zhong Shen (2019) present evidence that there was a further mass extinction approximately 260 mya associated with the eruption of the Emeishan flood basalts in China, which time corresponds to the ‘bump’ before the end-Permian great extinction in Figure 4.5.

⁹These can form layered hills looking a bit like stairs, called Traps.

¹⁰As noted above, recent behavior of the Black Sea shows how quickly a switch in the chemocline can lead to the anoxic layer reaching the surface (see Mee, 2006). The eruption of sulfur bacteria round China’s southern coast just before the 2008 Olympics is a more worrying example, although that too might have had similar local causes.

¹¹H₂S also attacks the ozone layer if driven to the upper atmosphere (possibly by volcanism), reducing protection from solar radiation.

quantities of CO₂, warming the atmosphere.

The fourth extinction at the end of the Triassic period, roughly 200 mya, helped the dinosaurs to emerge as a dominant life form in the Jurassic (see e.g., Stephen Brusatte, 2018). The cause is possibly the formation of a massive LIP called the Central Atlantic Magmatic Province, covering over eleven million square kilometers. The mass extinction could have been due to extensive CO₂ emissions from the dissociation of gas hydrates inducing intense global warming, as there is some evidence of a rise in atmospheric CO₂ near the Triassic–Jurassic boundary, or alternatively from sulfur dioxide emissions leading to cooling, but doubt remains about the cause and mechanism.

The fifth and perhaps best known major extinction occurred approximately 60 mya at the Cretaceous–Tertiary (K/T) boundary (now called Cretaceous–Paleogene, K-Pg), when the family of dinosaurs called saurischia went extinct. This extinction could plausibly be attributed to a meteoric impact, matching traces of iridium found between rocks separating dinosaur from mammalian epochs, and the discovery of the Chicxulub crater near the Yucatan peninsula. Even so, volcanism may also have played a role, as that time saw the formation of another LIP in India (the Deccan Traps), where a prolonged magma extrusion covered approximately 100,000 square kilometers by about 160 meters deep, and global temperatures that were about 4°C higher than currently (see e.g., Donald Prothero, 2008, for an evaluation).¹² Ward (2006) shows that the extinction at the end of the Triassic began when atmospheric CO₂ was just above 1,000 ppm, and that at the K-Pg boundary when CO₂ was just under 1,000 ppm, both far above the level of under 300 ppm at the end of the last ice age, and just over 400 ppm now.

The fossil record over the past 520 million years as presently known shows that terrestrial and marine biodiversity was related to sea-surface temperature, with biodiversity being relatively low during warm periods (see Peter Mayhew, Gareth Jenkins and Timothy Benton, 2009, and Andrew Clarke, 1993). Climate change, manifested by excessive global warming or cooling, has been a cause in all the above large-scale species extinctions. Indeed, it is difficult to imagine any other mechanism that would simultaneously exterminate both land and sea life other than large shifts in global cooling or warming. The key is *change*:

¹²The meteor may have struck an underwater oil deposit, ejecting huge quantities of smoke and soot into the atmosphere; also the impact could have played a role in the volcanism in India.

though not anthropogenic, both directions have led to major losses of species from the fossil record, although since life is abundant today, some species have clearly always managed to survive and evolve. However, a climate-change induced mass extinction could threaten the lives of millions of humans if species crucial to modern food chains were to vanish.

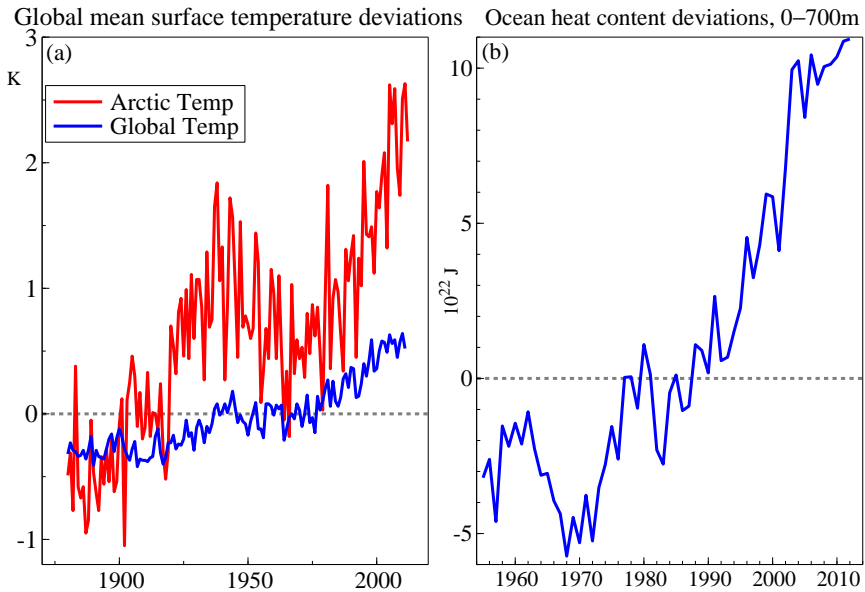


Figure 4.6: (a) Global and Arctic mean surface temperature deviations in degrees K since 1880; (b) global ocean heat content to a depth of 700m over 1957–2012.

Since all the great extinctions seem due to global climate change, albeit from possibly different causes, and since greenhouse gases lead to temperature changes, what is the evidence for the accumulation of CO_2 equivalents in the atmosphere? As discussed above, the records collected at Mauna Loa in Hawaii (see Keeling *et al.* 1976 and Sundquist and Keeling, 2009), show an unequivocal upward trend, with large seasonal variations around it. Figure 4.2 showed the recent increases in CO_2 levels from the low 300 parts per million (ppm) to near 400ppm since 1958. As a consequence global mean surface temperatures have been rising as seen in Figure 4.6(a), especially in the Arctic where a feedback from ice melting lowers albedo and accelerates warming. Global ocean heat content to a depth of 700m shown in Figure 4.6(b) has been

rising rapidly over 1957–2012.

The oceans currently contain approximately sixty times more carbon than the atmosphere, and from a geological perspective, some of that CO₂ can be exchanged quite rapidly between the atmosphere and oceans. Moreover, although oceans can probably absorb more CO₂ at present, that may have adverse effects for marine life (see Richard Stone, [2007](#)): acidification slows the growth of plankton and invertebrates, which are basic to the ocean food chain. Lower pH levels could prevent diatoms and coral reefs from forming their calcium carbonate shells (e.g., just from lowering the current pH level of 8.1 to pH of 7.9). Moreover, while oceans rapidly absorb CO₂ initially, much is evaporated straight back into the atmosphere (again think sparkling water left unsealed), and while later recycled, takes a long time before much is stored in deep ocean layers. So how has humanity created this potentially dire situation of adverse climate change? The next Chapter addresses that development.

5

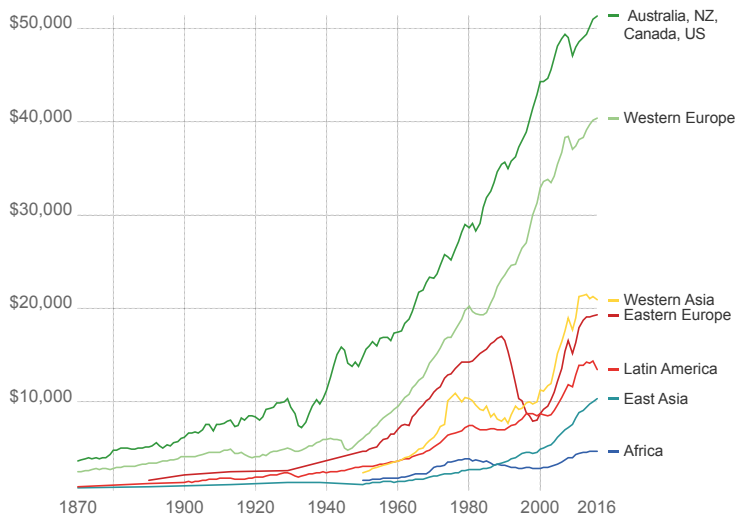
The Industrial Revolution and its consequences

The ‘Industrial Revolution’ began in the UK in the mid-18th Century for reasons well explained by Robert Allen (2009, 2017). While its antecedents lay several centuries earlier in the many scientific, technological and medical knowledge revolutions, the UK was the first country to industrialize on a large scale. The startling consequences of that step can be seen everywhere 250 years later: real income levels are **7–10** fold higher per capita as shown in Figure 5.1, many killer diseases have been tamed, and longevity has approximately doubled. The evidence recorded in <https://ourworldindata.org/economic-growth> shows great increases in living standards in many countries albeit these are far from evenly shared. Nevertheless, the Industrial Revolution and its succeeding developments have been of vast benefit to humanity, raising standards of living for billions of humans far above levels dreamt of by earlier generations: see the excellent discussion in Nick Crafts (2003) who demonstrates that the average individual living in the UK today would be unwise to swap their life for that of even one of the richest people several centuries ago.

Unfortunately, an unintended consequence of the Industrial Revolution was an explosion in anthropogenic CO₂ emissions. This occurred because the main source of non-human and non-animal power at the time came from steam engines fired by coal, following improvements to the earlier engine of

Average real GDP per capita across regions

The measures are adjusted for inflation (at 2011 prices) and also for price differences between regions (multiple benchmarks allow for cross-regional income comparisons).



Source: Maddison Project Database (2018)

OurWorldInData.org • CC BY-SA

Figure 5.1: Increases in average real GDP per capita across major regions from 1870. Source: <https://ourworldindata.org/>.

Thomas Newcomen by the separate condenser invented by James Watt, as well as his enhancing its versatility to generate rotary power. Moreover, at the time, coal was relatively available in the UK, and quite cheap to mine, so although transporting such a heavy substance by land was expensive, sea routes were widely used, and from the opening of the Bridgewater Canal in 1771 greatly reducing costs, a boom in canal building occurred.¹

However, the transport situation was even more radically altered at the ‘Rainhill trials’ in 1829 won by George and Robert Stephenson’s locomotive *Rocket* demonstrating the speed and power of steam-driven trains, a development that soon went global.² With improvements in steam engines, rail transport came to dominate for more than a century and of course produced volumes of CO₂ in the process.

¹ An invaluable spin off of this boom was the wonderful 1819 geological map by William Smith of the rock strata of the UK: see e.g., Simon Winchester (2001).

² See e.g., Richard Fullerton, Bruce Linster, Michael McKee and Stephen Slate (2002). As argued in Hendry (2011), prizes for methods to reduce, store or extract CO₂ from the atmosphere deserve serious consideration.

Oil consumption added considerably to CO₂ emissions from the late 19th Century with the invention of gasoline and diesel powered internal combustion engines for cars, and replacing coal in ships as the heat source for steam engines. In the 20th Century, air travel has further increased the demand for oil products, as have various chemical industries.

Nevertheless, the victory of coal and later oil was not guaranteed. Although electricity was known from ancient times as a shock that electric fish could deliver, and from ‘static electricity’ created by rubbing objects, the first understanding (and the English name) only came after William Gilbert (1600). Following many discoveries through its links to lightning by Benjamin Franklin, creating batteries by Alessandro Volta, and Hans Oersted’s finding that an electric current produces a magnetic field, the key breakthrough was Michael Faraday’s electric motor in the early 1830s, linking electricity with a moving magnet that allowed electricity to be generated as needed (see Stephen Blundell, 2012, for an excellent introduction). The first electricity generator in the UK in 1868 was hydro driven, but from 1882 till recently, coal-fired steam-driven power stations produced most of the UK’s electricity—adding to the already large use of coal in household fires, industry and rail transport.

Not only was hydro-electric power available in the 1860s, the first commercial photovoltaic solar panel was developed by Charles Fritts in 1881, building on the creation by Edmond Becquerel in 1839 of the first photovoltaic cell, a device that converted the energy of light directly to electricity. However, it took till the mid 1950s for really viable solar cells to be created by Bell Labs. Moreover, wind power has been used sporadically for more than 2000 years, growing in use in Iran from the 7th century with windmills, that idea reaching Europe about 400 years later and leading to their widespread use to generate power to grind grain and pump water. The first wind turbine to generate electricity was built by James Blyth in 1887, and by the 1930s wind-generated electricity was relatively common on US farms.

Finally, electric cars also date back before the 1880s but became a serious mode of transport when Thomas Parker built a vehicle with a high capacity and rechargeable battery. As they were quiet, comfortable, could travel fast for the time, and did not need gears, they were popular in the early 20th century, when internal combustion engines displaced them given their much greater travel range and lower cost after Henry Ford’s mass produced vehicles, again adding to greenhouse gas emissions (and many more noxious substances including

tiny particulates, nitrogen oxides and carbon monoxide).

Thus, not only are there long historical precedents for renewable energy generation and electric vehicles, they were first available in the second half of the 19th century: had these developments come a century earlier and seen the concomitant efficiency improvements achieved recently, coal-fired electric power and petrol cars need not have happened, an issue we return to in §7.13.

5.1 Climate does not change uniformly across the planet

Such large increases as 100 ppm in atmospheric CO₂ seen in Figure 4.2 have warmed the planet as shown in Figure 4.6.

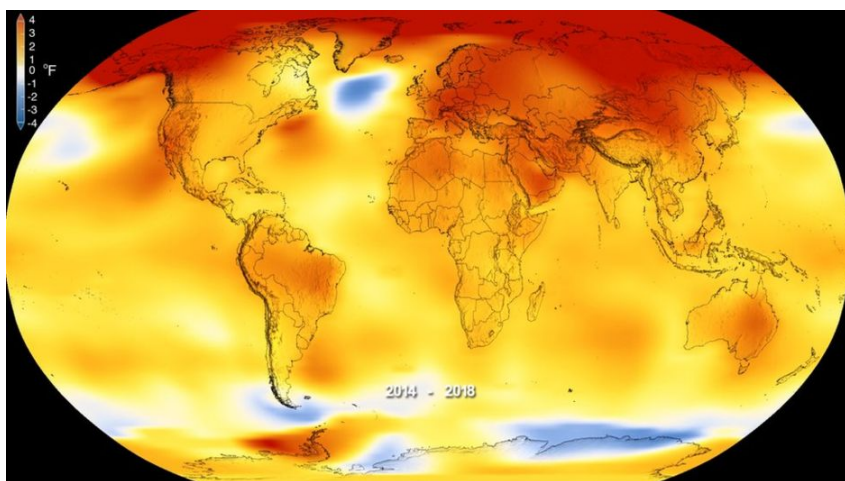


Figure 5.2: Changes in global temperature 2014–2018. Source: NASA

In addition to the faster Arctic warming,³ Figure 5.2 shows that temperature changes have varied between regions. As the tropics receive much more heat from the sun than the poles, that heat is distributed away from the equatorial regions towards the poles. In his excellent video, David Battisti explains that tropical cloud cover plays a key role in that process, and that using the average cloud cover in all the major climate systems greatly reduces

³see e.g. <https://www.msn.com/en-ca/weather/topstories/the-unexpected-link-between-the-ozone-hole-and-arctic-warming/ar-BB1058WF>

the differences between their simulations of future climate.⁴

The Arctic and other northern hemisphere regions have warmed the most over the period shown, but many parts of the planet have seen little change, whereas some ocean and Antarctic areas have cooled. Thus, the key is climate **change** induced by the rising temperatures that are fuelled by additional greenhouse gas emissions from human behavior, primarily burning fossil fuels, especially coal and oil, as well as reduced CO₂ take up from deforestation.

To summarize, we showed above that humanity can easily affect Earth's oceans and its atmosphere, and is doing so. Given the Earth's limited atmosphere, the share of CO₂ has risen by more than 30% to over 400ppm since 1860. Greenhouse gases are transparent to incoming short-wave radiation from the sun, but reflect back some outgoing long-wave radiation, warming the atmosphere, land and the oceans which respond to balance temperatures, also leading to sea-level rises. Over geological time, climate change has been responsible for the great extinctions, as life cannot adapt to some losses of habitat. Thus, understanding climate change is crucial to tackling its likely consequences. The large climate systems model how the Earth responds to changes in greenhouse gases, but the emissions thereof are a function of economic, social, technological, and political behavior. Empirical modeling is an essential addition which Chapter 2 described, but as we also discussed, can be prone to important difficulties. To illustrate how Climate Econometrics tackles these, we describe in detail econometric modeling of Ice Ages and past climate variability over the last 800,000 years in the next Chapter, then UK annual CO₂ emissions 1860–2017 in Chapter 7.

⁴<https://www.oxfordmartin.ox.ac.uk/videos/from-global-to-local-the-relationship-between-global-climate-and-regional-warming-with-prof-david-battisti/>

6

Identifying the causal role of CO₂ in Ice Ages

While many contributions led to the discovery of massive past glaciation on land, that by Louis Agassiz (1840), based on the contemporaneous movements of glaciers in his native Switzerland and using those to explain a number of previously puzzling features of the landscape in Scotland, was a major step forward in understanding the variability of past climate. Agassiz conceived of a ‘Great Ice Age’, an intense, global winter lasting ages, rather than multiple Ice Ages as now, but Archibald Geikie (1863) discovered plant fragments between different layers of glacial deposits, implying that sustained warm periods separated cold glacial periods in prehistory. The calculations by James Croll (1875) using just the variations in the Earth’s orbit then gave a theoretical mechanism for how ice ages could occur and a time line, where the changing albedo of ice coverage helped explain the relative rapidity with which glacial periods switched, although he predicted that the last ice age was older than observed. Recently, Kristina Pistone, Ian Eisenman, and Veerabhadran Ramanathan (2019) have shown that the complete disappearance of Arctic sea ice would be (in temperature terms) ‘equivalent to the effect of one trillion tons of CO₂ emissions’ (roughly 140 ppm) because an open ocean surface typically absorbs approximately six times more solar radiation than a high albedo surface covered with sea ice. Such an effect reducing ocean ice as the

climate gradually warmed after the peak of glacial extent would accelerate melting, and conversely for cooling. Croll's research was later amplified by Milutin Milankovitch (1969) (originally 1941) who calculated solar radiation at different latitudes from changes in obliquity and precession of the Earth as well as eccentricity. Milankovitch also corrected Croll's assumption that minimum winter temperatures mattered, to show that cooler summer maxima were more important in leading to glaciation.

Even a century after Agassiz, there was limited evidence to support such ideas and the timings of glacial episodes. However, these general explanations have since been corroborated by many empirical observations of past oceanic and atmospheric climate changes: see e.g., John Imbrie, Ed Boyle Steven Clemens and A. Duffy *et al.* (1992). As we show below, an important reason for analyzing what may seem like the distant past is its relevance today. The climate then was little affected by the activities of the various human species on the planet, partly as they were too sparse and partly did not have the technology. Consequently, any links between, say, CO₂ and temperature above the forces from the orbital drivers (which of course are still operating) must have been natural ones, so can help us understand their present interactions when CO₂ emissions are anthropogenic.

There are three main interacting orbital changes over time affecting incoming solar radiation (insolation) that could drive ice ages and inter-glacial periods. These are: (a) 100,000 year periodicity deriving from the non-circularity of the Earth's orbit round the Sun from the gravitational influences of other planets in the solar system where zero denotes circularity (eccentricity: *Ec* below); (b) a 41,000 year periodicity coming from changes in the tilt of the Earth's rotational axis relative to the ecliptic measured in degrees (obliquity: *Ob*); (c) about 23,000 and 19,000 year periodicities due to the precession of the equinox, also measured in degrees, which changes the season at which the Earth's orbit is nearest to the Sun, resulting in part from the Earth not being an exact sphere (*Pr*). These 3 are shown measured at 1000-year intervals in Figure 6.1(a), (b) and (c), together with summer-time insolation at 65° south (*St*) in Panel (d) (see Didier Paillard, Laurent Labeyrie and Pascal Yiou, 1996). The X-axes in such graphs are labelled by the time before the present in 1000-year intervals, starting 800,000 years ago. *Ec* and *St* show two major long-swings pre and post about -325 and within each, a number of shorter 'cycles' of varying amplitudes, levels and durations. *Ob* appears

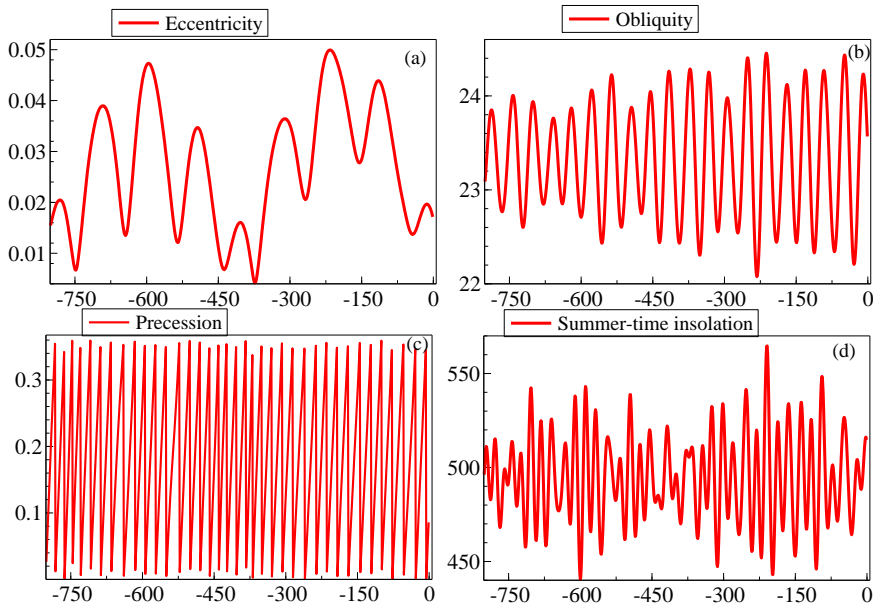


Figure 6.1: Ice-age orbital drivers: (a) eccentricity (*Ec*); (b) obliquity (*Ob*); (c) precession (*Pr*); (d) Summer-time insolation at 65° south (*St*).

to have increased in amplitude since the start of the sample, whereas it is difficult to discern changes in the patterns of *Pr*. The orbital series are strongly exogenous, and most seem non-stationary from shifting distributions, not unit roots.

Orbital variations are not the only forces that affect glaciation. The Earth's energy balance is determined by incoming and outgoing radiation: for a cointegrated econometric model thereof, see Pretis (2019). The role of *St* is to summarize changes in the former, but an exogenous summary measure of outgoing radiation is not clear, as changes that also affect climate include:

- (i) variations in the Sun's radiation output (radiative forcing);
- (ii) atmospheric water vapor and greenhouse gases (e.g., CO₂, N₂O, CH₄);
- (iii) volcanic eruption particulates in the atmosphere;
- (iv) albedo from alterations in ice cover, including from volcanic dust;
- (v) iron in wind-blown dust enabling Southern Ocean storage of CO₂;¹
- (vi) ocean temperatures (which lag behind land);

¹See Pearse Buchanan, Zanna Chase, Richard Matear, Steven Phipps *et al.* (2019).

- (vii) sea levels and induced ocean circulation patterns;
- (viii) cloud cover and its distribution in location and season;
- (ix) changes in the magnetic poles.

Of these, (i), (iii) and (ix) seem strongly exogenous, as do volcanic contributions to (iv), whereas (ii), the rest of (iv) and (v)–(viii) must be endogenously determined within the global climate system by the strongly exogenous drivers. However, anthropogenic greenhouse gas emissions are now ‘exogenously’ changing atmospheric composition: see Jean-François Richard (1980) for an analysis of modeling changes in a variable’s status as endogenous or exogenous, which here would just affect the last few (1000 year) observations.

That the distance from the Sun matters seems rather natural, as such variations change radiative forcing and hence global temperatures. However, the variations due purely to the eccentricity of the orbit are small. Obliquity also must matter: if the Northern Hemisphere directly faced the Sun, ice would usually be absent there; and if it never faced the Sun, would generally be frozen. Precession seems the smallest driving force of these, but interactions may be important: when the Earth is furthest from the Sun and tilts away in the Northern Hemisphere summer, that may cool faster: see Paillard (2010) for an excellent discussion of these interactions. Even so, a problem with the theory that ‘purely orbital’ variations drove ice ages over the last 800,000 years is that the known orbital variations should not result in sufficiently large changes in radiative forcing on the Earth to cause the rapid arrival and especially the rapid ending, of glacial periods: see Paillard (2001). Although St could provide some additional explanation, and in particular seems to help capture changes at peaks and troughs, we decided to only use the strongly exogenous orbital drivers. An equation regressing St just on these and its first lag produced $R^2 = 0.988$, so we leave to the reader the exercise of building a model with St included in the list of variables.²

There are several possible reasons for ‘rapid’ changes in the climate system, remembering that the observation frequency is 1000 years. The extent of Southern Ocean sea ice can substantively alter ocean ventilation by reducing the atmospheric exposure time of surface waters and by decreasing the vertical mixing of deep ocean waters, which Karl Stein, Axel Timmermann, Eun

²For more comprehensive systems that endogenously model measures for all the variables in (iii)–(vii), see Kaufmann and Juselius (2013) and Pretis and Kaufmann (2018). We are also grateful to those authors for providing the data series analyzed here.

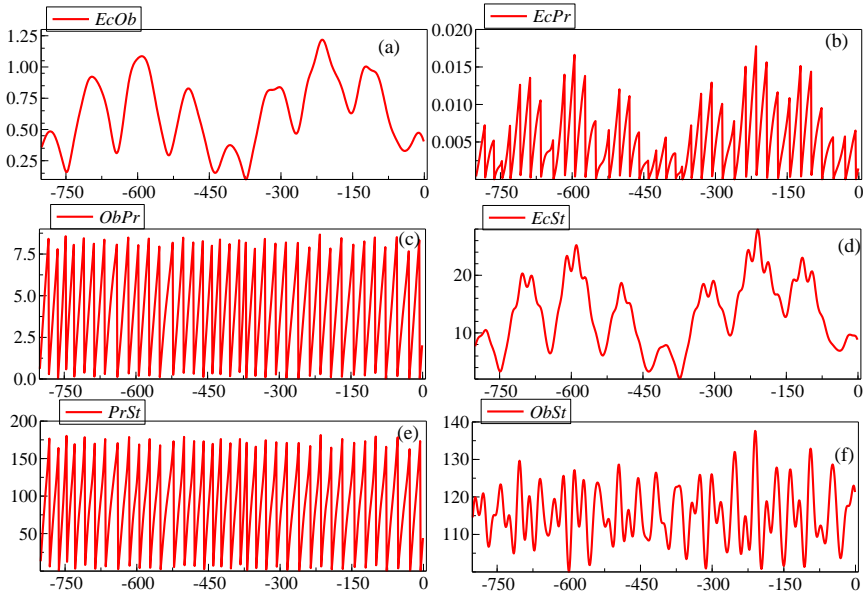


Figure 6.2: Ice-age orbital driver interactions: (a) $EcOb$; (b) $EcPr$; (c) $ObPr$; (d) $EcSt$; (e) $PrSt$; (f) $ObSt$.

Young Kwon, and Tobias Friedrich (2003) show can lead to 40ppm changes in atmospheric CO_2 . Another explanation is the presence of non-linear feedbacks or interactions between the drivers. Thus, Figure 6.2 shows their interactions in Panels (a) [$Ec \times Ob$], (b) [$Ec \times Pr$], (c) [$Ob \times Pr$], (d) [$Ec \times St$], (e) [$Pr \times St$] and (f) [$Ob \times St$] although the model developed here includes only the first three interactions together with the squares to capture non-linear influences. Explaining glaciation over the ice ages has garnered a huge literature, only a small fraction of which is cited here. The possibility of the Northern Hemisphere facing another ice age was still considered in the 1950s as the following quote illustrates:

We do not yet know whether the latest turn in our climatic fortunes, since the optimum years of the 1930s, marks the beginning of a serious downward trend or whether it is merely another wobble... Hubert Lamb (1959)

but by 1982, Lamb (1995) emphasized global warming as the more serious threat to climate stability.

The remainder of the chapter is as follows. Section 6.1 describes the data series over the past 800,000 years, then Section 6.2 models ice volume, CO₂ and temperature as jointly endogenous in a 3-variable system as a function of variations in the Earth's orbit. The general model is formulated in §6.2.1, and the simultaneous system estimates are discussed in §6.2.2. Their long-run implications are described in Section 6.3 with one hundred 1000-year 1-step and dynamic forecasts in §6.3.1. Then, §6.3.2 considers when humanity might have begun to influence climate, and discusses the potential exogeneity of CO₂ to identify its role during Ice Ages. Section 6.4 looks 100,000 years into the future using the calculable eccentricity, obliquity and precession of Earth's orbital path, to explore the implications for the planet's temperature of atmospheric CO₂ being determined by humans at levels far above those experienced during Ice Ages. Finally, Section 6.5 summarizes the conclusions on Ice-Age modeling.

6.1 Data series over the past 800,000 years

A vast international effort over many decades has been devoted to measuring the behavior of a number of variables over the ice ages. Naturally, proxies or indirect but closely associated observables that remain in the ground, ice, oceans and ocean floors are used based on well-established physical and chemical knowledge. Econometricians are essentially mere end users of this impressive research base.

Antarctic-based land surface temperature proxies (denoted *Temp* below) were taken from Jean Jouzel, Valérie Masson-Delmotte, O. Cattani, and Gabrielle Dreyfus *et al.* (2007). The paleo record from deep ice cores show that atmospheric CO₂ varied between 170ppm and 300ppm over the ice ages, where 1ppm = 7.8 gigatonnes of CO₂ (see Lüthil *et al.*, 2008). Ice volume estimates (denoted *Ice* below) were from Lorraine Lisiecki and Maureen Raymo (2005) (based on $\delta^{18}\text{O}$ as a proxy measure). To capture orbital variations, *Ec*, *Ob* and *Pr* and their interactions are conditioned on. All observations had been adjusted to the common EDC3 time scale and linearly interpolated for missing observations to bring all observations on a 1000 year time interval (EDC3 denotes the European Project for Ice Coring in Antarctica–EPICA–Dome C, where drilling in East Antarctica has been completed to a depth of 3260 meters, just a few meters above bedrock (see Frédéric Parrenin *et al.*, 2007).

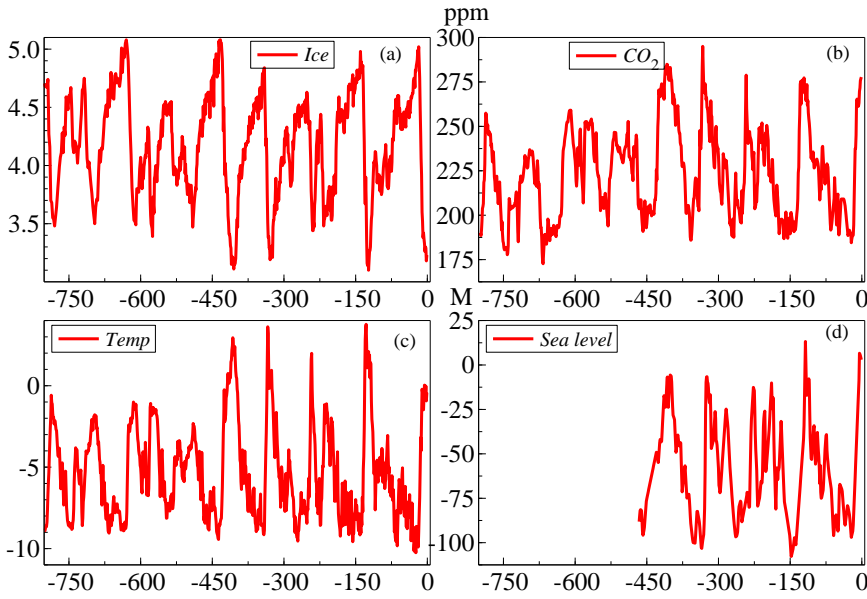


Figure 6.3: Ice-age time series: (a) Ice volume (*Ice*); (b) atmospheric CO₂ in parts per million (CO₂); (c) temperature (*Temp*); (d) shorter-sample sea level changes in meters.

Synchronization between the EPICA Dome C and Vostok ice core measures over the period $-145,000$ to the present was based on matching residues from volcanic eruptions (see Parrenin *et al.*, 2012). The total sample size in 1000 year intervals is $T = 801$ with the last 100 observations (i.e., 100,00 years, ending 1000 years before the present) used to evaluate the predictive ability of the estimated system. Figure 6.3 records a shorter sample of sea level data.³

We focus on modeling *Ice*, CO₂ and *Temp* as jointly endogenous functions of the orbital variables which we take to be strongly exogenous, so feedbacks onto their values from Earth's climate are negligible. The patterns of these time series are remarkably similar, all rising (or falling) at roughly the same times. Figure 6.4 emphasises how close these movements are by plotting pairs of time series: (a) CO₂ and the negative of ice volume (denoted *IceNeg*); (b) CO₂ and *Temp*; (c) *Temp* and *IceNeg*; (d) *IceNeg* and sea level (only for the

³Sea surface temperature data are available from Alfredo Martínez-García, Antoni Rosell-Melé, Walter Geibert and Rainer Gersonde *et al.* (2009) which could help explain oceanic CO₂ uptake and interactions with land surface temperature. Sea level data, based on sediments, can be obtained from M. Siddall *et al.* (2003), over a shorter sample, but are not analyzed here.

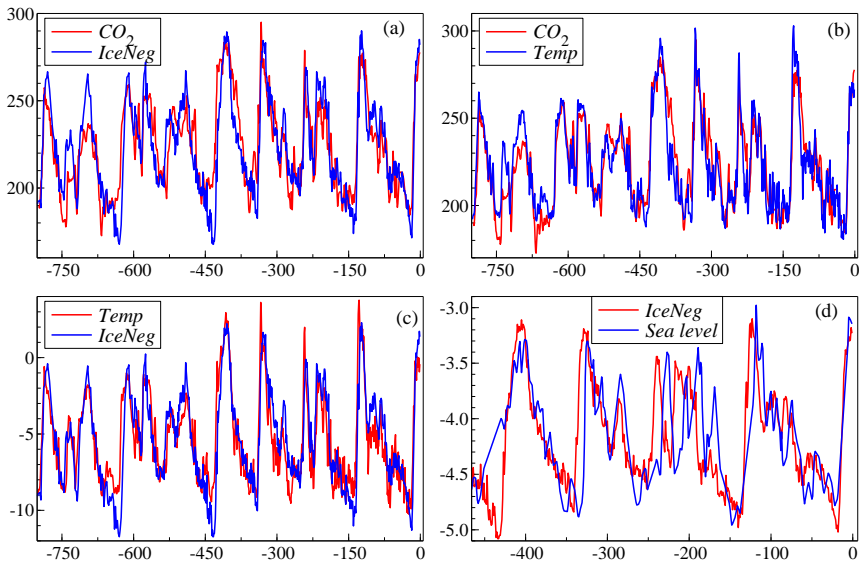


Figure 6.4: (a) CO_2 and the negative of ice volume (IceNeg); (b) CO_2 and temperature; (c) temperature and IceNeg ; (d) IceNeg and sea level (only for the last 465,000 years).

last 465,000 years). Atmospheric CO_2 levels closely track the negative of ice volume, the temperature record and sea level, as do other pairs.

If ice ages are due to orbital variations, why should atmospheric CO_2 levels also correlate so closely with ice volume? David Lea (2004) relates changes in tropical sea surface temperature to atmospheric CO_2 levels over the last 360,000 years to suggest that CO_2 was the main determinant of tropical climate. Conversely, in <https://climateaudit.org/2005/12/18/gcms-and-ice-ages/>, Stephen McIntyre argues that CO_2 should not be treated as a forcing variable in statistical models of ice-age climate, as it is an endogenous response. So is the mechanism not orbital variations, but instead that changes in atmospheric CO_2 levels alter global temperatures which in turn drive changes in ice volume? The answer lies in the deep oceans, in particular, the Southern Ocean, which acts as a carbon sink during cold periods, and releases some of that CO_2 as the planet warms, in turn enhancing cooling and warming: see e.g., Samuel Jaccard, Eric Galbraith, Martínez-García, and Robert Anderson (2016). Thus, the exogenous orbital variations drive temperature, which drives changes in ice volume and in turn CO_2 levels. By modeling the 3-variable

simultaneous-equations system estimated using full information maximum likelihood (FIML: see e.g., Hendry, 1976), treating all three as endogenous, the roles of *Temp* and CO_2 as endogenous determinants of *Ice* can be investigated. The approach used here is described in Section 2.8.

In addition to the many dozens of climatology-based studies, there are several econometric analyses of ice-age data, examining issues of cointegration and the adequacy of using orbital variables as the exogenous explanatory regressors. Kaufmann and Juselius (2010, 2013) analyze the late Quaternary ‘Vostok’ period of four ‘glacial cycles’ and Pretis and Kaufmann (2018) build and simulate a statistical climate model over the paleo-climate record of the 800,000 years of data investigated here. We now turn to system modeling of our three variables of interest.

6.2 System equation modeling of the ice-age data

Our focus is on modeling *Ice* allowing for the endogeneity of *Temp* and CO_2 , with dynamic feedbacks, non-linear impacts of the orbital variables and handling outliers. Consequently, the initial GUM is a VARX(1) for $\mathbf{y}_t = (\text{Ice } \text{CO}_2 \text{ Temp})_t$ conditional on the nine orbital measures and non-linear functions thereof where:

$$\mathbf{z}'_t = \left(\text{Ec } \text{Ob } \text{Pr } \text{EcOb } \text{EcPr } \text{PrOb } \text{Ec}^2 \text{ Ob}^2 \text{ Pr}^2 \right)_t \quad (6.1)$$

with a one-period lag (i.e., 1000 years earlier) on all variables. The lagged values are to capture dynamic inertia: when the ice covers a vast area, that will influence the ice sheet in the next period, even when periods are 1000 years apart. Moreover, that observation length is just 1% of the eccentricity periodicity, so the Earth will still be close to its previous position.⁴ System IIS at 0.1% was implemented with all the continuous variables retained, then after locating outliers, the regressor variables were selected at 1% to create a parsimonious VARX(1), denoted PVARX(1). Next, that system was transformed to a simultaneous-equations model of the PVARX(1), where only variables and outliers that were relevant in each equation were included, and finally contemporaneous links were investigated. Only retaining variables that are

⁴Residual autocorrelation suggests that a second lag or longer may also matter, despite such variables being at least 2,000 years earlier.

significant in a PVARX(1) avoids ‘spurious identification’ from using completely irrelevant variables that are then excluded differently in each equation to apparently achieve the order condition.

6.2.1 The general unrestricted model (GUM)

The GUM in this setting is a dynamic system with strongly exogenous regressors which can be written as:

$$\mathbf{y}_t = \gamma_0 + \mathbf{\Gamma}_1 \mathbf{y}_{t-1} + \mathbf{\Gamma}_2 \mathbf{z}_t + \mathbf{\Gamma}_3 \mathbf{z}_{t-1} + \mathbf{\Psi} \mathbf{d}_t + \boldsymbol{\epsilon}_t, \quad (6.2)$$

where \mathbf{d}_t denotes a vector of impulse indicators selected by system IIS. The difference from single-equation IIS described above is that indicators have to be significant at the target nominal significance level in the system, not just in any one equation therein.

First, all the \mathbf{y}_{t-1} , \mathbf{z}_t and \mathbf{z}_{t-1} in (6.2) are retained without selection when IIS is applied at $\alpha = 0.001$ for $T = 697$ keeping the last hundred observations for out-of-sample forecast evaluation. This led to 35 impulse indicators being selected, the earliest of which was $1_{\{-339\}}$. However, many of these were retained to avoid a failure of encompassing the first feasible GUM, and were not significant at $\alpha = 0.001$.

Table 6.1 records the correlations between the actual observations and the fitted values taking impulse indicators into account, so each variable can be explained in large measure by a model of the form (6.2). Table 6.2 shows the correlations between the residuals of the three equations, with residual standard deviations on the diagonal. There remains a high correlation between CO₂ and *Temp* residuals even conditional on all the orbital variables, but not between those of *Ice* and either CO₂ or *Temp*, although those correlations remain negative.

<i>Ice</i>	CO ₂	<i>Temp</i>
0.981	0.981	0.972

Table 6.1: Correlations between actual and fitted values in the VARX(1).

Next, the other regressors were selected at 1% resulting in a PVARX(1). Again, note that selection decisions are at the level of the system rather than individual equations. Finally, to avoid the spurious identification issue from

	<i>Ice</i>	<i>CO₂</i>	<i>Temp</i>
<i>Ice</i>	0.090	—	—
<i>CO₂</i>	−0.179	5.13	—
<i>Temp</i>	−0.180	0.574	0.711

Table 6.2: Correlations between VARX(1) residuals, with standard deviations on the diagonal.

indicators that were insignificant in the system, any that were also insignificant in every equation were manually deleted from the system, still leaving 32.

6.2.2 The simultaneous system estimates

Because many of the exogenous and lagged variables and impulse indicators were only significant in one equation, we reformulated the system as a simultaneous-equation model. This treats all three modeled variables as endogenous and was estimated by FIML. We then manually eliminated insignificant regressors in each equation in turn. The current dated values of *Temp* and *CO₂* in the *Ice* equation and of *Temp* in the *CO₂* equation were insignificant, but that of *CO₂* was significant in the equation for *Temp*. This delivered the system model in (6.3)–(6.5):

$$\begin{aligned} \widehat{Ice}_t = & 1.43 + 0.860 Ice_{t-1} - 0.020 Temp_{t-1} + 102 Ec_t - 101 Ec_{t-1} \\ & (0.34) \quad (0.015) \quad (0.002) \quad (31) \quad (32) \\ & - 0.040 Ob_{t-1} - 5.07 EcOb_t + 5.05 EcOb_{t-1} - 4.97 EcPr_t \\ & (0.014) \quad (1.30) \quad (1.36) \quad (1.03) \end{aligned} \quad (6.3)$$

$$\begin{aligned} \widehat{CO_{2,t}} = & 218 + 0.853 CO_{2,t-1} + 1.34 Temp_{t-1} + 1400 Ec_t \\ & (32) \quad (0.018) \quad (0.18) \quad (342) \\ & - 3070 Ec_{t-1} - 13.0 Ob_{t-1} + 70.7 EcOb_{t-1} + 0.232 Ob_t^2 \\ & (647) \quad (2.31) \quad (23) \quad (0.047) \end{aligned} \quad (6.4)$$

$$\begin{aligned} \widehat{Temp}_t = & - 2.49 + 0.879 Temp_{t-1} + 0.0080 CO_{2,t} - 301 Ec_t \\ & (0.69) \quad (0.023) \quad (0.0026) \quad (37) \\ & + 22.6 EcOb_t - 9.80 EcOb_{t-1} + 25.5 EcPr_t \\ & (2.45) \quad (1.94) \quad (7.1) \end{aligned} \quad (6.5)$$

The correlations between the actual and fitted values for the three variables in the SEM are virtually identical to those in Table 6.1, consistent with the likelihood-ratio test of the over-identifying restrictions against the PVARX(1) being $\chi^2_{OR}(64) = 69.7$, which is insignificant at even the 5% level. Although

the inertial dynamics play a key role in the three equations, all the eigenvalues of the system dynamics are less than unity in absolute value at (0.97, 0.86, 0.77). The test for excluding all the non-linear functions yields $\chi^2(8) = 155^{**}$, and that for dropping all the impulse indicators $\chi^2(44) = 439^{**}$, both of which reject at any viable significance level.⁵

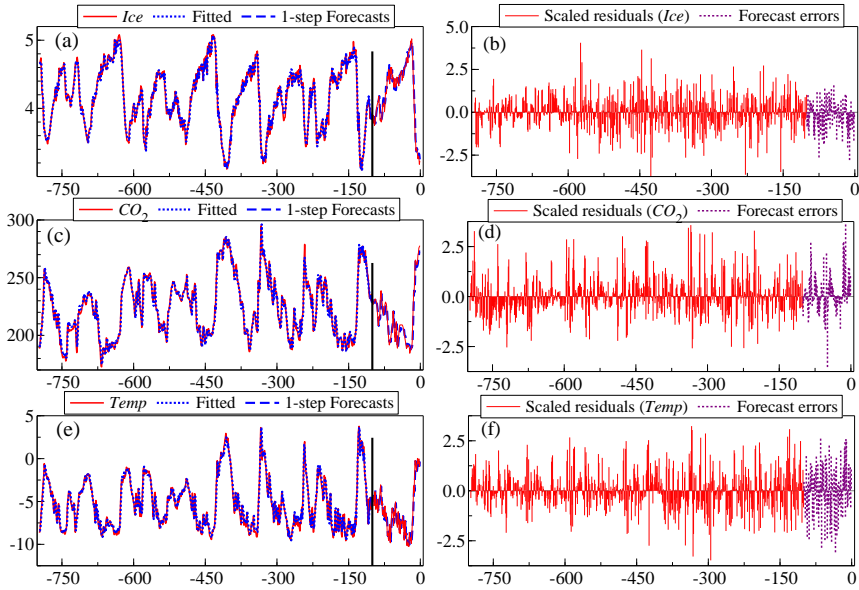


Figure 6.5: Actual, fitted and forecast values, with scaled residuals and forecast errors: (a) & (b) for *Ice* from (6.3); (c) & (d) for CO_2 from (6.4); (e) & (f) for *Temp* from (6.5). The vertical bar at $T = -100$ marks the start of the forecast period.

Figure 6.5 records the actual and fitted values and residuals scaled by their standard deviations for the three equations. The tracking is very close, including over the final 100 ‘out-of-sample’ observations, although the residuals show the occasional outlier: remember that IIS selection was at 0.1% to avoid overfitting. Figure 6.6 reports residual densities with a Normal matched by mean and variance, and correlograms. The densities are relatively close to the Normal for *Ice* and *Temp* after IIS, but less so for CO_2 , probably because the restriction to one lag has left some residual autocorrelation. Most of the formal mis-specification tests rejected, possibly also reflecting the many omitted influences noted above, although most of those seem to be endogenous

⁵The model recording retained impulse indicators is available on request from the authors.

responses as the climate changed, such as dust from wind storms and sea level changes both varying with temperature. In the present context, outliers as represented by impulse indicators could derive from measurement errors in the variables, super-volcanoes either dramatically lowering temperature by erupted particulates, or raising by emitting large volumes of CO_2 , or like wind-blown dust changing the albedo of ice sheets. Most indicators retained for *Ice* were negative around -0.2 , whereas for CO_2 they were primarily positive and around $+15$, and for *Temp* around 2 but mixed in sign. As outliers are relative to the model being estimated, those found here could also represent variables omitted from the system in (6.3)–(6.5).

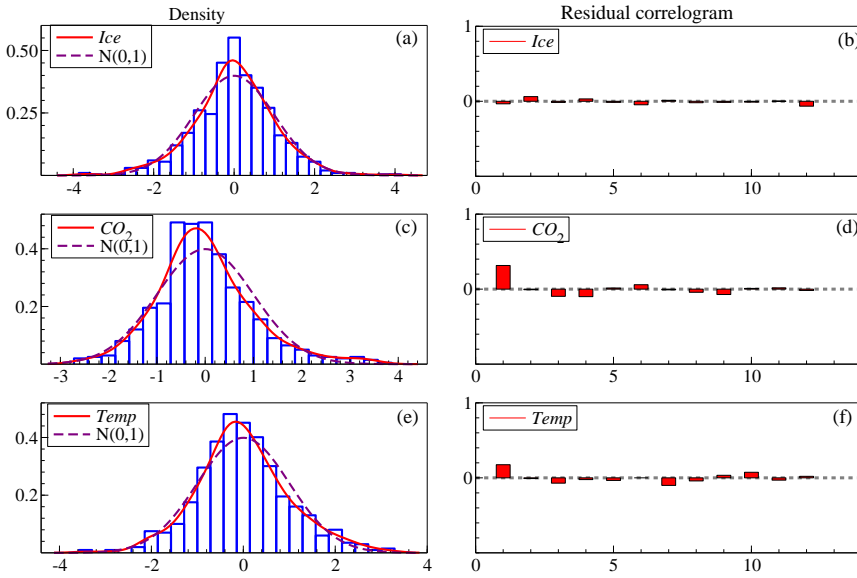


Figure 6.6: Residual densities and correlograms: (a) & (b) for *Ice* from (6.3); (c) & (d) for CO_2 from (6.4); (e) & (f) for *Temp* from (6.5).

Table 6.3 records the correlations between the residuals of the simultaneous equations model, with the residual standard deviations on the diagonal: these are close to those in Table 6.2.

Considering the equations in more detail, the volume of ice in (6.3) depends on its previous level and on previous temperatures, as well as on eccentricity, past obliquity, current and lagged interactions of eccentricity with obliquity, and with current precession. Although *Ec* and *EcOb* appear to enter

	<i>Ice</i>	CO ₂	<i>Temp</i>
<i>Ice</i>	0.086	—	—
CO ₂	−0.173	4.88	—
<i>Temp</i>	−0.184	0.509	0.667

Table 6.3: Correlations of the simultaneous model residuals, with standard deviations on the diagonal.

primarily as changes, the solved long-run outcome in Table 6.4 confirms they both also enter significantly as levels. CO₂ has a similar coefficient on its lag, a **positive** feedback from past temperature, current and past levels of eccentricity, past obliquity and their interaction, and squared obliquity. *Temp* responds to its previous value and **positively** to current CO₂: its coefficient in (6.5) entails that a 100 ppm increase (as seen since 1958) would raise temperatures by 0.8°C *ceteris paribus*. Neither current CO₂ nor *Temp* are significant in the equation for *Ice*; and current *Temp* is insignificant if added to that for CO₂.

6.3 Long-run implications

Table 6.4 solves out the dynamics and lags to express each endogenous variable as a function of the relevant orbital variables. The original coefficients are not easy to interpret as they depend on the units of measurement of the orbital variables, so CO₂ has been divided by 100 and *Temp* by 10 to align numerical coefficient values. Figure 6.7 graphs the computed time series of *Ice*, CO₂ and *Temp* from the long-run relationships in Table 6.4. These graphs include the last 100,000 years before the present which are outside the estimation sample.

	1	<i>Ec</i>	<i>Ob</i>	<i>EcOb</i>	<i>EcPr</i>	<i>ObSq</i>
<i>Ice</i>	−17.3	1162	1.80	−49.3	−111	−0.037
SE	(16.3)	(402)	(1.2)	(17)	(31)	(0.020)
CO ₂	32.0	−837	−2.19	35.6	47.2	0.039
SE	(11.9)	(276)	(0.87)	(11.7)	(19.1)	(0.016)
<i>Temp</i>	19.0	−798	−1.44	34.0	52.0	0.026
SE	(10.8)	(257)	(0.80)	(10.9)	(19.5)	(0.014)

Table 6.4: Long-run solutions as a function of the relevant strongly exogenous orbital variables where CO₂ has been divided by 100 and *Temp* by 10 to align numerical coefficient values.

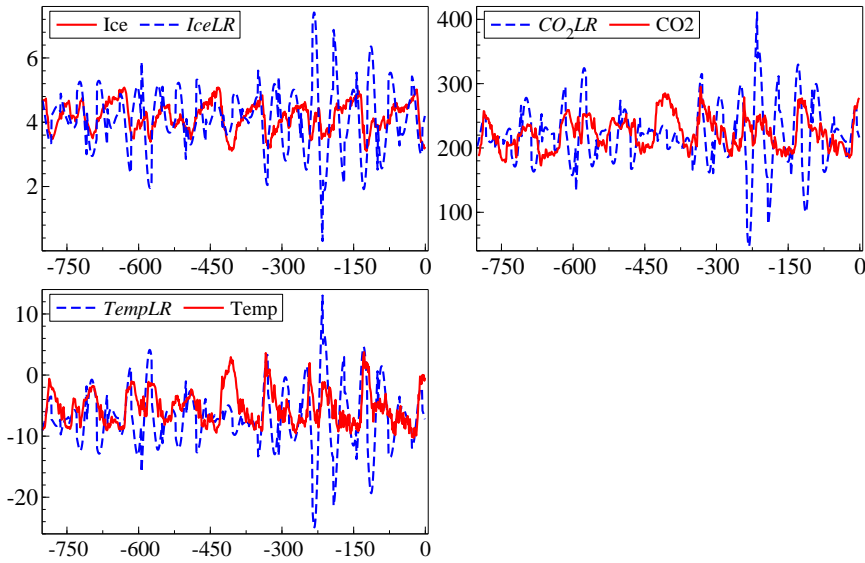


Figure 6.7: Computed time series of *Ice*, CO_2 and *Temp* from the relationships in Table 6.4.

Despite the different coefficients in the three long-run equations, the resulting time series in Figure 6.7 are relatively similar, and the correlations between them all exceed $|0.977|$. These graphs are just recombinations of the orbital drivers weighted by the coefficients in Table 6.4, so reflect the relatively volatile and quiescent periods seen in Figure 6.1. The increase in volatility from about 250,000 years ago is marked, though the inertial dynamics from the lagged dependent variables smooths that over time as seen in Figure 6.5.

6.3.1 1-step and long-run forecasts

Figure 6.8 records the hundred 1-step ahead forecasts with $\pm 2\text{SE}$ error bands based on coefficient estimation variances as well as the residual variances. The second column shows the resulting forecast errors (unscaled).

Table 6.5 reports their RMSFEs which are close to the in-sample standard deviations shown in the following row as $\hat{\sigma}_s$ (IIS) from Table 6.3 for comparison, or no IIS. The model for *Ice* provides a better description of the last 100 observations than the earlier sample, even though the in-sample residual standard deviations were calculated after outliers were removed by IIS. The forecast intervals in Figure 6.8 could be adjusted for the likely presence of out-

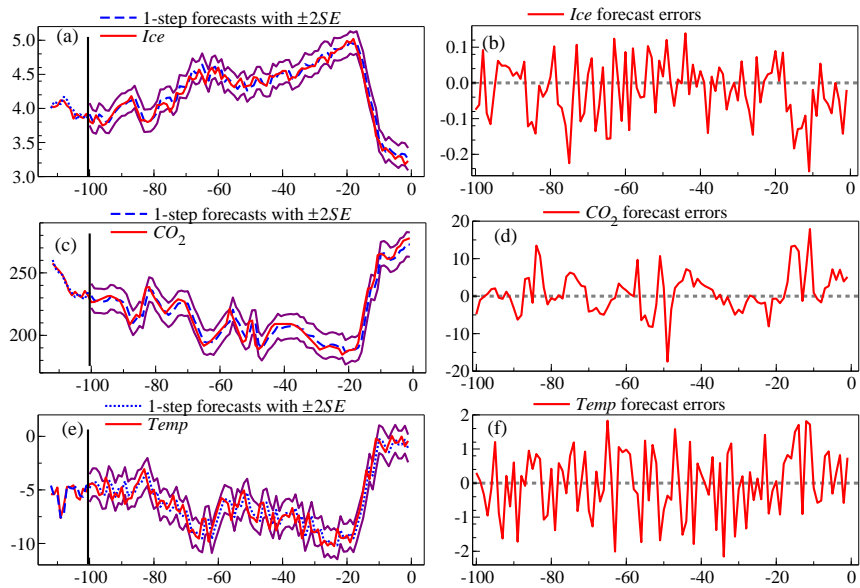


Figure 6.8: A hundred 1-step ahead forecasts at 1000-year measures with forecast intervals at $\pm 2SE$ shown by error bands: (a) for *Ice* from (6.3); (c) for *CO₂* from (6.4); (e) for *Temp* from (6.5). (b), (d) and (f) report the associated forecast errors.

liers in the future at roughly their rate of occurrence in the past by calculating the in-sample residual standard deviations excluding impulse indicators, as those after IIS understate the future uncertainty to some extent. The last row in Table 6.5 reports those ‘no-IIS’ $\tilde{\sigma}$ values.

	<i>Ice</i>	<i>CO₂</i>	<i>Temp</i>
RMSFE	0.084	5.35	0.958
$\hat{\sigma}$ s (IIS)	0.086	4.88	0.667
$\tilde{\sigma}$ s (no IIS)	0.091	5.38	0.748

Table 6.5: 1-step ahead root mean square forecast errors; in-sample model residual standard deviations after IIS; and model residual standard deviations without IIS.

The removal of outliers has not greatly improved the in-sample fit, and omitting impulse indicators would only increase the reported forecast intervals by about 10%. The table reveals that surprisingly, *Ice* provides a better description forecasting over the last 100 periods than in-sample with IIS, even though

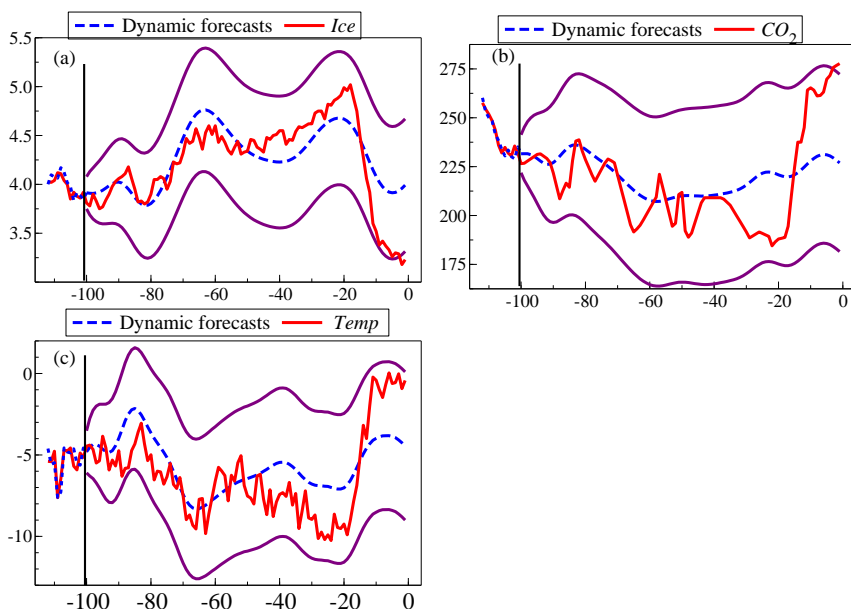


Figure 6.9: A hundred dynamic forecasts with $\pm 2.2\text{SE}$ error bands: (a) for *Ice* from (6.3); (b) for CO_2 from (6.4); (c) for *Temp* from (6.5).

IIS could not be applied to such a ‘future’ period, whereas CO_2 and *Temp* forecasts are worse. Also, the RMSFE for *Ice* is smaller than the in-sample fitted $\tilde{\sigma}$ without IIS, that for CO_2 is similar, whereas again *Temp* forecasts are worse. Looking back at Figure 6.5, the forecast errors for *Ice* seem less variable than the in-sample residuals on ‘ocular’ econometrics, less so those for CO_2 , whereas those for *Temp* look somewhat more volatile.

Figure 6.9 shows the 100 multi-period ahead forecasts with $\pm 2.2\text{SE}$ error bands to reflect the absence of indicators. These error bands assume that the coefficients in the model remain constant, and that no new forces intervene. With the Industrial Revolution, an additional driver of CO_2 was human fossil fuel emissions, and hence of temperature, so extending forecasts to that era by an unchanged model is likely to reveal failure. The first 60 periods are tracked quite well, but miss the changes around 20,000 years ago and all three sets of forecasts either cross or are close to an error band by the end. Compared to earlier changes seen in Figure 6.5, the profiles of *Ice* and CO_2 are similar over the last two cyclical periods, although the last cycle persisted for longer.

6.3.2 When did humanity first influence the climate?

William Ruddiman (2005) suggested that humanity began to influence the climate around the time of domesticating animals and starting farming, so we ‘zoom in’ on the last 10,000 years, and re-estimate the system up to -10 . The estimates are not much changed with $\chi^2_{\text{OR}}(64) = 67.0$, although the contemporaneous coefficient of CO_2 on Temp has increased to unity. Figure 6.10 records the multi-step forecasts over -10 to -1 for Ice , CO_2 and Temp . All the forecasts lie within their $\pm 2\text{SE}$ error bands, but they, and the fitted values for most of the previous 10,000 years, are systematically over for Ice and under for CO_2 and Temp . This matches the dynamic forecasts in Figure 6.9, and could reflect model mis-specification, or the slowly growing divergence that might derive from the increasing influence of humanity envisaged by Ruddiman (2005). Using the presence of proto-weeds that needed ground disturbance to grow in new areas, Ainit Snir, Dani Nadel, Iris Groman-Yaroslavski and Yoel Melamed *et al.* (2015) provide evidence of the origins of cultivation long before Neolithic farming, dating such events to around 23,000 years ago.

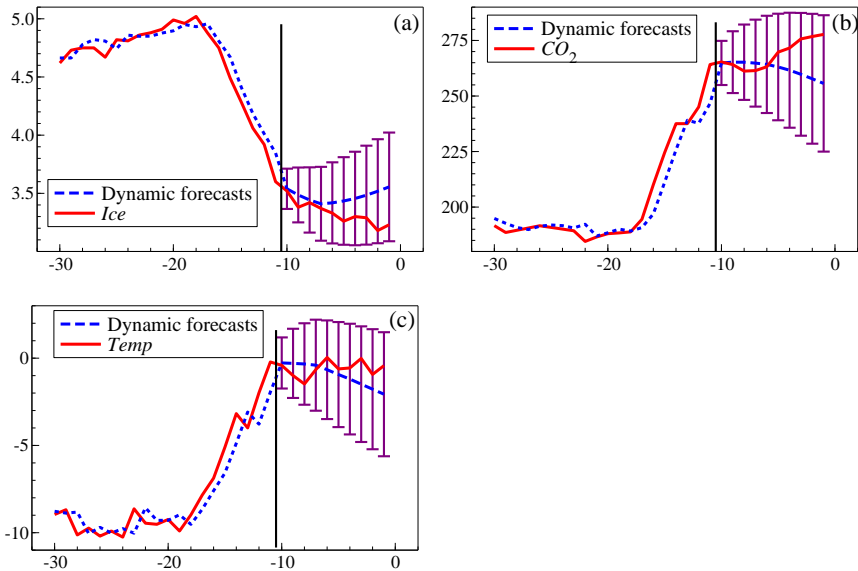


Figure 6.10: Multi-step forecasts over -10 to -1 of: (a) Ice ; (b) CO_2 ; (c) Temp .

Having estimated the system up to 10,000 years ago, we changed the status

of CO₂ to unmodeled and re-estimated the two-equation model for *Ice* and *Temp* conditional on CO₂. Neither fit was much improved, with $\hat{\sigma}_{Ice} = 0.085$ and $\hat{\sigma}_{Temp} = 0.688$, but now contemporaneous CO₂ is highly significant in the equation for *Ice*, with $t = -3.37^{**}$, and its coefficient in the equation for *Temp* has more than doubled to 0.024, which seems implausibly large with $t = 10.4^{**}$. Also, $\chi^2_{OR}(43) = 198^{**}$ strongly rejects. Thus, the evidence here favors CO₂ over the Ice Ages being an endogenous response to the orbital drivers jointly with *Ice* and *Temp*.

6.4 Looking ahead

The eccentricity of Earth's orbital path is calculable far into the future, as are its obliquity and precession, so we extended the data set for 100,000 years into the future: Figure 6.11 records these, where the dark vertical lines denote the present (see <https://biocycle.atmos.colostate.edu/shiny/Milankovitch>). The recent and future eccentricity is relatively quiescent compared to past values.

Extending the data allows us both to forecast over that horizon and to simulate the potential comparative climate should anthropogenically determined atmospheric CO₂ levels settle at (say) 400ppm.⁶ First, §6.3.2 suggested that humanity had been affecting climate since 10,000 years ago, so we commence the multistep forecasts from that date. Consequently, the first 10 forecasts are almost the same as those in Figure 6.10 but are based on a system that does not include any impulse indicators. Figure 6.12 shows these forecasts over -10 to 100 . The figure records the plot of the earlier time series back to 400kya (thousand years ago) to emphasize the comparison with the future period.

Given the relatively quiescent orbitals, these distant forecasts suggest a path well within the range of past data. Matching Pretis and Kaufmann (2020), we also find the next glacial maximum occurs in about 20,000 years. However, we know that the current value of CO₂—i.e., at time 0 on the graphs—is already greater than 400ppm, which is a value dramatically outside the ice-age range, so the location shift in CO₂ values has caused forecast failure in a model that treats CO₂ as remaining endogenously determined by natural factors.

Handling the impacts on *Ice* and *Temp* of a permanent jump from the

⁶We are grateful to Bingchen Wang for his excellent research assistance in data collection and curation. Any sequence of CO₂ values could be investigated: 400ppm is used as close to current levels.

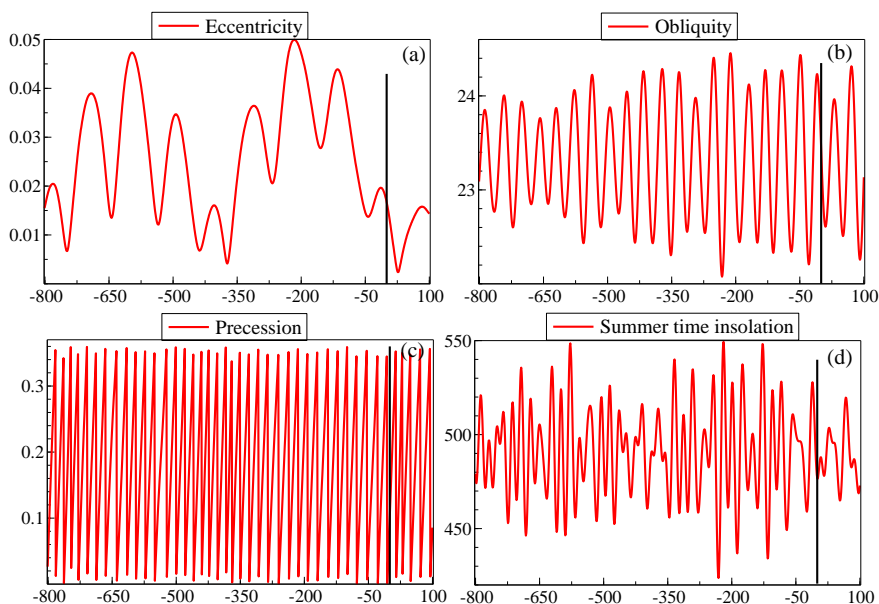


Figure 6.11: (a) Eccentricity; (b) Obliquity; (c) Precession; (d) Summer-time insolation at 65° south, all over -800 to +100.

highest Ice-Age value of CO₂ of around 300ppm to 400ppm or higher in recent times requires care. Arrhenius (1896) showed that the ‘greenhouse’ temperature response was proportional to the logarithm of CO₂. Our model is linear in CO₂, which does not matter over the ice-age period where $\log(\text{CO}_2)$ and CO₂ are correlated 0.99, but does later.

Panel (a) in Figure 6.13 plots $\log(\text{CO}_2)$ and CO₂, matched to have the same means and ranges, showing the very close match across all time periods once the future value representing 400ppm is set at CO₂ = 385ppm against $\log(400)$; Panel (b) shows the equivalent match for CO₂ = 440ppm against $\log(560)$, which is a doubling in CO₂ since the last in-sample observation. Thus, to establish the temperature and ice responses in the system, we set CO₂ to 385ppm and 440ppm to mimic the climatic effects of the anthropogenically exogenous values of 400ppm and 560ppm given the log-linear relation. Of course, the usual orbital drivers still operate, so will continue to influence all three variables in addition to human outputs, but that is switched off for CO₂

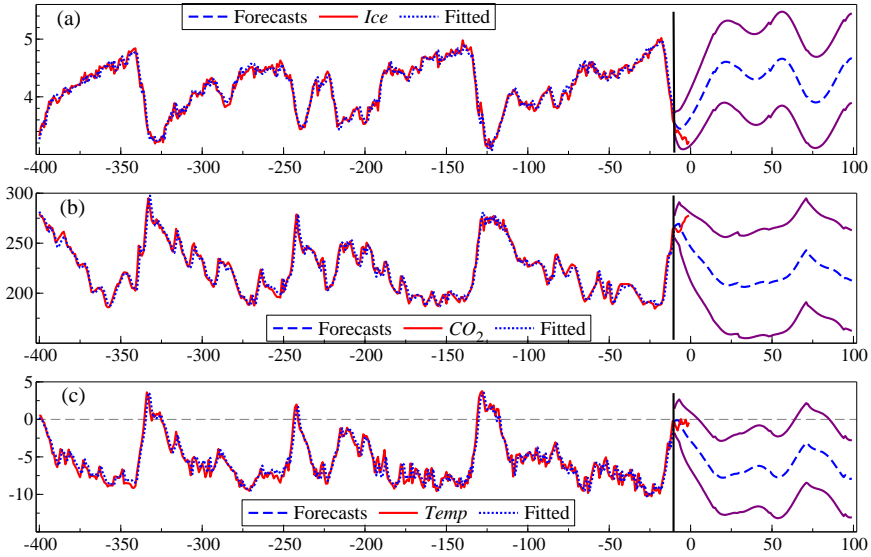


Figure 6.12: 110 dynamic forecasts with $\pm 2SE$ error bands: (a) for *Ice* from (6.3); (b) for CO_2 from (6.4); (c) for *Temp* from (6.5).

so the constant values are the net outcome.

To validly exogenize CO_2 , as the long-run relations between the variables and the orbital drivers should be unaffected by humanity’s intervention in CO_2 production, we fix the values of the parameters as those in equations (6.3) and (6.5), then assign an exogenous status to CO_2 . However, as the forecast period starts 90,000 years later than the estimated model, re-selection by IIS would be needed to include that sample followed by re-estimation. For comparability with the trajectories in Figure 6.12, we omitted all impulse indicators, which should have a negligible effect on these dynamic forecasts. The outcomes are shown in Figure 6.13 Panels (c) for *Ice* and (d) for *Temp*.

Even at 400ppm, ice volume in Panel (c) falls well below any previous values to a minimum around 75,000 years in the future, then increases somewhat from the increased eccentricity. The upper bound of the estimated uncertainty then lies below the lowest ice-age values. Increasing CO_2 to mimic a doubling simply magnifies such effects, leading to an almost ice-free Antarctic. Panel

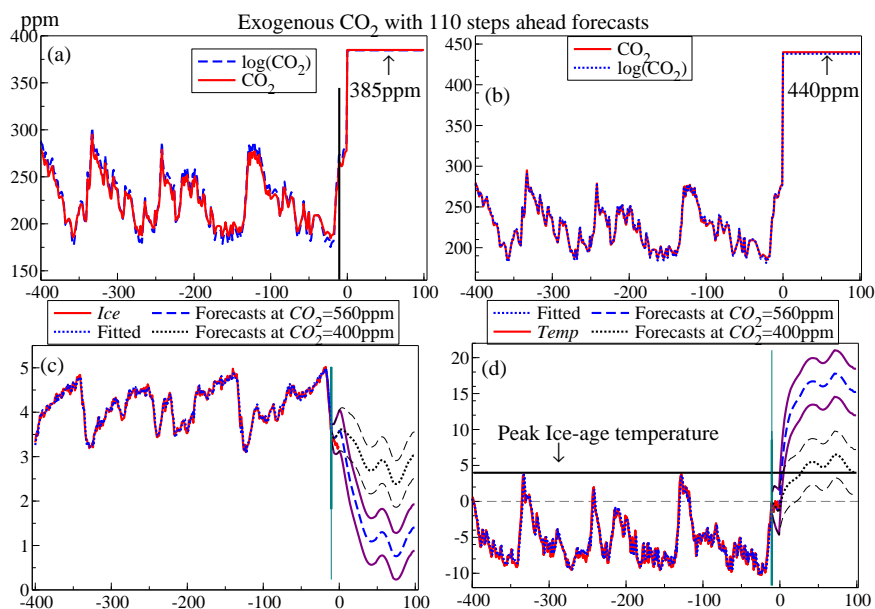


Figure 6.13: (a) & (b) CO_2 and $\log(\text{CO}_2)$ matched by means and ranges for scenarios of 400ppm and 560ppm respectively; 110 dynamic forecasts conditional on $\text{CO}_2 = 385\text{ppm}$ and 440ppm with $\pm 2\text{SE}$ error bands (c) for *Ice*; (d) for *Temp*.

(d) for *Temp* has a line at the peak ice-age temperature, showing that future values will be greater than that for 400ppm, at a maximum of almost 3°C more. Because the observation frequency is 1000 years, and the dynamic reactions occur with those lags given the slowly evolving orbitals, the impacts of the jump in CO_2 take a long time to work through. Nevertheless, for almost 100,000 years, the Antarctic temperature would be above zero (shown by the thin dashed line). Increasing CO_2 to 560ppm raises temperatures dramatically, rising to 13°C above the ice-age peak.

Table 6.6 records the resulting long-run relationships between *Ice* and *Temp* and their determinants: substituting the estimated equation for CO_2 from (6.4) and taking account of the scaling there would essentially deliver Table 6.4. As the model's parameters are fixed, there are no standard errors.⁷

⁷If just the coefficient of CO_2 is left free in this bivariate model with exogenous CO_2 , all others fixed, its long run value is estimated as 0.065 with a standard error of 0.001.

	1	<i>Ec</i>	<i>Ob</i>	<i>EcOb</i>	<i>EcPr</i>	CO ₂
<i>Ice</i>	13.2	365	−0.29	−15.5	−66	−0.0095
<i>Temp</i>	−20.5	−2476	0	106	210	0.066

Table 6.6: Long-run solutions as a function of the relevant strongly exogenous orbital variables and CO₂.

Using the solved long-run coefficient for CO₂ on *Temp* of 0.066, the simulated increase of 105ppm from 280ppm would eventually raise temperatures by about 6.9°C *ceteris paribus*, close to the forecast peak in the Panel (d). For the doubling of CO₂, equivalent to adding 160ppm allowing for the log-transform as seen in Panel (b), the increase would be about 13.2°C, somewhat lower than the peak simulated value of 17.8°C. Kaufmann and Juselius (2013) note that: ‘a permanent 180 ppm increase in atmospheric CO₂ increases the long-run Antarctic temperature by about 11.1C, which corresponds to a global value of about 5.6C [Masson-Delmotte *et al.*, 2006, 2010]’, although Masson-Delmotte *et al.* (2010) also question the 2-to-1 relation with global temperatures. However, our estimates of the effects of adding 120ppm are 7°C Antarctic warming and for adding 280ppm of between 13.2°C and 17.8°C which straddle their estimate. Reto Knutti, Maria Rugenstein and Gabriele Hegerl (2017) record a wide range of estimates of equilibrium climate sensitivity (ECS) with many larger than 6.

There are many *caveats*, from assuming the parameters of the models stay constant as anthropogenic warming increases despite the many implicit dynamic relationships between atmosphere and oceans, and how ice loss would impact those. Nevertheless, the simulated temperature responses to exogenous changes in atmospheric CO₂ are similar to but smaller than those that actually occurred during the ice ages: for example, 10°C between 252kya and 242kya from an 80ppm increase in CO₂, or 13°C between 156kya and 128kya from a 90ppm increase in CO₂, remembering that much of these temperature changes were driven directly by the orbitals.

6.5 Conclusions on Ice-Age modeling

The 3-equation model over the Ice Ages of ice volume, atmospheric CO₂ and Antarctic Temperature illustrates the approach to modeling a system.

While much simpler than the larger cointegrated systems in Kaufmann and Juselius (2013) and Pretis and Kaufmann (2018), the resulting estimated system provided a useful description of the available time series and strongly supported the view that the role of CO₂ was an endogenous response to the orbital drivers during the Ice Ages. The evidence also supported an impact of humanity on the Earth's climate starting at least 10,000 years ago. Extending the orbital data for 100,000 years ahead allowed multi-step forecasts with the system continuing as before so CO₂ was endogenously determined, as well as switching its status to exogenously determined by anthropogenic emissions, although the orbital drivers would still be operating. The resulting inferred global temperature rises would be dangerous at more than 5°C, with Antarctic temperatures positive for thousands of years. Thus, the aims of the Paris Accord remain crucial, so the next Chapter considers the UK's role in reducing its CO₂ emissions.

7

Econometric modeling of UK annual CO₂ emissions, 1860–2017

As described in Chapter 4, CO₂ and other greenhouse gas emissions influence the Earth's climate. Over the Ice Ages, such emissions were determined by natural forces as highlighted by Figure 6.4, but the model developed in Chapter 6 suggested that CO₂ then was primarily an intermediate determinant rather than an exogenous cause of climate variations. However, since the Industrial Revolution discussed in Chapter 5, although the same natural forces still operate, greenhouse gas emissions are now mainly by-products of energy production, manufacturing, and transport (all about a quarter of the UK's emissions), with agriculture, construction and waste making most of the rest.

As first into the Industrial Revolution, the UK initially produced a large share of global anthropogenic CO₂ emissions, albeit much of that was embodied in its exports of cloth production, steam engines, ships and iron products etc. Not only has its share of world CO₂ emissions shrunk to a tiny proportion following global industrialization, there has been a dramatic *drop* in its domestic emissions of CO₂, so that by 2017 they were back to 1890's levels: the country first into the Industrial Revolution is one of the first out. Indeed, on April 22, 2017, 'Britain has gone a full day without turning on its coal-fired power stations for the first time in more than 130 years',¹ and on May 26,

¹See <https://www.ft.com/content/8f65f54a-26a7-11e7-8691-d5f7e0cd0a16>.

2017 generated almost 25% of its electrical energy from solar,² and now goes weeks without burning coal for electric energy production.

The data analyzed here are aggregate, but as the UK population has more than doubled since 1860, in 2013 the UK's CO₂ emissions in per capita terms actually dropped below the level of 1860 (see Figure 7.1c), and are now just 55% of their level in 1894, despite per capita real incomes being around 7-fold higher. Thus, although the UK now 'imports' substantial embodied CO₂—reversing the Industrial Revolution direction—major domestic emissions reductions have occurred but have obviously not involved substantive sacrifice: see Catherine Brinkley (2014) for an empirical analysis of decoupling growth and CO₂ emissions. Much remains to reduce CO₂ emissions towards the net zero level that will be required to stabilize temperatures, an issue we address in Section 7.12.

The aim of this Chapter is to model the UK's CO₂ emissions to establish the determinants of the UK's remarkable drop accomplished with rising real incomes. We again use *Autometrics* to jointly select relevant variables, their lags, possible non-linearities, outliers and location shifts in putative relationships, and also rigorously test selected equations for being well-specified representations of the data.

The structure of this Chapter is as follows. Section 7.1 defines the variables and records their sources, then Section 7.2 describes the UK time-series data under analysis, initially using only data over 1861–2011 for estimation and selection to allow an end-of-sample parameter-constancy test to 2017, updating estimation to 2013 in Section 7.7. Section 7.3 formulates the econometric model, where §7.3.1 considers the choice of functional forms of the regressors. Then Section 7.4 evaluates a simple model formulation, and highlights the inadequacy of such specifications facing wide-sense non-stationary data. The four stages of model selection from an initial general model are described in Section 7.5, then Section 7.6 addresses selecting indicators in the general model. Section 7.7 describes selecting relevant regressors given the retained indicators, and implementing a cointegration reduction, where the non-integrated formulation is estimated in Section 7.8. Section 7.9 conducts an encompassing test of the linear-semilog model versus a linear-linear one. Section 7.10 presents conditional 1-step 'forecasts' and multi-step forecasts

²See <https://www.ft.com/content/c22669de-4203-11e7-9d56-25f963e998b2>.

from a VAR, then Section 7.11 addresses the policy implications of the empirical analysis, then Section 7.12 considers whether the UK can reach its 2008 Climate Change Act (CCA) CO₂ emissions targets for 2050, and the more recent aim of net zero greenhouse gas (GHG) emissions. Finally, Section 7.13 estimates a ‘climate-environmental Kuznets curve’.

7.1 Data definitions and sources

The variables used in the analysis of UK CO₂ emissions are defined as follows:

E_t	=	CO ₂ emissions in millions of tonnes (Mt)	[1], [2]
O_t	=	Net oil usage, millions of tonnes	[3].
C_t	=	Coal volumes in millions of tonnes	[4].
G_t	=	real GDP, £10 billions, 1985 prices	[5], [7], p.836, [8]a,b.
K_t	=	total capital stock, £billions, 1985 prices	[6], [7], p.864, [8]b,c.
Δx_t	=	$(x_t - x_{t-1})$ for any variable x_t	
$\Delta^2 x_t$	=	$\Delta x_t - \Delta x_{t-1}$	

Sources:

[1] World Resources Institute <http://www.wri.org/our-work/project/cait-climate-data-explorer> and <https://www.gov.uk/government/collections/final-uk-greenhouse-gas-emissions-national-statistics>;

[2] Office for National Statistics (ONS)

<https://www.gov.uk/government/statistics/provisional-uk-greenhouse-gas-emissions-national-statistics-2015>;

[3] Crude oil and petroleum products: production, imports and exports 1890 to 2017 Department for Business, Energy and Industrial Strategy (Beis);

[4] Beis and Carbon Brief <http://www.carbonbrief.org/analysis-uk-cuts-carbon-record-coal-drop>;

[5] ONS <https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/timeseries>;

[6] ONS <https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/bulletins/capitalstocksconsumptionoffixedcapital/2014-11-14#capital-stocks-and-consumption-of-fixed-capital-in-detail>;

[7] Mitchell (1988) and Feinstein (1972);

[8] Charles Bean (from (a) *Economic Trends Annual Supplements*, (b) *Annual Abstract of Statistics*, (c) *Department of Employment Gazette*, and (d)

National Income and Expenditure).

See Hendry and Ericsson (1991) and Hendry (2001, 2015) for discussions about G_t and K_t . There are undoubtedly important measurement errors in all these time series, but James Duffy and Hendry (2017) show that strong trends and large location shifts of the form prevalent in the data analyzed here help offset potential biases in the long-run relation's estimated coefficients.

7.2 UK CO₂ emissions and its determinants

As already noted, energy production, manufacturing, and transport each account for roughly 25% of UK CO₂ emissions, the rest coming mainly from agriculture, construction and waste in approximately equal shares. While other greenhouse gas emissions matter, CO₂ comprises about 80% of the UK total, with methane, nitrous oxide and hydrochlorofluorocarbons (HCFCs) making up almost all the rest in CO₂ equivalents. However, the various fossil fuels have different CO₂ emissions per unit of energy produced and how efficiently fuels are burnt also matters, from coal on an open fire or in a furnace, through gasoline-powered vehicles with different engine efficiencies, to a gas-fired home boiler or a power station. A standard approach to estimate country fossil fuel emissions is to use the product of the volumes of fuels produced, the proportion of each fuel that is oxidized, and each fuels' carbon content (see Greg Marland and Ralph Rotty, 1984). Table 7.1 records the average CO₂ emissions per million British thermal units (Btu) of energy produced for the main fossil fuels.³

As rough approximations for interpreting CO₂ reductions, coal has a relative weight of around 2.2, oil 1.6 and natural gas 1.1, depending on the units of measurements. Thus, switching energy production from coal to natural gas would reduce emissions by about 45%–50% for the same amount of energy. Of course, switching to renewable sources would effect a 100% reduction, and is an essential step to reach a net-zero emissions target.

³Variations on such data are used in David Erickson, Richard Mills, Jay Gregg and Terence Blasing *et al.* (2008), Chris Jones and Peter Cox (2005), James Randerson, Matthew Thompson, Thomas Conway and Inez Fung *et al.* (1997), and Cynthia Nevison, Natalie Mahowald, Scott Doney and Ivan Lima *et al.* (2008). Data using this methodology are available at an annual frequency in Marland, Thomas Boden, and Robert Andres (2011). CO₂ emissions from cement production are estimated to make up about 5% of global anthropogenic emissions (see Ernst Worrell, Lynn Price, Nathan Martin and Chris Hendriks *et al.*, 2001).

Coal (anthracite)	228.6
Coal (bituminous)	205.7
Coal (lignite)	215.4
Coal (sub-bituminous)	214.3
Diesel fuel & heating oil	161.3
Gasoline	157.2
Propane	139.0
Natural gas	117.0

Table 7.1: Pounds of CO₂ emitted per million British thermal units (Btu) of energy produced. Source: US Department of Energy & <https://www.eia.gov/tools/faqs/faq.php?id=73&t=11>

The main data over 1860–2017 on UK CO₂ emissions, energy volumes, and the relation of CO₂ emissions to the capital stock are shown in Figure 7.1.

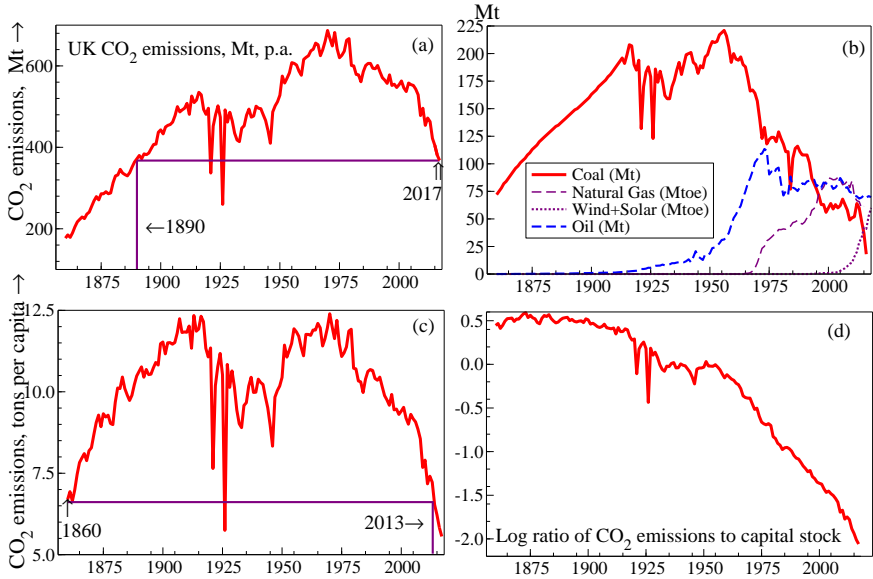


Figure 7.1: (a) UK CO₂ emissions in millions of tonnes (Mt); (b) UK fuel sources: coal (Mt), oil (Mt), natural gas (millions of tonnes of oil equivalent, Mtoe) and wind+solar (Mtoe); (c) CO₂ emissions per capita, in tons per annum; (d) ratio of CO₂ emissions to the capital stock on a log scale, all series to 2017.

Panel (a) shows that UK CO₂ emissions rose strongly and quite steadily

from 1860 till about 1916, oscillated relatively violently till about 1946 from the sharp depression at the end of World War I, the General Strike, Great Depression starting in 1930, and World War II, then resumed strong growth till 1970. Following another somewhat turbulent period till 1984, emissions began to fall slowly, accelerating after 2005 to the end of our time series in 2017, by which time they were below levels first reached in 1890. Panel (c) plots CO₂ emissions per capita, revealing that by 2013 they had fallen below the level at the start of our data period in 1860.

Panel (b) records the time series for coal volumes and net oil usage (imports plus domestic production less exports), natural gas and renewables. Coal volumes behave similarly to CO₂ emissions till 1956 at which point they turn down and continue falling from then onwards, dropping well below the volumes mined in 1860. The sharp dips from miners' strikes in 1921, 1926 and 1984 are clearly visible. Conversely, oil volumes are essentially zero at the start, but rise rapidly in the period of cheap oil after World War II, peak in 1973 with the first Oil Crisis, but stabilize from 1981 on, despite a doubling in vehicle travel to more than 500 billion kilometers p.a. Natural gas usage rises quickly from the late 1960s, but has recently fallen slightly, and renewables have been growing rapidly this century.

Finally Panel (d) plots the log-ratio of CO₂ emissions to the capital stock and shows that it started to decline in the 1880s, and has dropped by more than 92% over the hundred and thirty years since. As capital embodies the vintage of technology prevalent at the time of its construction, tends to be long lasting, and is a key input to production, the volumes of CO₂ produced by production are likely to be strongly affected by the capital stock: see e.g., Alexander Pfeiffer, Richard Millar, Cameron Hepburn and Eric Beinhocker (2016). Hence, 'stranded assets' could be a potential problem if legislation imposed much lower CO₂ emissions targets, as looks likely for the UK.

To highlight the massive changes that have occurred in the UK, Figure 7.2 reports a scatter plot of CO₂ emissions against the quantity of coal, showing the dates of each pair of points, and a 3-dimensional plot of E_t against K_t and C_t . As with Figure 7.1(a), there is strong growth in emissions as coal output expands until the mid 1950s when coal production peaks, but emissions continue to grow till the mid 1970s despite a substantial reduction in coal volumes, and only then start to decline, falling noticeably after 2008. Referring back to Figure 7.1(b), the rapid rise in oil use initially offsets the fall in coal, but

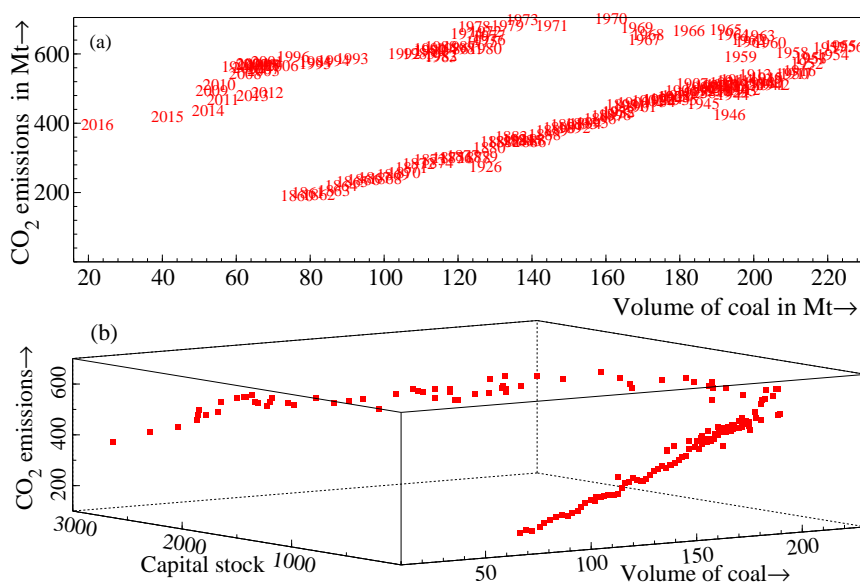


Figure 7.2: (a) Scatter plot of CO₂ emissions against the quantity of coal by date; (b) 3-dimensional plot of E_t against K_t and C_t .

after the two Oil Crises of the 1970s, the fall in coal is reflected in the decline in emissions. Panel (b) shows the major role of the capital stock in changing the link between coal and CO₂ emissions, reflecting the efficiency gains seen in Figure 7.1(d). Figure 7.3 shows the distributional shifts in CO₂ emissions that have occurred historically, using approximately 40-year sub-periods.

All the above graphs show non-linear relationships at the bivariate level (i.e., between CO₂ emissions and coal production, say), as well as shifts in relations. An immediate implication is that simple correlations between pairs of variables change over time, so will be poor guides to what matters in a multivariable relationship, as Table 7.2 shows. Coal volumes have the smallest correlation with CO₂ emissions, yet were manifestly one of its main determinants.⁴

Figure 7.4 shows recursive estimates of the relation $E_t = \hat{\beta}_0 + \hat{\beta}_1 C_t + \hat{\nu}_t$, confirming the dramatic non-constancy of that overly simple model, illustrating the problems of not modeling non-stationarity.

⁴Correlations are not well defined for non-stationary variables, as they are not constant over time.

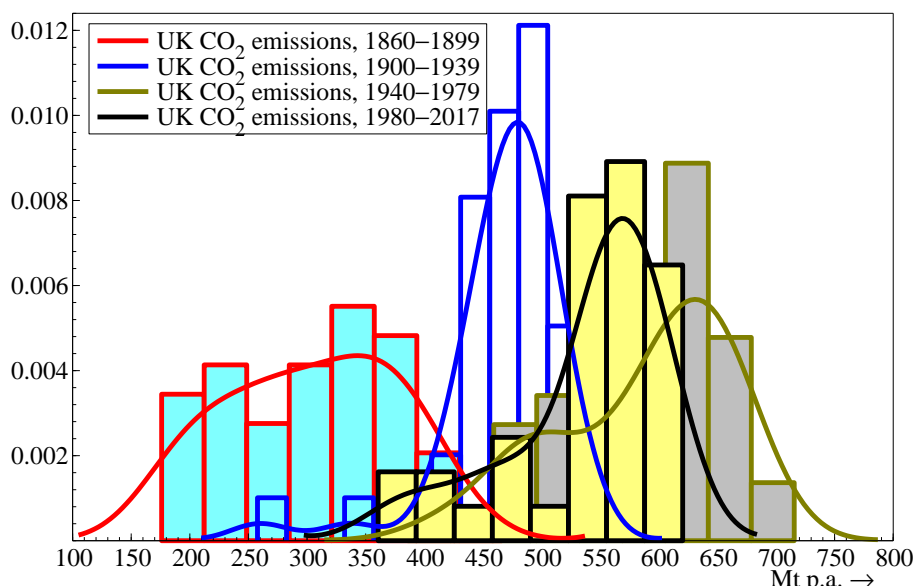


Figure 7.3: Sub-period distributions of UK CO₂ emissions.

Correlations :	CO ₂ emissions	Coal	Oil	real GDP	capital
CO ₂ emissions	1.000	0.243	0.734	0.528	0.506
Coal		1.000	-0.424	-0.598	-0.624
Oil			1.000	0.829	0.822
real GDP				1.000	0.997

Table 7.2: Whole-sample correlations.

7.3 Model formulation

Following the formulation in (2.1) above, the general model is the system characterizing the LDGP. Here, we are interested in modeling UK CO₂ emissions given the volumes of coal and oil the UK used and the main representations of the scale of the economy and its productive capacity, namely GDP and the capital stock. Over most of our sample period, there would not be any contemporaneous or lagged feedbacks from CO₂ emissions to the explanatory regressors, although by the middle of the 20th century with ‘Clean Air’ Acts of Parliament, that is a possibility, increasingly so by the first decade of the

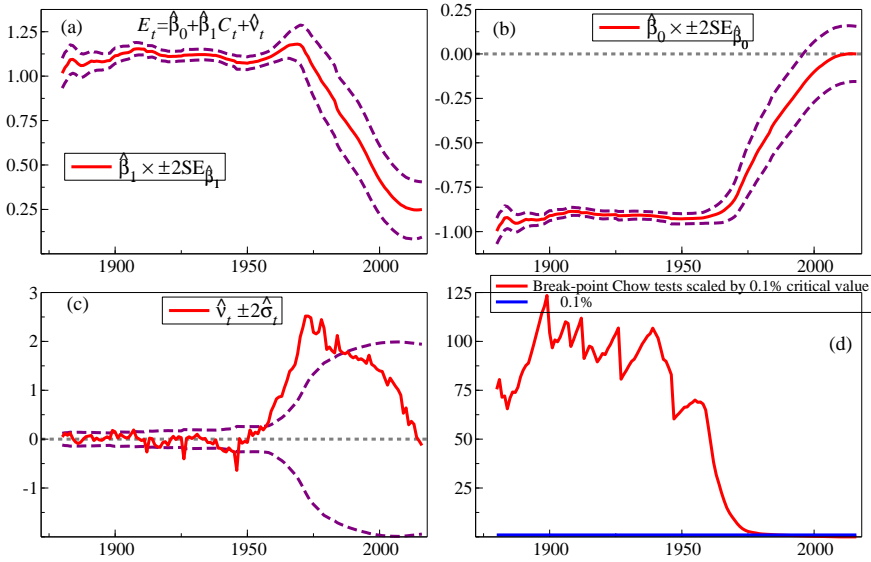


Figure 7.4: (a) Recursive $\hat{\beta}_{1,t}$ with $\pm 2SE_{\hat{\beta}_{1,t}}$; (b) Recursive $\hat{\beta}_{0,t}$ with $\pm 2SE_{\hat{\beta}_{0,t}}$; (c) 1-step recursive residuals $\hat{\nu}_t$ with $\pm 2\hat{\sigma}_t$; (d) Break-point Chow tests scaled by their 0.1% critical values.

21st century as climate change concerns grow, but overall a conditional model seems a viable representation here.

Combining all the above information, neither of the two ‘polar’ approaches to modeling the UK’s CO₂ emissions, namely as (a) decomposed into its sources (coal, oil, gas etc.), or (b) as a function of economic variables (capital and output) alone, seems likely to be best. On (a), not all sources have been recorded historically, especially their carbon compositions, which will have varied over time with the type of coal used, and how oil was refined to achieve which products (*inter alia*). On (b), that changing mix will entail non-constancy in the relation between emissions and the capital stock and GDP. To capture the changing mix and its relation to the economic variables, we included the two main emitters, coal and oil, with the capital stock and GDP. The latter then explain the emissions not accounted for by the former: the solved long-run relationship in equation (7.5) below finds all four variables play a significant role, and the coefficients for coal and oil are also consistent with that interpretation. In turn, the additive nature of emissions suggests a

linear relation with coal and oil, although that leaves open how the economic variables might enter, considered in §7.3.1.

A further obvious feature of Figure 7.1(a) is the number of very large ‘outliers’ occurring during the inter-war and immediate post-war periods. Consequently, the general set of variables from which the model for CO₂ emissions will be selected comprises its lagged value and current and first lagged values of coal and oil volumes, real GDP and the capital stock. These variables are all retained without selection while selecting over both impulse and step indicators at $\alpha = 0.1\%$ significance. First, however, we address the functional forms for G_t and K_t .

7.3.1 Functional forms of the regressors

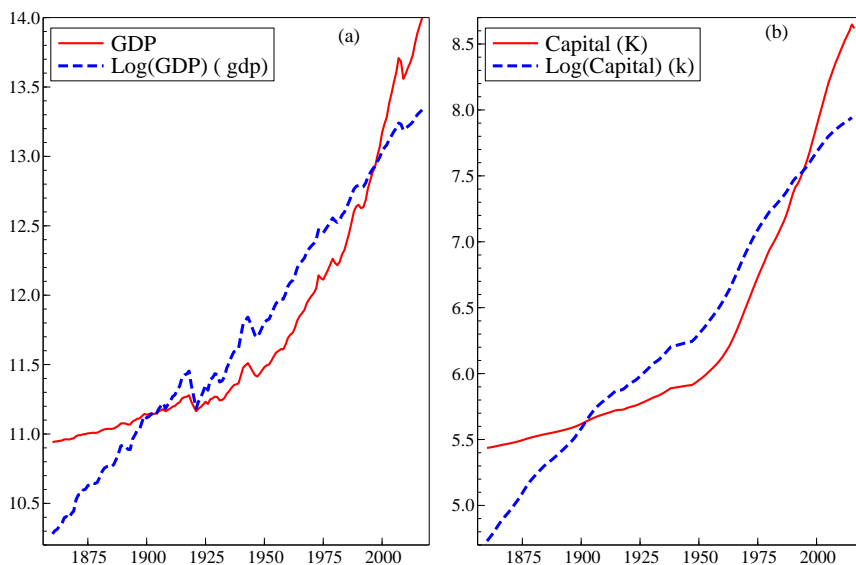


Figure 7.5: Graphs on a logarithmic scale matched by means and ranges of linear (capitals) and log (lower case) transforms of (a) GDP; (b) Capital stock.

In §2.2.2 we considered a low-dimensional representation of non-linearity, but here a more specific issue is whether to transform the various regressors to logarithms or leave as linear. CO₂ emissions depend linearly on the volumes of fossil fuels consumed with the weights shown in Table 7.1. Moreover, it

is the volume of CO₂ emitted that has to be reduced to net zero, so we use that as the dependent variable. In turn, it is natural to include coal and oil volumes linearly as well. Nevertheless, both linear and log linear relations were investigated. As oil was used in negligible quantities in the 19th Century, early volumes were increased by unity (to ensure positive values), but the log transform still seemed to distort rather than help.

Equivalent linear and log-linear equations were formulated as:

$$E_t = \beta_0 + \beta_1 E_{t-1} + \beta_2 C_t + \beta_3 C_{t-1} + \beta_4 O_t + \beta_5 O_{t-1} + \beta_6 G_t + \beta_7 G_{t-1} + \beta_8 K_t + \beta_9 K_{t-1} + u_t \quad (7.1)$$

and the same form with all variables in logs, then estimated with IIS+SIS selecting at 0.001 retaining all the regressors in (7.1). The log-linear version had a residual standard deviation of 2.6%, whereas dividing the residual standard deviation of the linear form (reported in <https://voxeu.org/article/driving-uks-capita-carbon-dioxide-emissions-below-1860-levels>) by the mean value of E_t yielded 2.0%, so the linear representation dominated on the criterion proposed by Sargan (1964). By way of comparison, even after IIS, the ‘Kuznets curve’ formulation in (7.7) below had a residual standard deviation of 5.5%.

However, that leaves open the choice of log or linear just for G_t and K_t . Figure 7.5 graphs those variables in linear and log transforms, matched by means and ranges to highlight any relative curvature. Given the large increase in both since 1860, £100billion corresponds to very different percentage changes, illustrated by the apparently small fall in G after World War I, yet the largest drop in g , with the opposite after 2008. Consequently, we will model with the logs, denoted g and k , scaled by 100 so coefficients are between ± 10 , and Δg and Δk are percentage changes. The encompassing test in Section 7.9 checks how well the two possibilities of linear and semi-log compare. Outliers and location shifts detected by super saturation estimation may well differ between these specifications.

7.4 Evaluating a model without saturation estimation

Thus, the baseline relationship between emissions and its main determinants was formulated as:

$$E_t = \beta_0 + \beta_1 E_{t-1} + \beta_2 C_t + \beta_3 C_{t-1} + \beta_4 O_t + \beta_5 O_{t-1} + \beta_6 g_t + \beta_7 g_{t-1} + \beta_8 k_t + \beta_9 k_{t-1} + v_t. \quad (7.2)$$

To demonstrate why a simple-to-general methodology is inadequate, we will first estimate and evaluate the relation in (7.2) over 1861–2011 with six observations retained as an end-of-sample constancy test for 2012–2017, given in (7.3) where estimated coefficient standard errors (SEs) are shown in parentheses below estimated coefficients with heteroskedastic and autocorrelation consistent standard errors (HACSEs) shown below those in brackets (see Whitney Newey and Kenneth West, 1987, and Donald Andrews, 1991).

$$\begin{aligned} \hat{E}_t = & 0.79 E_{t-1} + 2.58 C_t - 2.21 C_{t-1} + 2.05 O_t - 1.53 O_{t-1} \\ & (0.054) \quad (0.14) \quad (0.18) \quad (0.43) \quad (0.43) \\ & [0.070] \quad [0.38] \quad [0.40] \quad [0.53] \quad [0.53] \\ & + 0.81 g_t - 0.99 g_{t-1} + 1.67 k_t - 1.39 k_{t-1} + 61 \quad (7.3) \\ & (0.53) \quad (0.53) \quad (2.67) \quad (2.62) \quad (133) \\ & [0.49] \quad [0.57] \quad [2.65] \quad [2.57] \quad [109] \end{aligned}$$

$$\hat{\sigma} = 16.2 \quad R^2 = 0.985 \quad F_{AR}(2, 139) = 8.44^{**} \quad \chi_{nd}^2(2) = 64.4^{**}$$

$$F_{ARCH}(1, 149) = 18.9^{**} \quad F_{Het}(18, 132) = 2.95^{**}$$

$$F_{Reset}(2, 139) = 14.3^{**} \quad F_{Chow}(6, 141) = 0.96 \quad t_{ur} = -3.91$$

Despite the high R^2 induced by the non-stationarities in the variables, the model is completely inadequate. Every mis-specification test rejects, the key economic variables g and k are insignificant, and t_{ur} does not reject the null hypothesis of no cointegration. The solved long-run equation for E in Table 7.3 also has the ‘wrong’ relative coefficients of coal and oil.

Variable	1	C	O	g	k
Coefficient	289	1.77	2.45	-0.86	1.35
SE	635	0.17	0.64	1.05	0.97

Table 7.3: Solved static long-run equation for E .

The HACSEs do not alter the significance or insignificance of the regressors, and given the substantive rejections on F_{AR} and F_{Het} , are surprisingly close to the conventional SEs (see the critiques of HACSEs in Castle and Hendry, 2014a, and Aris Spanos and Reade, 2015), so do not alert investigators who fail to compute mis-specification tests as to the problems.

Finally, the recursively-estimated coefficients $\hat{\beta}_{i,t}$ with $\pm 2SE_{i,t}$, the residuals with $\pm 2\hat{\sigma}_t$, and the recursive F_{Chow} test are shown in Figure 7.6 revealing

considerable non-constancy. The coefficient of E_{t-1} is converging towards unity, often signalling untreated location shifts (see Castle *et al.*, 2010).

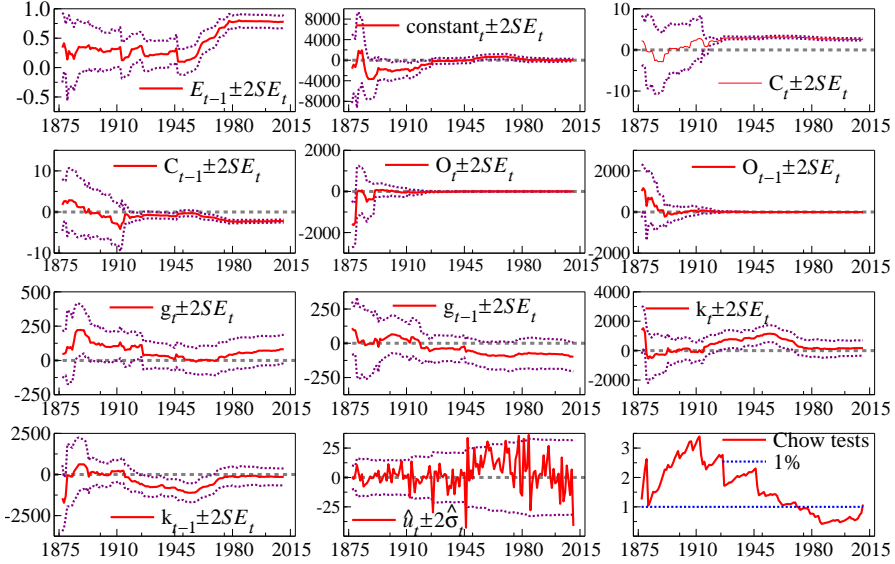


Figure 7.6: Graphs of $\hat{\beta}_{i,t}$, $i = 0, \dots, 9$ with $\pm 2SE_{i,t}$; \hat{u}_t with $\pm 2\hat{\sigma}_t$, and F_{Chow_t} over 1875–2011.

The dilemma confronting any investigator after fitting (7.3), and facing so many test rejections, is how to proceed. Mis-specification tests can reject against a number of different alternatives to those for which they were originally derived, so implementing that particular alternative is a non-sequitur. For example, residual autocorrelation need not entail error autocorrelation but may arise from incorrect dynamics, unmodeled location shifts or other parameter changes, data measurement errors and omitted variables, so adopting a recipe of the form often attributed to Guy Orcutt and Donald Cochrane (1949) can be counter-productive (see e.g. Mizon, 1995). Indeed, once there is residual heteroskedasticity and non-constancy, it is unclear what other rejections mean, except to confirm that something is wrong. The obvious alternative of general-to-specific is what we now explore for modeling UK CO₂ emissions.

7.5 Four stages of single-equation model selection

In this Section, we consider the four stages of conditional model selection from (7.2) extended by using super saturation (namely IIS+SIS), fitting to data over 1861–2011 to allow an end-of-sample parameter-constancy test to 2017.

First, in Section 7.6 we select both impulse and step indicators at a tight nominal significance level α , which is the theoretical gauge, retaining all of the other regressors in (7.2) without selection. The studies referenced in Chapter 2 have established that the theoretical and empirical gauges are generally close for IIS, and have derived the uncertainty around the latter, which is almost negligible for very small $\alpha_{0.001} = 0.001$. Less is known analytically about the gauge of SIS or super saturation, but the simulation studies noted earlier suggest the gauge should be set around $1/2T$. Since there are $T = 151$ observations, there will be $M \approx 300$ indicators in the candidate set (T impulse indicators and $T - 2$ step indicators), so under the null hypothesis that no indicators are needed, $\alpha_{0.001}M = 0.001 \times 300 = 0.3$ of an indicator will be significant by chance. Even doubling that, $\alpha_{0.001}2M$ can be interpreted that one indicator will be retained adventitiously approximately 3 out of every 5 times these choices are applied to new data sets with the same configuration of T , so over-fitting seems unlikely. As shown above, estimating (7.2) without indicator variables is unsuccessful as all mis-specification tests strongly reject. Diagnostic tests will be applied to check that the finally selected equation is well specified, with non-autocorrelated, homoskedastic and nearly Normal residuals, constant parameters, and no remaining non-linearity: (7.4) records that outcome.

Second, in Section 7.7 we select over the other nine regressors at $\alpha_{0.01}$ (indicators already selected are bound to be significant at this second stage). Almost none of the 9 regressors will be retained by chance if in fact they are irrelevant.

Third, also in Section 7.7 we solve this selected model for the cointegrating, or long-run, relation implicit in it, and reparametrize the non-deterministic variables to differences. In doing this mapping to a non-integrated specification, step indicators are included in the cointegration relation, so that they do not cumulate to trends, leaving impulse indicators and differenced step indicators unrestricted. While this may seem somewhat complicated, the reasons for doing so are explained in the survey articles by Hendry and Juselius (2000,

2001) and in Hendry and Pretis (2016). Finally, we re-estimate that non-integrated formulation in Section 7.8.

7.6 Selecting indicators in the general model

Following this path, we find for $T = 1862\text{--}2011$, retaining all the regressors and selecting impulse and step indicators jointly at 0.1%, testing constancy over 2012–2017:

$$\begin{aligned}
 \hat{E}_t = & \underset{(0.06)}{0.52} E_{t-1} - \underset{(13)}{47} 1_{\{1921\}} - \underset{(20)}{163} 1_{\{1926\}} - \underset{(10)}{44} 1_{\{1946\}} \\
 & + \underset{(11)}{56} 1_{\{1947\}} + \underset{(9.8)}{29} 1_{\{1996\}} - \underset{(14)}{42} S_{\{1925\}} + \underset{(13)}{72} S_{\{1927\}} \\
 & - \underset{(7.5)}{31} S_{\{1969\}} + \underset{(10)}{47} S_{\{2010\}} - \underset{(89)}{158} + \underset{(0.13)}{1.86} C_t \\
 & - \underset{(0.18)}{0.88} C_{t-1} + \underset{(0.26)}{1.71} O_t - \underset{(0.28)}{1.07} O_{t-1} + \underset{(0.33)}{0.95} g_t \\
 & - \underset{(0.33)}{1.13} g_{t-1} + \underset{(1.8)}{7.64} k_t - \underset{(1.8)}{7.02} k_{t-1} \tag{7.4}
 \end{aligned}$$

$$\hat{\sigma} = 9.58 \quad R^2 = 0.995 \quad F_{AR}(2, 130) = 2.93 \quad \chi_{nd}^2(2) = 5.97$$

$$F_{ARCH}(1, 149) = 3.42 \quad F_{Het}(20, 123) = 0.82$$

$$F_{Reset}(2, 130) = 2.30 \quad F_{Chow}(6, 132) = 1.40 \quad F_{nl}(27, 105) = 1.04$$

where F_{nl} tests for non-linearity (see §2.2.2). All of these mis-specification tests are insignificant, including F_{Reset} and F_{nl} so all of the non-linearity has been captured by (7.4), but the tests are applied to $I(1)$ data, so correct critical values are not known: see Vanessa Berenguer-Rico and Jesus Gonzalo (2014) for a test of non-linear cointegration applied in this context.

Five impulse and four step indicators have been selected despite the very tight significance level. Combining the indicators in (7.4) allows some simplification by transforming $1_{\{1926\}}$ and $S_{\{1927\}}$ to $\Delta 1_{\{1926\}}$, and $1_{\{1947\}} - 1_{\{1946\}} = \Delta 1_{\{1947\}}$. This reduces the number of genuine location shifts to three, an intermediate modeling stage that was implemented before selecting over the 9 regressors. The resulting $\hat{\sigma}$ was unaffected by these transformations.

The remaining step shifts capture major events with long-term impacts that

are not otherwise captured by the variables in the model. These could reflect changes in the improving efficiency of fuel use, or the effects of omitting other sources of emissions with key technological changes, or usage shifts not taken into account in calculating emissions. Since steps in the *Autometrics* implementation of SIS terminate at the dates shown, their reported signs reflect what happened **earlier**, so a positive coefficient for $S_{\{1925\}}$ entails a higher level prior to 1926. That is the date of the 1926 Act of Parliament that created the UK's first nationwide standardized electricity distribution grid, greatly enhancing the efficiency of electricity, but also witnessed the General Strike probably captured by $\Delta 1_{\{1926\}}$. Then 1969 saw the start of the major conversion of UK gas equipment from coal gas (about 50% hydrogen) to natural gas (mainly methane) with a considerable expansion in its use. The coefficients of both these location shifts have the appropriate signs of reducing and increasing emissions respectively. Although the UK's Clean Air Act of 1956 did not need a step indicator, probably because it was captured by the resulting fall in coal use, we interpret the step shift $S_{\{2010\}}$ showing a higher level of emission of 47Mt before then as the reaction to the Climate Change Act of 2008 (see <https://www.legislation.gov.uk/ukpga/2008/27/contents>) and the European Union's Renewables Directive of 2009, discussed in Section 7.11. Thus, we doubt the explanation is the Great Recession of 2008–2012, since the previous largest GDP fall in 1921–22 did not need a step, but just had an impulse indicator for the large outlier in 1921. As coal volumes are included, indicators for miners' strikes should only be needed to capture changes in inventories, which might explain part of the large impulse indicator for 1926.

7.7 Selecting regressors and implementing cointegration

Secondly, all 9 regressors are retained when selecting at 1% significance.

Third, we solve for the long-run cointegrating relationship, justified by the *PcGive* (2018) unit-root t-test value of $t_{ur} = -8.99^{**}$ which strongly rejects the null hypothesis of no cointegration (see Ericsson and MacKinnon, 2002, for the appropriate critical values, which are programmed into *PcGive*). The resulting cointegration relation defines the equilibrium-correction trajectory $\tilde{Q}_t = E_t - \tilde{E}_{LR,t}$ (adjusted to a mean of zero in-sample). Step indicators need to be **led** by one period as \tilde{Q}_{t-1} will be entered in the transformed model.

However, being at the end of the initial sample up to 2011 from the

definition of step indicators here, $1 - S_{\{2010\}}$ only has 2 observations in sample. Consequently, it was decided to extend the estimation sample by two observations to 2013 since the current full sample now ended in 2017, to enable the cointegrating relation to include $S_{\{2010\}}$. This led to closely similar estimates to (7.4) with $t_{ur} = -9.34^{**}$ and the solved long-run:

$$\begin{aligned} \tilde{E}_{LR} = & \quad 2.0 \ C + \quad 1.4 \ O + \quad 1.18 \ k - \quad 0.27 \ g + \quad 63 \ S_{\{1924\}} \\ & (0.06) \quad (0.18) \quad (0.27) \quad (0.28) \quad (6) \\ & - 64.0 \ S_{\{1968\}} + \quad 70 \ S_{\{2009\}} - \quad 328 \\ & (14) \quad (13) \quad (165) \end{aligned} \quad (7.5)$$

All variables are significant at 1% other than g , which is ‘wrong signed’. The coefficient of coal is close to the current standard estimate of ≈ 2.1 – 2.3 , as is that of oil to its estimate, though somewhat lower than the 1.6 in Table 7.1.

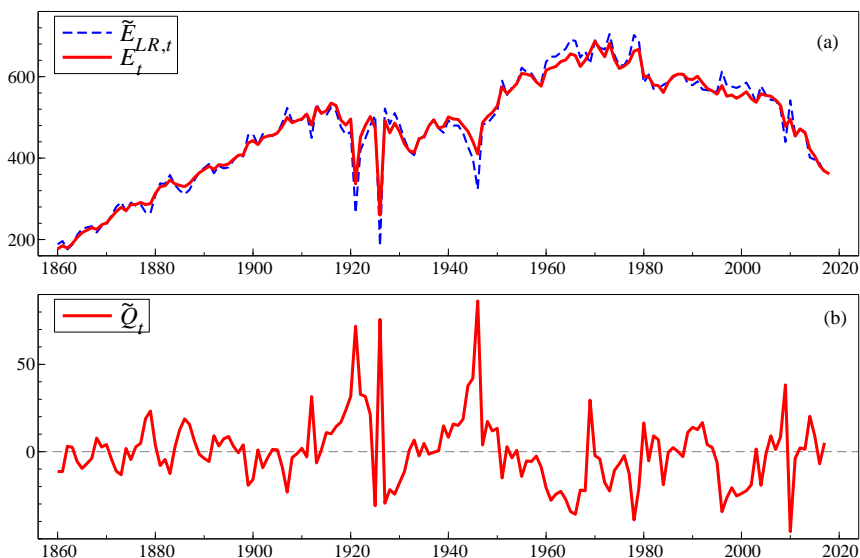


Figure 7.7: (a) E_t and $\tilde{E}_{LR,t}$; (b) $\tilde{Q}_t = E_t - \tilde{E}_{LR,t}$ centered on a mean of zero.

Figure 7.7(a) shows how closely the long-run derived relation $\tilde{E}_{LR,t}$ in (7.5) tracks E_t . Panel (b) records the resulting time series for \tilde{Q}_t centered on a mean of zero. While \tilde{Q}_t is not stationary from a changing variance—unsurprising given the omission of the impulse indicators, matching the visible spikes—a unit root is rejected.

Because the units in which the different variables in (7.5) are measured are not directly comparable, their relative importance as determinants of the level of E_t is hard to judge. However, Figure 7.2(b) provided a 3-dimensional plot of E_t against K_t and C_t to show that while the rise then fall of coal usage in Figure 7.2(a) explains much of the behavior of CO₂ emissions, the increases in the capital stock track the shift in the mid 20th Century to higher emissions for the same volumes of coal as in the 19th (a similar picture emerges when plotting E_t against C_t and O_t , but with a more erratic spread). Moreover, in the relatively similar long-run solution in a log-linear formulation, where coefficients are elasticities, the two dominant influences were 0.42 from coal and 0.40 from capital stock, with much smaller effects from GDP and oil. These effects match prior anticipations as discussed above. Indeed, in the linear and log-linear models, the long-run effect of GDP is *negative*, possibly reflecting the move from manufacturing to a service-based economy, although it is insignificant in the semi-log form (7.5).

7.8 Estimating the cointegrated formulation

Fourth, transforming to a model in differences and the lagged cointegration relation from (7.5) then re-estimating revealed a couple of additional outliers (significant at 1% but not the original 0.1%), and adding those indicators yielded (7.6) for 1861–2013, testing constancy over 2014–2017.

$$\begin{aligned}
 \widehat{\Delta E}_t = & \underset{(0.10)}{1.88} \Delta C_t + \underset{(0.21)}{1.71} \Delta O_t + \underset{(1.09)}{7.15} \Delta k_t + \underset{(0.28)}{0.89} \Delta g_t \\
 & - \underset{(0.05)}{0.50} \tilde{Q}_{t-1} - \underset{(2.4)}{15.2} - \underset{(8.8)}{79.4} \Delta 1_{\{1926\}} + \underset{(6.4)}{50.2} \Delta 1_{\{1947\}} \\
 & - \underset{(11.1)}{45.8} 1_{\{1921\}} - \underset{(8.9)}{27.5} 1_{\{1912\}} + \underset{(8.9)}{26.8} 1_{\{1978\}} + \underset{(8.9)}{28.4} 1_{\{1996\}}
 \end{aligned} \tag{7.6}$$

$$\hat{\sigma} = 8.87 \quad R^2 = 0.94 \quad F_{AR}(2, 139) = 0.49 \quad \chi^2_{nd}(2) = 1.67$$

$$F_{Het}(14, 134) = 1.03 \quad F_{ARCH}(1, 151) = 0.53$$

$$F_{Reset}(2, 139) = 1.50 \quad F_{nl}(15, 126) = 1.35 \quad F_{Chow}(4, 141) = 1.75$$

Increases in oil, coal, k and g all lead to increases in emissions, which then equilibrate back to the long-run relation in (7.5). There are very large perturbations from this relationship, involving step shifts, impulses and blips. Archival research revealed that 1912 saw the first **national** strike by coal miners in Britain causing considerable disruption to train and shipping schedules, although nothing obvious was noted for 1978.

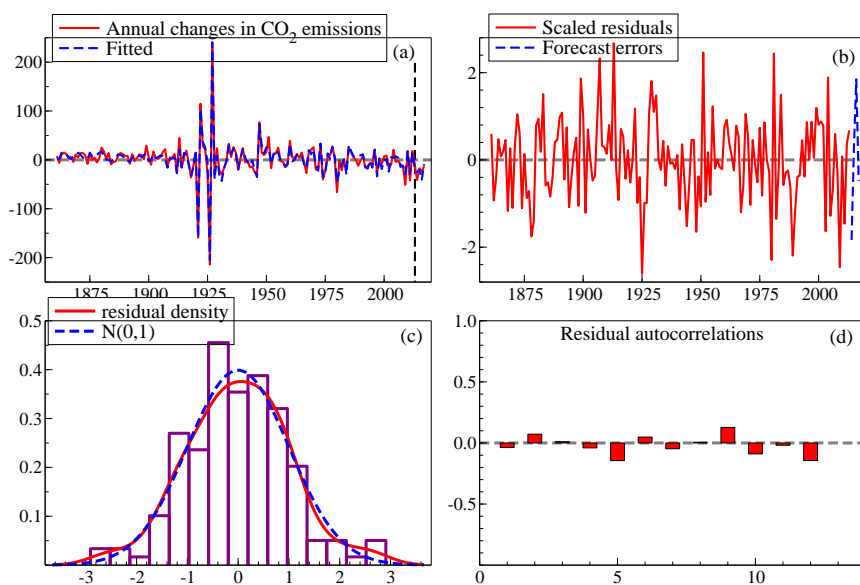


Figure 7.8: (a) Actual and fitted values for ΔE_t from (7.6); (b) residuals scaled by $\hat{\sigma}$; (c) residual density and histogram with a Normal density for comparison; (d) residual autocorrelation.

The turbulent periods create such large changes it is difficult to ascertain how well the model describes the data from Figure 7.8, so Figure 7.9 records the implied levels' fitted values and outcomes. The match is extremely close, although the sudden lurches are only 'modeled' by indicator variables, as are several of the step shifts. Possible explanations for the need for impulse indicators, some discussed above, include the role of gas, changes in stocks of coal and oil leading to divergences from measured output (so having different effects on emissions), the changing efficiency of production and usage (e.g., replacing electric fires by central heating), and general changes such as better insulation. All of the diagnostic statistics remain insignificant.

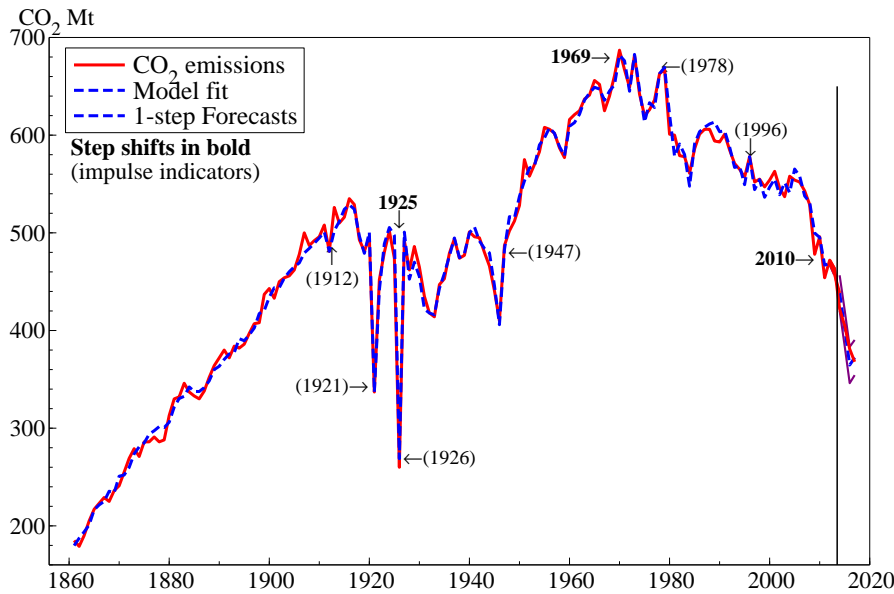


Figure 7.9: Actual and fitted values for UK CO₂ emissions with indicator dates.

7.9 Encompassing of linear-semilog versus linear-linear

Encompassing tests were applied in §3.2 to discriminate between the roles of trend and previous distances driven. In this section, we apply them to test the linear-semilog model in (7.6) denoted M_1 against the earlier linear-linear model reported in <https://voxeu.org/article/driving-uks-capita-carbon-dioxide-emissions-below-1860-levels> denoted M_2 . As in the application of encompassing in Table 3.1, the instruments are the combined regressors of the two models. Table 7.4 records the outcome.

Test	Form	M_1 vs. M_2	Form	M_2 vs. M_1
Cox (1962)	$N[0,1]$	-3.74^{**}	$N[0,1]$	-5.40^{**}
Ericsson (1983) IV	$N[0,1]$	3.26^{**}	$N[0,1]$	4.53^{**}
Sargan (1964)	$\chi^2(4)$	10.0^*	$\chi^2(4)$	17.9^{**}
Joint model	$F(4,134)$	2.62^*	$F(4,134)$	5.00^{**}

Table 7.4: Encompassing test statistics where M_1 is (7.6) with $\sigma_{M_1} = 8.89$, M_2 is the linear model with $\sigma_{M_2} = 9.18$ and $\sigma_{\text{Joint}} = 8.69$.

The instruments used were $S_{\{2010\}}$, Δg_t , Δk_t , Constant, $\Delta 1_{\{1947\}}$, \tilde{Q}_{t-1} , ΔC_t , ΔO_t , $1_{\{1912\}}$, $1_{\{1921\}}$, $1_{\{1978\}}$, $1_{\{1996\}}$, ΔK_t , $1_{\{1970\}}$, $\Delta S_{\{1983\}}$, $\Delta 1_{\{1926\}}$, and $\tilde{Q}_{GK,t-1}$, where \tilde{Q}_{t-1} and $\tilde{Q}_{GK,t-1}$ denote the equilibrium correction terms of the versions with log and linear GDP and Capital respectively. Although M_1 is rejected against M_2 , the $F(4,134)$ parsimonious encompassing test against the joint model is equivalent to adding the four variables from the linear model, and is not significant at the 1% level used for selection, nor are any of those variables individually significant at 1%. Conversely, \tilde{Q}_{t-1} , Δg_t , Δk_t , and $1_{\{1921\}}$ are highly significant at less than 0.1% if added to M_2 .

7.10 Conditional 1-step ‘forecasts’ and system forecasts

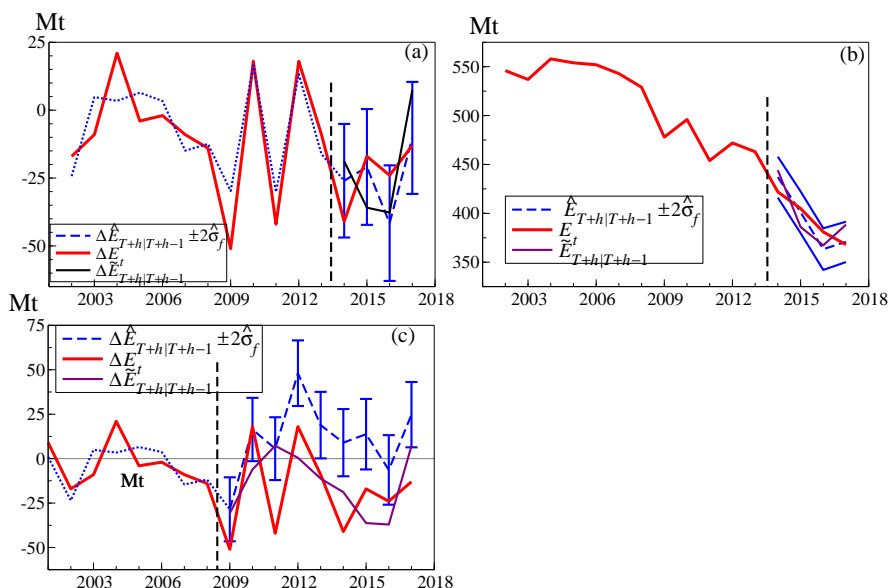


Figure 7.10: (a) Outcomes ΔE_t , fitted values, and 1-step conditional ‘forecasts’ $\widehat{\Delta E}_{T+h|T+h-1}$ with $\pm 2\hat{\sigma}_f$ shown as bars, and robust ‘forecasts’ $\tilde{\Delta E}_{T+h|T+h-1}$; (b) implied $\tilde{E}_{T+h|T+h-1}$ from (a) with $\pm 2\hat{\sigma}_f$, and corresponding robust ‘forecasts’ $\tilde{E}_{T+h|T+h-1}$, both from 2013; (c) ΔE_t , fitted values, and 1-step conditional ‘forecasts’ $\widehat{\Delta E}_{T+h|T+h-1}$ with $\pm 2\hat{\sigma}_f$ shown as bars, and robust ‘forecasts’ $\tilde{\Delta E}_{T+h|T+h-1}$ commencing in 2008.

To check the constancy of the model after 2013, Figure 7.10 (a) records the

four 1-step ahead ‘forecasts’ $\widehat{\Delta E}_{T+h|T+h-1}$ for ΔE_{T+h} from (7.6), from $T = 2013$ with $h = 1 \dots 4$, conditional on the realized values for the regressors, where $\hat{\sigma}_f$ denotes the forecast standard error. We also report ‘forecasts’ from the robust device (2.27), denoted $\widehat{\Delta E}_{T+h|T+h-1}$ (see Hendry, 2006, and Section 2.9). The derived ‘forecasts’ $\widehat{E}_{T+h|T+h-1}$ for the levels E_{T+h} are also shown in Panel (b). The robust devices have slightly larger RMSFEs of 14.9, as against $\widehat{\Delta E}_{T+h|T+h-1}$ of 13.7, so the conditional ‘forecasts’ suggest no substantive shift in the relationship, despite describing the lowest levels of CO₂ emissions seen since the 19th century. However, Panel (c) shows the important role of the step-indicator for 2010 as the forecasts resulting when it is absent are systematically too high.

Re-estimating the CO₂ model up to 2017 showed almost no change in $\hat{\sigma}$ to 8.99, consistent with constancy. However, dropping S_{2010} then re-estimating to 2017 leads to a jump in $\hat{\sigma}$ to 11.1 and rejection on almost all the diagnostic tests, as does commencing forecasts from 2008, at which point the effects of the Climate Change Act would not be known. Now the advantages of the robust device come into their own as panel (c) shows. The mis-specified model’s ‘forecasts’ suffer systematic failure when $S_{\{2010\}}$ is not included (all other indicators were included), lying outside the $\pm 2\hat{\sigma}_f$ error bars for the last four observations, with a RMSFE of 36, whereas despite that omission, the robust ‘forecasts’ track the downward trend in emissions with RMSFE=25.

To obtain unconditional forecasts and evaluate the role of IIS and SIS in model development and forecasting, a vector autoregression (VAR) with 2 lags was estimated for the five variables, E_t , C_t , O_t , g_t , and k_t , over 1862–2011 with and without the indicators found for (7.4). In the former, those indicators were included in all equations. The VARs were estimated unrestrictedly without any selection to eliminate insignificant variables as that would lead to different specifications between the systems: Clements and Hendry (1995) demonstrate the validity of forecasting in this setting. Figure 7.11 Panel (a) reports the outcomes for 1-step ahead forecasts with and without the step indicators and Panel (b) the multi-step forecasts going 1, 2, . . . 6 steps ahead. In both cases, indicator-based forecast intervals are shown as fans, and without step indicators by bars. In-sample, impulse indicators only have an impact on forecasts to the extent that they change estimated parameters, whereas step indicators can have lasting effects. As can be seen, including the step indicators

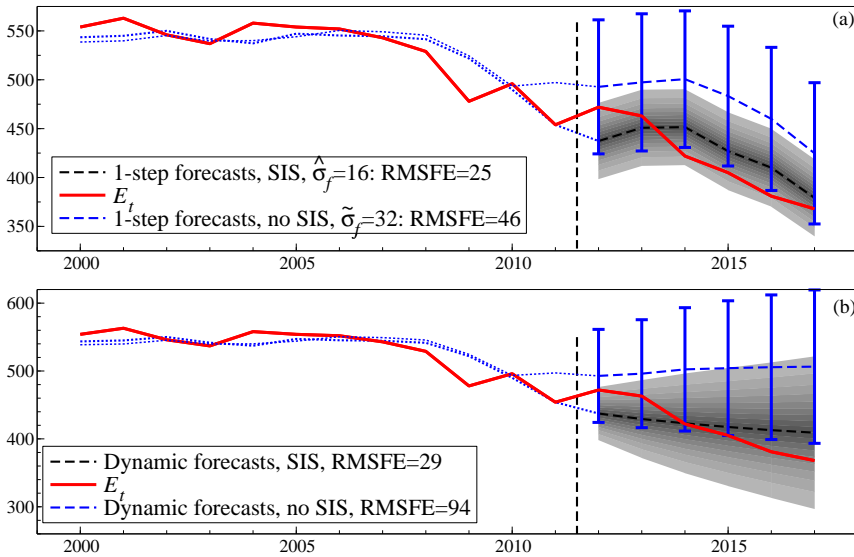


Figure 7.11: (a) Outcomes fitted values, and 1-step forecasts with and without step indicators, with $\pm 2\hat{\sigma}_f$ respectively shown as bars and fans, plus RMSFEs; (b) Outcomes fitted values, and multi-step forecasts with and without step indicators, with $\pm 2\hat{\sigma}_f$ respectively shown as bars and fans, plus RMSFEs.

greatly reduces $\pm 2\hat{\sigma}_f$ for the 1-step forecasts in Panel (a) and leads to much smaller RMSFEs in both cases as compared to when no step indicators are included. The outcomes do not always lie within the uncertainty intervals for forecasts without SIS, which under-estimate the uncertainty despite a larger $\hat{\sigma}$.

7.11 Policy implications

The most important implication of the above evidence is that substantial CO₂ reductions have been feasible, so far with little apparent impact on GDP. The UK's 2008 Climate Change Act (CCA) established the world's first legally-binding climate-change target to reduce the UK's greenhouse-gas emissions by at least 80% by 2050 from the 1990 baseline (the UK carbon budget counts six GHG emissions, not just CO₂). A range of policy initiatives was implemented, with an updated carbon plan in 2011 (again covering more than just CO₂ emissions), with carbon budgets to limit GHG emissions to 3018 Mt CO₂-equivalent over the five years 2008–2012 and 2782 over 2013–2017.

While only counting the CO₂ component, which is approximately 80% of the total, emissions over 2008–2012 cumulated to 2477 Mt, and to date over 2013–2017, to 2039 Mt, both below the sub-targets, allowing 20% for other GHG emissions while still hitting those overall targets.

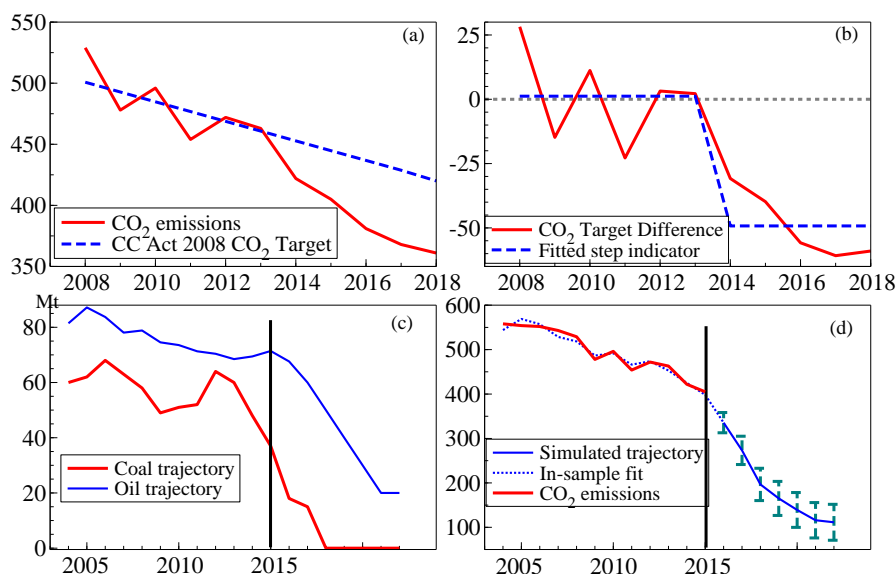


Figure 7.12: (a) UK CO₂ emissions and Climate Change Act 2008 CO₂ targets; (b) deviations from targeted values with a step indicator; (c) scenario reductions required in coal and oil use for original 2050 target; (d) resulting reductions in CO₂ emissions from (7.6). In (c) & (d), the horizon is compressed to 5-year intervals after 2017.

To test the UK's achievement of its 2008 CCA targets for CO₂, the above 5-year total targets were translated into annual magnitudes, starting 20Mt above and ending 20Mt below the average target for the period. However, our test does not depend greatly on the within-period allocation, which affects any apparent residual autocorrelation (not significant, but the sample is small). We then scaled these annual targets by 0.8 as the share of CO₂ in total greenhouse gases emitted by the UK, shown in Figure 7.12 (a). As a decade has elapsed since the Act, there were 10 annual observations on CO₂ emissions to compare to the targets, and we calculated a test of the difference between targets and outcomes being zero, but starting in 2009 as the Act could not have greatly influenced the emissions in its year of implementation. A graph of those

differences is shown in Figure 7.12 (b).

The null of “emissions=targets” is strongly rejected on the negative side with a mean of -18 and a zero-innovation error t -test value of -2.67 ($p < 0.03$; $t = -1.99$ correcting estimated standard errors for residual autocorrelation and heteroskedasticity), or as in panel (b), a downward step of -46.8 starting in 2013 with a t of -5.9 . A similar approach could be used to evaluate the extent to which countries met their Paris Accord Nationally Determined Contributions (NDCs), given the relevant data. Thus, the UK has reduced its emissions faster than the targets and in 2017 was already below the implicit target for 2018. Indeed, the budget for 2018–2022 of 2544 Mt, roughly 410 Mt p.a. of CO₂, is undemanding given the 2017 level of 368 Mt, but should not induce complacency, as the easiest reductions have been accomplished with coal use now almost negligible. The NDCs agreed at COP21 in Paris are insufficient to keep temperatures below 2°C so must be enhanced, and common time frames must be adopted to avoid a lack of transparency in existing NDCs: see Sam Rowan (2019). Since the baseline dates from which NDCs are calculated is crucial, 5-year NDC reviews and evaluation intervals are needed.

7.12 Can the UK reach its CO₂ emissions targets for 2050?

For CO₂ emissions to meet their share by 2050 of the 80% drop from the 1990 baseline of 590 Mt, as required by the 2008 CCA, they would need to fall to about 120 Mt pa. To illustrate, we simulate a scenario with no coal usage, quite a possibility now that coal is banned for electricity generation from 2025, and a 70% fall in oil use, to around 20 Mt p.a., from greatly increased use of non-gasoline vehicles sustained by expanded renewables and alternative engines. The outcome is shown in Figure 7.12 panel (c). The horizon is compressed after 2017, as the timing of such dramatic reductions is highly uncertain. Implicitly, reduced dependence on natural gas to under 35 Mtoe p.a. (a 75% reduction) is required, potentially replaced by hydrogen as the UK used to burn (but then from coal gas) before the switch starting in 1969 discussed above. With about a quarter of CO₂ emissions coming from agriculture, construction and waste (currently about 100 Mt p.a.) a serious effort to much more than halve those must also be entailed. Panel (d) records the resulting trajectory for CO₂ emissions, falling from around the 2015 level of 400 Mt p.a. to about 120 Mt p.a., or around 1.8 tonnes per capita p.a., down

from 12.4 tonnes per capita p.a. in 1970. The point and interval ‘forecasts’ are at constant K and G , and assume the parameters of (7.6) remain constant despite the major shift: increases in K and G would make the targets harder to achieve unless they were carbon neutral. However, given the key role of the capital stock in explaining the UK’s CO₂ emissions since 1860, as K embodies the vintage of the technology at the time of its construction and is long lived, transition to zero carbon has to be gradual, and necessitates that new capital, and indeed new infrastructure in general, must be zero carbon producing. As a ‘policy’ projection, together these measures would reach the UK’s 2008 announced 2050 target—but only if such reductions, perhaps with offsetting increases, could actually be achieved.

The rapidly falling costs of renewable-energy sources like solar cells and wind turbines (see e.g., Doyne Farmer and François Lafond, 2016) combined with improved storage methods should substantially reduce oil and gas use in electricity production. Table 7.5 records recent estimates of electricity generating costs in £/MWh by different technologies.

Power generating technology costs £/MWh	Low	Central	High
Nuclear PWR (Pressurized Water Reactor)(a)	82	93	121
Solar Large-scale PV (Photovoltaic)	71	80	94
Wind Onshore	47	62	76
Wind Offshore	90	102	115
Biomass	85	87	88
Natural Gas Combined Cycle Gas Turbine	65	66	68
CCGT with CCS	102	110	123
Open-Cycle Gas Turbine	157	162	170
Advanced Supercritical Coal Oxy-comb. CCS	124	134	153
Coal IGCC with CCS (b)	137	148	171

Table 7.5: Electricity generating technology costs in £/MWh (megawatt hour). (a) New nuclear power guaranteed strike price of £92.50/MWh for Hinkley Point C in 2023; (b) IGCC =Integrated Gasification Combined Cycle. Lowest cost alternatives shown in bold. Source: *Electricity Generation Costs*, Department for Business, Energy and Industrial Strategy (BEIS), November 2016.

Onshore wind turbines have fallen in cost and increased in efficiency so rapidly over the past two decades that for the UK at least they offer the lowest

cost alternative (with the incidental benefit of creating marine reserves and saltwater fish sanctuaries), even below natural gas combined-cycle turbines before the costs of carbon capture and storage (CCS) are included.⁵ Solar photovoltaics come next (and this is the UK!) if CCS is enforced, though both require large backup electricity storage systems for (e.g.) windless nights. Renewables' share of overall electricity generation reached a peak of 60.5 per cent at one stage in April 2020, according to National Grid data. Increased outputs of renewable electricity will reduce the volume of emissions for a given level of energy production by also reducing usage of oil in transport through electric car use, but should not influence the constancy of the empirical models above conditional on the volumes of coal and oil included.

A probable reason for the sharp fall in coal use in 2017 is a rise in its price relative to those of other energy sources, with the UK carbon tax doubling in 2015 to £18 per tonne of CO₂. Conversely, natural gas use has increased 3.5 fold since the mid-1980s, so although producing less than half the CO₂ emissions of coal per Btu, still contributes about 140 Mt p.a. to CO₂ emissions. The use of oil in transport will take longer to reduce, but more efficient engines (with diesel being phased out completely given its toxic pollutants), and most vehicles powered from renewable sources, combined with much higher taxes on gasoline, offer a route to the next stage of CO₂ emissions reductions.

In 2019, the UK Government amended the original CCA target to zero net emissions by 2050. Then all the sources of CO₂ emissions must go to a level such that carbon capture and sequestration (CCS), possibly combined with atmospheric CO₂ extraction methods, would remove the rest. Facing an almost certain irreducible non-zero minimum demand for oil and gas (e.g. for chemicals), to achieve the Paris COP21 target of zero net emissions before 2050 requires really major technological change, almost certainly involving development of current research avenues into removing or using existing CO₂: see <https://phys.org/news/2014-09-carbon.html>. Net zero is an excellent target, but incredibly difficult to achieve, and as yet there is no sensible strategy to do so...although we now discuss some highly speculative routes.

To meet the net zero target, overall natural gas use would need to be reduced to near zero, like coal. Natural gas is mainly used for electricity

⁵<https://institutions.newscientist.com/article/2217235-the-cost-of-subsidising-uk-wind-farms-has-dropped-to-an-all-time-low/> show a more recent cost of £40MWh, now highly competitive.

production and household indoor and water heating. The former could be handled in part by increased renewable sources, but as potentially serious storage problems remain (more on this shortly), research effort should be devoted to developing safe small modular nuclear reactors (SMRs) based on well developed nuclear-powered engines in submarines.⁶ These SMRs might be able to use thorium or the ‘spent’ uranium fuel rods from older reactors, helping reduce the serious problem of existing nuclear-waste disposal. Household use could be reduced by increased taxes on natural gas and oil usage, encouraging the adoption of solar panels and (e.g.) air heat pumps, as well as switching the national gas system back to its pre-1969 hydrogen basis.⁷ Over the next 20–30 years with ever improved technologies, and consequential cost reductions in generating electricity by renewables, a zero target does not seem impossible for electricity and gas without requiring reductions in GDP growth, perhaps even increasing it with new opportunities.

Next, road transport decarbonisation is progressing slowly using lithium-ion battery powered electric vehicles, with relatively short journey capacity yet take a non-negligible time to recharge, discouraging the replacement of internal combustion engines. Huge advances have occurred in recent years both in understanding the properties of graphene, and in its cost of production (see ‘graphene in a flash’ from plastic waste in <https://phys.org/news/2020-01-lab-trash-valuable-graphene.html>). Graphene nanotubes (GNTs) can act as electrode supercapacitors (see e.g.: <https://www.nature.com/articles/s41598-020-58162-9>). Thus, one could imagine an electric vehicle that sandwiched an array of GNTs between two Faraday cages over the roof and above the inside of a vehicle, or by a prefabricated modular unit fitted to the existing roof (perhaps even retrofitted on existing car roofs) to power an electric motor, so the vehicle becomes the battery. GNTs seem capable of rapid charging, and should be able to sustain viable distances on a single charge.

Moreover, if successful, once most vehicles are like that, then mandating them to be plugged in when not in use, a vast electric storage system would be available for no additional investment, so renewable sources of electricity could be widely adopted without worrying about security of supply. There are

⁶www.world-nuclear.org/info/Nuclear-Fuel-Cycle/Power-Reactors/Small-Nuclear-Power-Reactors/

⁷Such taxes are to change behavior, not to raise revenue, so should be redistributed to families facing fuel poverty.

undoubtedly many key technical issues needing research as to how such a system would work in practice, some ongoing such as developing 2-dimensional tri-layers of graphene as an insulator, superconductor and magnet.⁸

The potential benefits of such a power source could be huge as a 'sensitive intervention point' (SIP: see <https://science.sciencemag.org/content/364/6436/132>). By not demonising road transport for its CO₂ footprint and dangerous pollution, cars with internal combustion engines could be replaced at a rate matching the increased need for storage from the extension of renewables. The basics of electric engines are established, so employment can be maintained in vehicle manufacture and all its ancillary industries as well as in new graphene-based ones. Two side benefits are a major reduction in both mining for lithium, and later disposal of, or recycling, the resulting toxic battery waste; and eliminating the need for expensive catalytic converters, cutting production costs markedly, eliminating a target for theft (which then exacerbates air pollution), and reducing palladium mining.

Indirect consequences could solve the UK's rail systems problem of a lack of electrification across much of the network by replacing diesel-electric trains by GNT-supplied electric ones, although some progress is occurring with hydrogen driven trains in Germany and the UK (see <https://www.birmingham.ac.uk/research/spotlights/hydrogen-powered-train.aspx>); and even more speculatively, as GNTs are so light, possibly short-haul electric aircraft. As an historical aside, we noted above that electric cars date back before the 1880s, so an all-electric transport system is just going back to where society might have been 140 years ago.

However, agriculture, construction, chemical industry and waste management look more problematic, although there is progress in efficiency improvements. Inner-city vertical and underground farms economize on water, fertilizer and energy (especially from transport reductions) and are increasingly viable given the falls in costs for LED lighting (see e.g., www.scientificamerican.com/article.cfm?id=the-rise-of-vertical-farms). There is considerable research on altering farm mammal diets to reduce methane emissions, including adding dietary fumaric acid (from plants like lichen and Iceland moss), where lambs showed a reduction by up to

⁸See <https://www.graphene-info.com/graphene-triples-superconducting-insulating-and-ferromagnetic>

70% (e.g., <https://phys.org/news/2008-03-scientists-cow-flatulence.html>). Changes to human diets are also en route, and need encouraging: on a small scale, see <http://www.climateeconometrics.org/2020/03/18/nuffield-colleges-decreasing-food-emissions/>. Prefabrication of highly insulated dwellings must be a priority, as well as using less GHG-intensive building materials. Recycling more, using more waste for fuel, and landfilling less to reduce methane are all essential.

The UK's total 'consumption induced' CO₂ equivalent emissions are higher than the domestic level through CO₂ embodied in net imports,⁹ although the large reductions achieved to date have a major domestic component, and of course 'consumption induced' CO₂ will fall as the CO₂ intensity of imports falls with reductions in exporting countries. However, targeting consumption rather than production emissions has the unwanted consequence of removing any incentives for emitting industries or exporting countries to improve their performance, as these would not be counted against them (e.g., if NDCs used a consumption basis). Border carbon taxes have a role to play in improving both exporters and importers performance. Similarly, allocating emissions from transport and packaging to (say) the food sector would again alleviate those intermediate sectors of the responsibility to invest to reduce what are in fact their emissions by attributing them to retail outlets or consumers. Conversely, the purchasing clout of large retail chains can pressure suppliers to improve, as (e.g.) Walmart is doing.¹⁰

The aggregate data provide little evidence of high costs to the reductions achieved in CO₂ emissions, which have dropped by 186Mt from 554Mt to 368Mt (34%) so far this century, during which period real GDP has risen by 35%, despite the 'Great Recession' but before the pandemic. Historically, those in an industry that was being replaced (usually by machines) lost out and bore what should be the social costs of change, from cottage spinners, weavers and artisans in the late 18th–early 19th centuries (inducing 'Luddites'), to recent times (from a million coal miners in 1900 to almost none today). There is a huge difference in the impacts of substitutes and complements for existing methods: motor vehicles were a huge advance, and created many new jobs

⁹See <http://www.emissions.leeds.ac.uk/chart1.html> and <https://www.biogeosciences.net/9/3247/2012/bg-9-3247-2012.html>.

¹⁰See <https://corporate.walmart.com/2016grr/enhancing-sustainability/reducing-energy-intensity-and-emissions>.

directly and indirectly, mainly replacing horses but indirectly destroying their associated workforce. Although not a direct implication of the aggregate model here, greater attention needs to be focused on the local costs of lost jobs as new technologies are implemented: mitigating inequality impacts of climate induced changes ought to matter centrally in policy decisions.

Given the important role of the capital stock in the model above, ‘stranded assets’ in carbon producing industries are potentially problematic as future legislation imposes ever lower CO₂ emissions targets to achieve zero net emissions (see Pfeiffer *et al.* 2016). As argued by Farmer, Hepburn, Matthew Ives and Thomas Hale *et al.* (2019) exploiting sensitive intervention points in the post-carbon transition could be highly effective, and they cite the UK’s Climate Change Act of 2008 as a timely example that had a large effect.

An excellent ‘role model’ that offers hope for reductions in other energy uses is the dramatic increases in lumen-hours per capita consumed since 1300 of approximately 100,000 fold yet at one twenty-thousandth the price per lumen-hour (see Roger Fouquet and Peter Pearson, 2006).

7.13 Climate-environmental Kuznets curve

The ‘environmental Kuznets curve’ is assumed to be a \cap shaped relationship between pollution and economic development: see Susmita Dasgupta, Benoit Laplante, Hua Wang and David Wheeler (2002) and David Stern (2004). For a ‘climate-environmental Kuznets curve’, we estimated a regression of the log of CO₂ emissions, denoted e_t (lower case denotes logs) on the log of real GDP, g_t and its square g_t^2 , which delivered:

$$\hat{e}_t = - \underset{(1.6)}{31.5} + \underset{(0.27)}{6.13} g_t - \underset{(0.012)}{0.247} g_t^2 \quad (7.7)$$

$$\begin{aligned} \hat{\sigma} &= 0.091 \quad R^2 = 0.91 \quad F_{AR}(2, 145) = 37.3^{**} \quad F_{ARCH}(1, 148) = 0.26 \\ F_{Het}(3, 146) &= 1.70 \quad \chi_{nd}^2(2) = 68.3^{**} \quad F_{Reset}(2, 145) = 10.61^{**} \\ F_{Chow}(5, 147) &= 3.06^* \quad F_{nl}(6, 141) = 8.72^{**} \end{aligned}$$

Many of the diagnostic tests are significant, and both F_{reset} and F_{nl} reveal that all of the non-linearity has not been captured by (7.7). Indeed, (7.7) has a borderline rejection on the parameter-constancy test, but the rejections on the other mis-specification tests makes that difficult to interpret. Full-sample

impulse-indicator saturation (IIS) selected **17** indicators at a significance level of 0.1%, but still led to $F_{nl}(6, 128) = 11.5^{**}$, and $\hat{\sigma} = 0.055$.

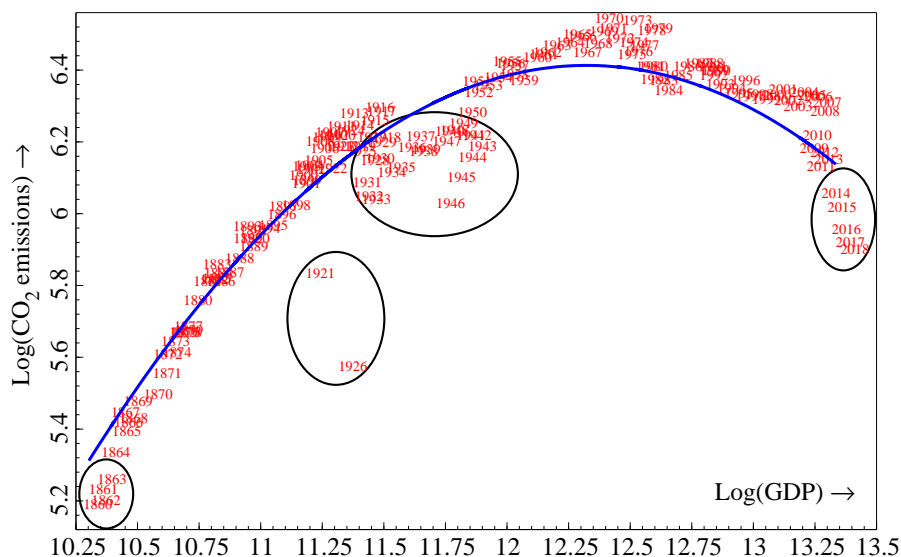


Figure 7.13: Scatter plot of log of CO₂ emissions against the log of GDP (shown by dates) with the fitted values from equation (7.7) (shown by the line).

The relationship between log CO₂ emissions and log real GDP is plotted in Figure 7.13. The large drop in CO₂ emissions while GDP more than doubled is notable, and reflects improved technology in energy use as well as a changing mix of fuels. Although the non-linearity is marked, there are large and systematic deviations from the fitted curve, shown inside ellipses for the start and end of the sample, 1921 & 1926, and the 1930s & 1940s.

Since the final model in (7.6) is linear in CO₂ emissions, and log-linear in GDP, a natural question is whether it can account for the non-linearity of the ‘climate Kuznets curve’ in Figure 7.13. This is answered in Figure 7.14 where the log of the fitted values from (7.6) is cross plotted against log(GDP) together with log(CO₂) data, to reveal the same non-linearity even though log(GDP) enters the equilibrium-correction mechanism in (7.5) negatively and is insignificant. The regression of log(CO₂) on the log of the fitted values from (7.6) had $\hat{\sigma} = 0.019$. Of course that better explanation is greatly enhanced by using coal and oil, but conversely is after translation into logs.

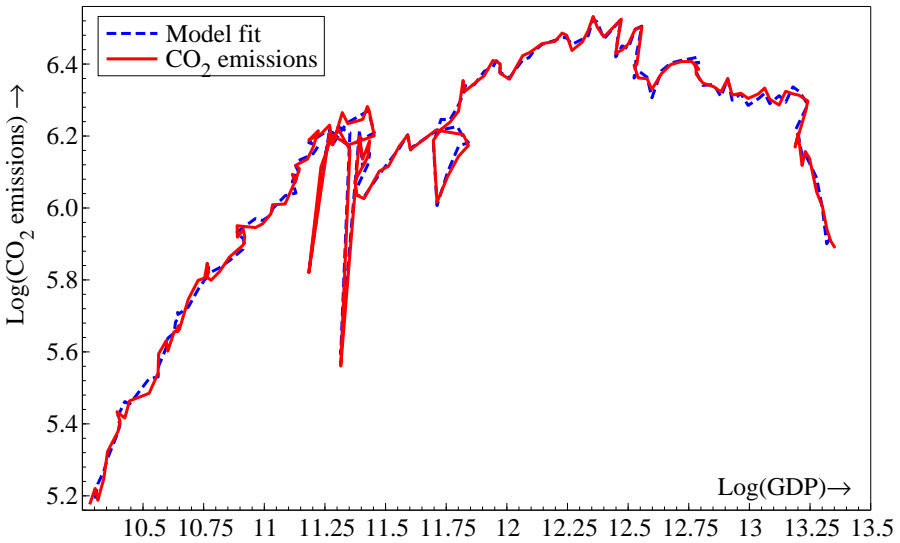


Figure 7.14: Plot of UK $\log(\text{CO}_2)$ emissions and log fitted values against $\log(\text{GDP})$: re-creating a ‘climate Kuznets Curve’.

Thus, the ‘curvature’ of an eventually declining relationship between $\log \text{CO}_2$ emissions and \log real GDP is an artefact of both being correlated with technology. Had electricity been discovered in 1300, batteries several decades later, rather than waiting for Volta in 1800, and solar cell technologies a few decades after that and so on, all of which depended on knowledge and understanding rather than income levels per se, an electrical world economy may have circumvented the need for coal. Conversely, if neither electricity nor the internal combustion engine had been discovered, leaving only coal as a fuel source, efficiency improvements or lower usage would have been the only routes to reductions in CO_2 emissions. Relative costs of energy provision matter, and Table 7.5 showed recent power generating costs, but the metaphor suggests a ‘climate Kuznets curve’ is mainly a technology-driven relation. Income levels may matter more for other environmental relations such as clean air and less polluted rivers.

8

Conclusions

Climate econometrics aims to apply econometric methods to augment our understanding of climate change and the interactions between human actions and climate responses. The field has evolved rapidly over the last few years from a pressing need to understand the science, economics, health and policy implications of climate change. Time-series econometrics is well placed to offer insights on these issues, as the methodology has been developed to model complex, evolving and shifting interactions over time due to human behavior. Although originally applied to economic data, the methodology is applicable to climate-economic research as anthropogenic forces play a key role in determining climate and vice versa. This review aimed to explain the tools developed at Climate Econometrics (<http://www.climateeconometrics.org/>) to disentangle complex relationships between human actions and climate responses and their associated economic effects, masked by stochastic trends and breaks. Empirical applications to climate problems demonstrate the benefits of applying data-based methods jointly with underlying theory to improve our understanding of climate phenomena, to assist in forecasting future outcomes and to provide policy guidance.

We described novel modeling approaches to climate-economic data, combining insights from climate theory with empirical evidence to discover new

results. We embed theory models in far larger information sets, allowing new features, dynamics, outliers, breaks and non-linearities to be discovered, while retaining established theory. One of the fundamental aspects of the modeling approach is the use of multi-path selection that enables more candidate variables than observations to be explored. This opens the door to a wide range of indicator saturation estimators to model outliers, breaks, distributional shifts and non-constancies. We discuss the costs of selection relative to mis-specification and show the remarkably small costs associated with searching over large numbers of candidate variables, thus enabling wide-sense non-stationary data to be modeled.

The monograph emphasises the importance of taking account of the non-stationary nature of time series, both stochastic trends and distributional shifts. The example in Chapter 3 looking at distances traveled by car and human road fatalities illustrates the hazards of not correctly dealing with non-stationary time series. Such an example highlights that mistaken inferences and possibly false causal attribution can occur if the data properties are not carefully taken into account and relationships modeled with rigorous testing.

Our short excursion into climate science in Chapter 4 discusses the Earth's atmosphere and oceans, demonstrates that humanity can easily alter these, and shows it is doing so. The composition of the atmosphere, the roles of CO₂ and other greenhouse gases, and the consequences of changes in atmospheric composition are discussed, focusing on the impacts of climate change on the 'great extinctions' over geological time. Chapter 5 notes that the consequences of the Industrial Revolution in the UK during the 17th century have been both good and bad, greatly raising living standards worldwide, but leading to dangerous levels of CO₂ emissions from using fossil fuels.

The econometric approach outlined in Chapter 2 was illustrated by two detailed applications. Chapter 6 applies the modeling approach to the last 800,000 years of ice ages to illustrate its practical details. The theoretical model of ice ages is based on variations in the Earth's orbit, which determine the solar radiation reaching the planet, where and when it is most concentrated and hence the speed with which glacial periods occurred and later retreated. But the question then arises that if ice ages are due to orbital variations, why should CO₂ levels correlate so highly negatively with land ice volume? Ice, CO₂ and temperature are modeled as jointly endogenous functions of the orbital variables in a 3-variable simultaneous equations system, applying saturation

estimation on the system to model outliers, along with dynamics, and non-linearities to capture interaction effects. The approach embeds the theory and allows for dynamics, non-linearities, non-stationarities and endogeneity in a system, rebutting concerns that the modeling approach is inherently single equation. The evidence suggests that CO_2 was an endogenous response to orbital drivers over Ice Ages, jointly with ice volume and temperature, albeit now mainly determined by anthropogenic sources. Looking into the future with CO_2 changing to an exogenously determined value set by anthropogenic emissions points to temperatures dangerously in excess of the peak values measured over the Ice Ages.

Chapter 7 developed an explanation of the United Kingdom's CO_2 emissions data over 1860–2017 in terms of coal and oil usage, capital stock and GDP, taking account of their non-stationary nature, with many turbulent periods and major shifts over the 157 years. Having been first into the Industrial Revolution that has transformed the world's wealth at the cost of climate change, the UK is one of the first out in terms of its CO_2 emissions; the UK's total CO_2 emissions have dropped below the level first reached in 1894, and per capita UK CO_2 emissions are now below their level in 1860, when the UK was the 'workshop of the world', and yet per capita real incomes are more than 7-fold higher. The econometric approach to modeling such dramatic changes was explained in four steps. The key explanatory variables were coal and oil usage and capital stock, whereas GDP had an insignificant effect in levels given the other explanatory variables, possibly reflecting a move away from manufacturing to a service economy, notwithstanding which, the model implies a non-linear 'climate Kuznets curve' between emissions and GDP. Compared to directly fitting a 'climate Kuznets curve' as in (7.7), the resulting model highlights the benefits of the more general methodology. Improvements in multi-step forecasts also highlighted the advantages of taking account of in-sample outliers and shifts using impulse- and step-indicator saturation, despite those creating more candidate variables to select over than observations.

The UK's climate policy has been effective, and the resulting large emissions reductions have not yet involved major aggregate sacrifices. The UK's reductions (e.g., of 24 Mt in 2016) are the more impressive against recent recorded global annual increases of 3.3ppm (remember that 1ppm = 7.81 gigatonnes of CO_2). However, local losses in incomes and employment from changes in fuel production have not been addressed and 'stranded assets' could

be a potential problem if future legislation imposes lower CO₂ emissions targets. The UK's target of a 100% reduction from the 1990 baseline of 590 Mt of emissions is only achievable with complete reductions in oil and gas use and in other sources net of increased re-absorption of CO₂.

We conclude by briefly describing some of the other applications of methods developed at Climate Econometrics, including to panels and cross-sections, to demonstrate the diverse array of problems to which our approaches can be applied, including to health care as well as economic modeling. Their common feature is to emphasize the role of human behavior in climate change additional to well-established physical processes.

To determine any anthropogenic signature in global CO₂ concentrations, Hendry and Pretis (2013) modeled anthropogenic and natural contributions to atmospheric CO₂ using a large number of potential explanatory variables: their general unrestricted model had 492 variables for 246 monthly observations. Despite approximately 10^{148} possible models to select over, *Autometrics* estimated just 571 models at $\alpha = 0.001$, and finally selected 14 variables. No deterministic terms were retained, so all the explanation of changes in global CO₂ concentrations came from stochastic sources, demonstrating that the modeling approach does not impose any particular stance on likely causes. Increases in cumulated changes in vegetation reduced atmospheric levels of CO₂ by about 2ppm with large seasonal fluctuations, whereas the Southern Oscillation Index (SOI) increased it by about 3ppm, both over 1981(7)–2002(12), revealing small net effects from natural sources. However, the cumulation of anthropogenic sources over that period produced a strong trend from 340ppm to almost 380ppm, closely matching the CO₂ measured at Mauna Loa, and consistent with isotope-based measures.

Pretis *et al.* (2015a) applied SIS to objectively investigate a slowdown in the rise of global mean surface temperature (called the hiatus in warming). Their results indicated that when temperature can be modeled by anthropogenic forcings and natural variability such as solar insolation, the hiatus was not statistically different from the rest of the 1950–2012 temperature record. They also found no evidence that the slowdown in temperature increases was uniquely tied to episodes of La Niña-like cooling.

Pretis and Max Roser (2017a) compared socio-economic scenarios created in 1992 and 2000 against the observational record to investigate the possible decoupling of economic growth and fossil-fuel CO₂ emissions. They showed

that global emission intensity (fossil fuel CO₂ emissions relative to GDP) rose in the first part of the 21st century, primarily from the under-predicted rapid growth in Asia, counter to some climate projections foreseeing a decline. Nevertheless, the wide spread of temperature changes in climate projections did not predominately originate from uncertainty across climate models, but from the broad range of different global socio-economic scenarios and their implied energy production. As discussed in Section 2.9 above, forecasting the future when human behavior is involved is always prone to unanticipated shifts.

Kaufmann, Mann, and Sucharita Gopal *et al.* (2017) demonstrated that spatial heterogeneity of climate-change experiences mattered, since skepticism about whether the Earth was warming was greater in areas exhibiting cooling relative to areas that had warmed. Moreover, recent cooling could offset historical warming to enhance skepticism. While *climate change* is due to global warming, the former is a better epithet since not all areas warm uniformly as a consequence of changes.

Pretis, Moritz Schwarz, Kevin Tang, Karsten Haustein and Allen (2018b) applied a panel-data version of IIS to analyze the potential impacts on economic growth of stabilizing global temperatures at either 1.5°C or 2°C above pre-industrial levels. They estimated a non-linear response of changes in global annual per capita GDP growth to average annual temperatures, both without and with IIS. They found that stabilizing temperature increases at 1.5°C did not lead to too many major costs in reducing per capita GDP growth across countries, although some already poor countries in the tropics would suffer. However, temperatures rising by 2°C would inflict serious losses on most already-hot countries, lowering their projected GDP per capita growth by up to 2.5% per annum, consistent with other empirical estimates. At a global level, 1.5°C is preferable to 2°C above pre-industrial levels and is still (possibly) achievable.

Tropical cyclones, including hurricanes in the West Atlantic and typhoons in the Pacific, are often highly destructive: see Kerry Emanuel (2005). Four of the five costliest US natural disasters have been caused by relatively recent hurricanes, so Martinez (2020) applied the model selection methods and saturation estimators described above to examine whether improvements in forecasting could reduce their damages as earlier warnings would allow more time to prepare or evacuate. A measure of forecast uncertainty was added

to the main natural and human channels determining damages, the former including highest wind speeds, minimum pressure, maximum storm surge, maximum rainfall, seasonal cyclone energy, historical frequency of hurricanes, soil moisture content and air temperature; and the latter including income, population, and the number of housing units in the area at risk, together with other candidate variables such as time of year and the location hit etc. Applying *Autometrics* with IIS to the 98 hurricane strikes to the Eastern USA since 1955, minimum pressure, maximum storm surge, maximum rainfall, income, housing units, and forecast uncertainty were selected.¹

All of these studies had theory guidance in formulating their approaches, but the econometric techniques discussed above also allow for direct linking of climate models with empirical data, as in Pretis (2019), to further improve econometric research on human responses to climate variability. This monograph has emphasized the need for handling all modeling decisions jointly, allowing for key forms of wide-sense non-stationarity, facing possibly incorrect theories and mis-measured data. Few approaches in either climate or economic modeling as yet consider all such effects jointly, but a failure to do so can lead to mis-specified models and hence incorrect theory evaluation, mis-leading forecasts and poor policy analyses. The software to implement our approach is available as *Autometrics* in *OxMetrics8*, in *XLModeler* for Excel, and as *Gets* in R, so can be readily accessed.

¹His web site <https://sites.google.com/view/andrewbmartinez/current-research/damage-prediction-tool> provides a tool based on his model for predicting the damages after a hurricane strike to the Atlantic coast of the USA, which has proved accurate for most of the hurricanes in recent years.

References

- Agassiz, L. 1840. *Études sur les glaciers*. Digital book on Wikisource accessed on July 22, 2019: https://fr.wikisource.org/w/index.php?title=%C3%89tudes_sur_les_glaciers&oldid=297457. Neuchâtel: Imprimerie de OL Petitpierre.
- Akaike, H. 1973. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Second International Symposium on Information Theory*. Ed. by B. N. Petrov and F. Csaki. Budapest: Akademia Kiado. 267–281.
- Allen, R. C. 2009. *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press.
- Allen, R. C. 2017. *The Industrial Revolution: A Very Short Introduction*. Oxford: Oxford University Press.
- Andrews, D. W. K. 1991. “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation”. *Econometrica*. 59: 817–858.
- Arrhenius, S. A. 1896. “On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground”. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (fifth series)*. 41: 237–275. <http://www.globalwarmingart.com/images/1/18/Arrhenius.pdf>.
- Beenstock, M., Y. Reingewertz, and M. Paldor. 2012. “Polynomial cointegration tests of anthropogenic impact on global warming”. *Earth Systems Dynamics*. 3: 173–188.

- Berenguer-Rico, V. and I. Wilms. 2020. "Heteroscedasticity testing after outlier removal". *Econometric Reviews*: DOI: 10.1080/07474938.2020.1735749.
- Berenguer-Rico, V. and J. Gonzalo. 2014. "Co-summability: From Linear to Non-linear Co-integration". *Working paper*. Oxford University: Economics Department.
- Blundell, S. 2012. *Magnetism: A Very Short Introduction*. Oxford: Oxford University Press.
- Bontemps, C. and G. E. Mizon. 2008. "Encompassing: Concepts and Implementation". *Oxford Bulletin of Economics and Statistics*. 70: 721–750.
- Boumans, M. A. and M. S. Morgan. 2001. "Ceteris Paribus Conditions: Materiality and the Applications of Economic Theories". *Journal of Economic Methodology*. 8: 11–26.
- Bralower, T. J. 2008. "Earth science: Volcanic cause of catastrophe". *Nature*. 454: 285–287. doi:10.1038/454285a.
- Brinkley, C. 2014. "Decoupled: successful planning policies in countries that have reduced per capita greenhouse gas emissions with continued economic growth". *Environment and Planning C: Government and Policy*. 32: 1083–1099.
- Brusatte, S. 2018. "How the dinosaurs got lucky". *Scientific American*. 318(5): 18–25.
- Buchanan, P. J., Z. Chase, R. J. Matear, S. J. Phipps, and N. L. Bindoff. 2019. "Marine nitrogen fixers mediate a low latitude pathway for atmospheric CO₂ drawdown". *Nature Communication*. 10. <https://doi.org/10.1038/s41467-019-12549-z>.
- Burke, M., W. M. Davis, and N. S. Diffenbaugh. 2018. "Large potential reduction in economic damages under UN mitigation targets". *Nature*. 557: 549–553.
- Burke, M., S. M. Hsiang, and E. Miguel. 2015. "Global non-linear effect of temperature on economic production". *Nature*. 527: 235–239.
- Caceres, C. 2007. "Asymptotic Properties of Tests for Mis-Specification". *Unpublished Doctoral Thesis*. Oxford University: Economics Department.
- Castle, J. L., M. P. Clements, and D. F. Hendry. 2015a. "Robust Approaches to Forecasting". *International Journal of Forecasting*. 31: 99–112.
- Castle, J. L., J. A. Doornik, and D. F. Hendry. 2011. "Evaluating Automatic Model Selection". *Journal of Time Series Econometrics*. 3 (1): DOI: 10.2202/1941–1928.1097.

- Castle, J. L., J. A. Doornik, and D. F. Hendry. 2012. "Model Selection when there are Multiple Breaks". *Journal of Econometrics*. 169: 239–246.
- Castle, J. L., J. A. Doornik, and D. F. Hendry. 2019a. "Modelling Non-stationary 'Big Data'". *Working Paper*. Oxford University: Nuffield College.
- Castle, J. L., J. A. Doornik, and D. F. Hendry. 2019b. "Multiplicative-indicator Saturation". *Working Paper*. Oxford University: Nuffield College.
- Castle, J. L., J. A. Doornik, and D. F. Hendry. 2019c. "Robustly Selecting Regression Models". *Working Paper*. Oxford University: Economics Department.
- Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis. 2015b. "Detecting Location Shifts During Model Selection by Step-Indicator Saturation". *Econometrics*. 3(2): 240–264.
- Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis. 2019d. "Trend-indicator Saturation". *Working Paper*. Oxford University: Nuffield College.
- Castle, J. L., N. W. P. Fawcett, and D. F. Hendry. 2010. "Forecasting with Equilibrium-correction Models during Structural Breaks". *Journal of Econometrics*. 158: 25–36.
- Castle, J. L. and D. F. Hendry. 2010. "A Low-Dimension Portmanteau Test for Non-linearity". *Journal of Econometrics*. 158: 231–245.
- Castle, J. L. and D. F. Hendry. 2014a. "Model Selection in Under-specified Equations with Breaks". *Journal of Econometrics*. 178: 286–293.
- Castle, J. L. and D. F. Hendry. 2014b. "Semi-Automatic Non-linear Model Selection". In: *Essays in Nonlinear Time Series Econometrics*. Ed. by N. Haldrup, M. Meitz, and P. Saikkonen. Oxford: Oxford University Press. 163–197.
- Castle, J. L. and D. F. Hendry. 2019. *Modelling our Changing World*. <https://link.springer.com/book/10.1007%2F978-3-030-21432-6>. Palgrave Mcmillan: London.
- Castle, J. L., D. F. Hendry, and A. B. Martinez. 2017. "Evaluating Forecasts, Narratives and Policy using a Test of Invariance". *Econometrics*. 5(39): DOI:10.3390/econometrics5030039.
- Castle, J. L. and N. Shephard, eds. 2009. *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.
- Cheng, L., J. Abraham, Z. Hausfather, and K. E. Trenberth. 2019. "How fast are the oceans warming?" *Science*. 363(6423): 128–129.

- Chow, G. C. 1960. "Tests of Equality between Sets of Coefficients in Two Linear Regressions". *Econometrica*. 28: 591–605.
- Clarke, A. 1993. "Temperature and Extinction in the Sea: A Physiologist's View". *Paleobiology*. 19: 499–518.
- Clements, M. P. and D. F. Hendry. 1995. "Forecasting in Cointegrated Systems". *Journal of Applied Econometrics*. 10: 127–146. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.
- Clements, M. P. and D. F. Hendry. 1998. *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- CO2 Program Scripps. 2010. "The Keeling Curve". http://scrippsco2.ucsd.edu/history_legacy/keeling_curve_lessons. La Jolla, CA: Scripps Institution of Oceanography.
- Cox, D. R. 1962. "Further Results on Tests of Separate Families of Hypotheses". *Journal of the Royal Statistical Society*. B, 24: 406–424.
- Crafts, N. F. R. 2003. "Is Economic Growth Good for Us?" *World Economics*. 4(3): 35–49.
- Croll, J. 1875. *Climate and Time in Their Geological Relations, A Theory of Secular Changes of the Earth's Climate*. New York: D. Appleton.
- Dasgupta, S., B. Laplante, H. Wang, and D. Wheeler. 2002. "Confronting the Environmental Kuznets Curve". *Journal of Economic Perspectives*. 16: 147–168.
- Davis, W. M. 2019. "Dispersion of the temperature exposure and economic growth: panel evidence with implications for global inequality". *Thesis submitted in partial fulfilment for the MPhil degree*. Oxford: Economics Department.
- Dickey, D. A. and W. A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root". *Econometrica*. 49: 1057–1072.
- Doob, J. L. 1953. *Stochastic Processes*. 1990 edition. New York: John Wiley Classics Library.
- Doornik, J. A. 2008. "Encompassing and Automatic Model Selection". *Oxford Bulletin of Economics and Statistics*. 70: 915–925.
- Doornik, J. A. 2009. "Autometrics". In: *The Methodology and Practice of Econometrics*. Ed. by J. L. Castle and N. Shephard. Oxford: Oxford University Press. 88–121.

- Doornik, J. A. and K. Juselius. 2018. *CATS 3 for OxMetrics*. London: Timberlake Consultants Press.
- Doornik, J. A. and H. Hansen. 2008. "An Omnibus Test for Univariate and Multivariate Normality". *Oxford Bulletin of Economics and Statistics*. 70: 927–939.
- Doornik, J. A. and D. F. Hendry. 2015. "Statistical Model Selection with Big Data". *Cogent Economics and Finance*: DOI:10.1080/23322039.2015.1045216. URL: <https://www.tandfonline.com/doi/full/10.1080/23322039.2015.1045216#.VYE5bUYsAsQ>.
- Doornik, J. A. and D. F. Hendry. 2017. "Automatic Selection of Multivariate Dynamic Econometric Models". *Unpublished typescript*. University of Oxford: Nuffield College.
- Doornik, J. A. and D. F. Hendry. 2018. *Empirical Econometric Modelling using PcGive: Volume I*. 8th. London: Timberlake Consultants Press.
- Duffy, J. A. and D. F. Hendry. 2017. "The Impact of Near-Integrated Measurement Errors on Modelling Long-run Macroeconomic Time Series". *Econometric Reviews*. 36: 568–587.
- Emanuel, K. 2005. *Divine Wind: The History and Science of Hurricanes*. Oxford: Oxford University Press.
- Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity, with Estimates of the Variance of United Kingdom Inflation". *Econometrica*. 50: 987–1007.
- Engle, R. F. and D. F. Hendry. 1993. "Testing Super Exogeneity and Invariance in Regression Models". *Journal of Econometrics*. 56: 119–139.
- Engle, R. F., D. F. Hendry, and J.-F. Richard. 1983. "Exogeneity". *Econometrica*. 51: 277–304.
- Erickson, D., R. Mills, J. Gregg, T. J. Blasing, F. Hoffmann, R. Andres, M. Devries, Z. Zhu, and S. Kawa. 2008. "An estimate of monthly global emissions of anthropogenic CO₂: Impact on the seasonal cycle of atmospheric CO₂". *Journal of Geophysical Research*. 113.
- Ericsson, N. R. 1983. "Asymptotic Properties of Instrumental Variables Statistics For Testing Non-nested Hypotheses". *Review of Economic Studies*. 50: 287–303.
- Ericsson, N. R. 2012. "Detecting crises, jumps, and changes in regime". *Working paper*. Federal Reserve Board of Governors, Washington, D.C.

- Ericsson, N. R. 2017a. "How biased are U.S. Government Forecasts of the Federal Debt?" *International Journal of Forecasting*. 33: 543–559.
- Ericsson, N. R. 2017b. "Interpreting estimates of forecast bias". *International Journal of Forecasting*. 33: 563–568.
- Ericsson, N. R. and J. G. MacKinnon. 2002. "Distributions of Error Correction Tests for Cointegration". *Econometrics Journal*. 5: 285–318.
- Erwin, D. H. 1996. "The Mother of Mass Extinctions". *Scientific American*. 275(1): 72–78.
- Erwin, D. H. 2006. *Extinction: How Life on Earth Nearly Ended 250 Million Years Ago*. Princeton: Princeton University Press.
- Estrada, F., P. Perron, and B. Martínez-López. 2013. "Statistically derived contributions of diverse human influences to twentieth-century temperature changes". *Nature Geoscience*. 6: 1050–1055.
- Farmer, J. D., C. Hepburn, M. C. Ives, T. Hale, T. Wetzer, P. Mealy, R. Rafaty, S. Srivastav, and R. Way. 2019. "Sensitive intervention points in the post-carbon transition". *Science*. 364(6436): 132–134. <https://science.sciencemag.org/content/364/6436/132.summary>.
- Farmer, J. D. and F. Lafond. 2016. "How predictable is technological progress?" *Research Policy*. 45: 647–665.
- Feinstein, C. H. 1972. *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge: Cambridge University Press.
- Fisher, F. M. 1966. *The Identification Problem in Econometrics*. New York: McGraw Hill.
- Fouquet, R. and P. J. G. Pearson. 2006. "Seven Centuries of Energy Services: The Price and Use of Light in the United Kingdom (1300–2000)". *Energy Journal*. 27: 139–178.
- Fullerton, R. L., B. G. Linster, M. McKee, and S. Slate. 2002. "Using Auctions to Reward Tournament Winners: Theory and Experimental Investigations". *RAND Journal of Economics*. 33: 62–84.
- Gamber, E. N. and J. P. Liebner. 2017. "Comment on "How biased are US government forecasts of the Federal Debt?"". *International Journal of Forecasting*. 33: 560–562.
- Geikie, A. 1863. "On the Phenomena of the Glacial Drift of Scotland". *Transactions of the Geological Society of Glasgow*. 1: 1–190.

- Gilbert, W. 1600. *De Magnete, Magnetisque Corporoibus, et de Magno Magne-
nete Tellure: Physiologia noua, Plurimis & Argumentis, & Experimentis
Demonstrata*. Translated by Mottelay, P. F. (1893) 'On the Loadstone and
Magnetic Bodies, and on That Great Magnet the Earth: A New Physiology,
Demonstrated with Many Arguments and Experiments'. New York: John
Wiley & Sons. London: Peter Short.
- Godfrey, L. G. 1978. "Testing for Higher Order Serial Correlation in Regres-
sion Equations when the Regressors include Lagged Dependent Variables".
Econometrica. 46: 1303–1313.
- Haavelmo, T. 1943. "The Statistical Implications of a System of Simultaneous
Equations". *Econometrica*. 11: 1–12.
- Hannan, E. J. and B. G. Quinn. 1979. "The Determination of the Order of an
Autoregression". *Journal of the Royal Statistical Society*. B, 41: 190–195.
- Harvey, A. C. and J. Durbin. 1986. "The effects of seat belt legislation on
British road casualties: A case study in structural time series modelling".
Journal of the Royal Statistical Society, Series B. 149: 187–227.
- Hendry, D. F. 1976. "The Structure of Simultaneous Equations Estimators".
Journal of Econometrics. 4: 51–88.
- Hendry, D. F. 1995. *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F. 1999. "An Econometric Analysis of US Food Expenditure,
1931–1989". In: *Methodology and Tacit Knowledge: Two Experiments in
Econometrics*. Ed. by J. R. Magnus and M. S. Morgan. Chichester: John
Wiley and Sons. 341–361.
- Hendry, D. F. 2001. "Modelling UK Inflation, 1875–1991". *Journal of Applied
Econometrics*. 16: 255–275.
- Hendry, D. F. 2006. "Robustifying Forecasts from Equilibrium-Correction
Models". *Journal of Econometrics*. 135: 399–426.
- Hendry, D. F. 2009. "The Methodology of Empirical Econometric Modeling:
Applied Econometrics Through the Looking-Glass". In: *Palgrave Hand-
book of Econometrics*. Ed. by T. C. Mills and K. D. Patterson. Basingstoke:
Palgrave MacMillan. 3–67.
- Hendry, D. F. 2011. "Climate Change: Possible Lessons for our Future from
the Distant Past". In: *The Political Economy of the Environment*. Ed. by
S. Dietz, J. Michie, and C. Oughton. London: Routledge. 19–43.

- Hendry, D. F. 2015. *Introductory Macro-econometrics: A New Approach*. <http://www.timberlake.co.uk/macroeconometrics.html>. London: Timberlake Consultants.
- Hendry, D. F. 2018. "Deciding between Alternative Approaches in Macroeconomics". *International Journal of Forecasting*. 34: 119–135, with 'Response to the Discussants', 142–146.
- Hendry, D. F. and J. A. Doornik. 2014. *Empirical Model Discovery and Theory Evaluation*. Cambridge, Mass.: MIT Press.
- Hendry, D. F. and N. R. Ericsson. 1991. "An Econometric Analysis of UK Money Demand in 'Monetary Trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz". *American Economic Review*. 81: 8–38.
- Hendry, D. F. and S. Johansen. 2015. "Model Discovery and Trygve Haavelmo's Legacy". *Econometric Theory*. 31: 93–114.
- Hendry, D., S. Johansen, and C. Santos. 2008. "Automatic selection of indicators in a fully saturated regression." *Computational Statistics & Data Analysis*. 33: 317–335.
- Hendry, D. F. and K. Juselius. 2000. "Explaining Cointegration Analysis: Part I". *Energy Journal*. 21: 1–42.
- Hendry, D. F. and K. Juselius. 2001. "Explaining Cointegration Analysis: Part II". *Energy Journal*. 22: 75–120.
- Hendry, D. F. and H.-M. Krolzig. 2005. "The Properties of Automatic Gets Modelling". *Economic Journal*. 115: C32–C61.
- Hendry, D. F., M. Lu, and G. E. Mizon. 2009. "Model Identification and Non-unique Structure". In: *The Methodology and Practice of Econometrics*. Ed. by J. L. Castle and N. Shephard. Oxford: Oxford University Press. 343–364.
- Hendry, D. F. and G. E. Mizon. 1993. "Evaluating Dynamic Econometric Models by Encompassing the VAR". In: *Models, Methods and Applications of Econometrics*. Ed. by P. C. B. Phillips. Oxford: Basil Blackwell. 272–300.
- Hendry, D. F. and G. E. Mizon. 2011. "Econometric Modelling of Time Series with Outlying Observations". *Journal of Time Series Econometrics*. 3 (1): DOI: 10.2202/1941–1928.1100.
- Hendry, D. F., A. J. Neale, and F. Srba. 1988. "Econometric Analysis of Small Linear Systems using PC-FIML". *Journal of Econometrics*. 38: 203–226.

- Hendry, D. F. and F. Pretis. 2013. "Anthropogenic Influences on Atmospheric CO₂". In: *Handbook on Energy and Climate Change*. Ed. by R. Fouquet. Cheltenham: Edward Elgar. 287–326.
- Hendry, D. F. and F. Pretis. 2016. *All Change! The Implications of Non-stationarity for Empirical Modelling, Forecasting and Policy*. Oxford University: Oxford Martin School Policy Paper.
- Hendry, D. F. and C. Santos. 2005. "Regression Models with Data-based Indicator Variables". *Oxford Bulletin of Economics and Statistics*. 67: 571–595.
- Hendry, D. F. and C. Santos. 2010. "An Automatic Test of Super Exogeneity". In: *Volatility and Time Series Econometrics*. Ed. by M. W. Watson, T. Bollerslev, and J. Russell. Oxford: Oxford University Press. 164–193.
- Hettmansperger, T. P. and S. J. Sheather. 1992. "A Cautionary Note on the Method of Least Median Squares". *The American Statistician*, 46: 79–83.
- Heydari, E., N. Arzani, and J. Hassanzadeh. 2008. "Mantle plume: The invisible serial killer—Application to the Permian-Triassic boundary mass extinction". *Palaeogeography, Palaeoclimatology, Palaeoecology*. 264: 147–162.
- Hoffman, P. F. and D. P. Schrag. 2000. "Snowball Earth". *Scientific American*. 282: 68–75.
- Hsiang, S. 2016. "Climate econometrics". *Annual Review of Resource Economics*. 8(1): 43–75.
- Imbrie, J. E. 1992. "On the structure and origin of major glaciation cycles, 1, Linear responses to Milankovitch forcing". *Paleoceanography*. 7: 701–738.
- Jaccard, S. L., E. D. Galbraith, A. Martínez-García, and R. F. Anderson. 2016. "Covariation of deep Southern Ocean oxygenation and atmospheric CO₂ through the last ice age". *Nature*. 530: 207–210. doi:10.1038/nature16514.
- Jackson, L. P. and D. F. Hendry. 2018. "Risk and Exposure of Coastal Cities to Future Sea-level Rise". *Working Paper*. Oxford University: INET Oxford.
- Johansen, S. 1995. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Johansen, S. and B. Nielsen. 2009. "An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator". In: *The Methodology and Practice of Econometrics*. Ed. by J. L. Castle and N. Shephard. Oxford: Oxford University Press. 1–36.

- Johansen, S. and B. Nielsen. 2016. "Asymptotic theory of outlier detection algorithms for linear time series regression models". *Scandinavian Journal of Statistics*. 43: 321–348.
- Jones, C. and P. Cox. 2005. "On the significance of atmospheric CO₂ growth rate anomalies in 2002–2003". *Journal of Geophysical Research*. 32.
- Jouzel, J., V. Masson-Delmotte, O. Cattani, G. Dreyfus, S. Falourd, and G. E. Hoffmann. 2007. "Orbital and millennial Antarctic climate variability over the past 800,000 years". *Science*. 317: 793–797.
- Kaufmann, R. K. and K. Juselius. 2010. "Glacial cycles and solar insolation: The role of orbital, seasonal, and spatial variations". *Climate of the Past Discussions*. 6: 2557–2591. doi:10.5194/cpd-6-2557-2010.
- Kaufmann, R. and K. Juselius. 2013. "Testing hypotheses about glacial cycles against the observational record". *Paleoceanography*. 28: 175–184.
- Kaufmann, R. K., H. Kauppi, M. L. Mann, and J. H. Stock. 2011. "Reconciling anthropogenic climate change with observed temperature 1998–2008". *Proceedings of the National Academy of Science*. 108: 11790–11793.
- Kaufmann, R. K., H. Kauppi, M. L. Mann, and J. H. Stock. 2013. "Does temperature contain a stochastic trend: linking statistical results to physical mechanisms". *Climatic Change*. 118: 729–743.
- Kaufmann, R. K., M. L. Mann, S. Gopala, J. A. Liederman, P. D. Howe, F. Pretis, X. Tanga, and M. Gilmore. 2017. "Spatial heterogeneity of climate change as an experiential basis for skepticism". *Proceedings of the National Academy of Sciences*. 114(1): 67–71.
- Keeling, C. D., R. B. Bacastow, A. E. Brainbridge, C. A. Ekdahl, P. R. Guenther, L. S. Waterman, and J. F. S. Chin. 1976. "Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii". *Tellus*. 6: 538–551.
- Kitov, O. I. and M. N. Tabor. 2015. "Detecting structural changes in linear models: A variable selection approach using multiplicative indicator saturation". *Unpublished Paper*. University of Oxford.
- Knutti, R., M. A. A. Rugenstein, and G. C. Hegerl. 2017. "Beyond equilibrium climate sensitivity". *Nature Geoscience*. 10: 727–736. <https://doi.org/10.1038/ngeo3017>.
- Koopmans, T. C. 1949. "Identification Problems in Economic Model Construction". *Econometrica*. 17: 125–144.

- Koopmans, T. C., ed. 1950a. *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph. No. 10. New York: John Wiley & Sons.
- Koopmans, T. C. 1950b. "When is an equation system complete for statistical purposes?" In: *Statistical Inference in Dynamic Economic Models*. Ed. by T. C. Koopmans. Cowles Commission Monograph. No. 10. New York: John Wiley & Sons. Chap. 17.
- Koopmans, T. C. and O. Reiersøl. 1950. "The Identification of Structural Characteristics". *The Annals of Mathematical Statistics*. 21: 165–181.
- Kurle, J. K. 2019. "Essays in Climate Econometrics". *Unpublished M. Phil. Thesis*. University of Oxford: Economics Department.
- Lamb, H. H. 1959. "Our changing climate, past and present". *Weather*. 14: 299–317.
- Lamb, H. H. 1995. *Climate, History and the Modern World*. Second edition (First ed., 1982). London: Routledge.
- Lea, D. W. 2004. "The 100,000-yr cycle in tropical SST, greenhouse forcing and climate sensitivity". *Journal of Climate*. 17: 2170–2179.
- Lisiecki, L. E. and M. E. Raymo. 2005. "A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records". *Paleoceanography*. 20: doi:10.1029/2004PA001071.
- Lüthil, D., M. Le Floch, B. Bereiter, T. Blunier, U. Barnola J-M. and Siegenthaler, D. Raynaud, J. Jouzel, H. Fischer, K. Kawamura, and T. F. Stocker. 2008. "High-resolution Carbon Dioxide Concentration Record 650,000–800,000 years before Present". *Nature*. 453. doi:10.1038/nature06949.
- Marland, G., T. Boden, and R. Andres. 2011. "Global, Regional, and National Fossil Fuel CO₂ Emissions". In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. <http://cdiac.ornl.gov/trends/emis/overview.html>.
- Marland, G. and R. Rotty. 1984. "Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950-1982". *Tellus*. 36B: 232–261.
- Martinez-Garcia, A., A. Rosell-Melé, W. Geibert, R. Gersonde, P. Masqué, and V. E. Gaspari. 2009. "Links between iron supply, marine productivity, sea surface temperature, and CO₂ over the last 1.1 Ma". *Paleoceanography*. 24: doi:10.1029/2008PA001657.

- Martinez, A. B. 2020. "Forecast Accuracy Matters for Hurricane Damages". *Econometrics*: <https://www.mdpi.com/2225-1146/8/2/18>.
- Martinez, A. B., J. L. Castle, and D. F. Hendry. 2019. "Smooth Robust Multi-Step Forecasting Methods". *Unpublished paper*. Oxford University: Nuffield College.
- Masson-Delmotte, V., et al. 2006. "Past and future polar amplification of climate change: Climate model intercomparisons and ice-core constraints". *Climate Dynamics*. 26: 513–529. <https://doi.org/10.1007/s00382-005-0081-9>.
- Masson-Delmotte, V., et al. 2010. "EPICA Dome C record of glacial and interglacial intensities". *Quaternary Science Reviews*. 29: 113–128. <https://doi.org/10.1016/j.quascirev.2009.09.030>.
- Mayhew, P. J., G. B. Jenkins, and T. G. Benton. 2009. "A long-term association between global temperature and biodiversity, origination and extinction in the fossil record". *Proceedings of the Royal Society, B*. 275:1630: 47–53.
- Mee, L. 2006. "Reviving Dead Zones: How can we restore coastal seas ravaged by runaway plant and algae growth caused by human activities?" *Scientific American*. 295: 78–85.
- Meinshausen, M., N. Meinshausen, W. Hare, S. C. Raper, K. Frieler, R. Knutti, D. J. Frame, and M. R. Allen. 2009. "Greenhouse-gas emission targets for limiting global warming to 2°C". *Nature*. 458: 1158–1162.
- Milankovitch, M. 1969. *Canon of insolation and the ice-age problem*. English translation by the Israel Program for Scientific Translations of *Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem*, Textbook Publishing Company, Belgrade, 1941. Washington, D.C: National Science Foundation.
- Mitchell, B. R. 1988. *British Historical Statistics*. Cambridge: Cambridge University Press.
- Mizon, G. and J. Richard. 1986. "The encompassing principle and its application to non-nested hypothesis tests". *Econometrica*. 54: 657–678.
- Mizon, G. E. 1995. "A simple message for autocorrelation correctors: Don't". *Journal of Econometrics*. 69: 267–288.
- Myhre, G., A. Myhre, and F. Stordal. 2001. "Historical evolution of radiative forcing of climate". *Atmospheric Environment*. 35: 2361–2373.

- Nevison, C., N. Mahowald, S. Doney, I. Lima, G. van der Werf, J. Randerson, D. Baker, P. Kasibhatla, and G. McKinley. 2008. "Contribution of ocean, fossil fuel, land biosphere, and biomass burning carbon fluxes to seasonal and interannual variability in atmospheric CO₂". *Journal of Geophysical Research*. 113.
- Newey, W. K. and K. D. West. 1987. "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation-Consistent Covariance Matrix". *Econometrica*. 55: 703–708.
- Nielsen, B. 2006. "Order determination in general vector autoregressions". In: *Time series and related topics: In memory of Ching-Zong Wei*. Ed. by H.-C. Ho, C.-K. Ing, and T. L. Lai. Vol. 52. *Lecture Notes–Monograph Series*. Beachwood, OH: Institute of Mathematical Statistics. 93–112.
- Nielsen, B. and M. Qian. 2018. "Asymptotic properties of the gauge of step-indicator saturation". *Discussion Paper*. University of Oxford: Economics Department.
- Nielsen, B. and A. Rahbek. 2000. "Similarity Issues in Cointegration Analysis". *Oxford Bulletin of Economics and Statistics*. 62: 5–22.
- Orcutt, G. H. and D. Cochrane. 1949. "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis". *Journal of the American Statistical Association*. 44: 356–372.
- Paillard, D. 2001. "Glacial cycles: towards a new paradigm". *Reviews of Geophysics*. 39: 325–346.
- Paillard, D. 2010. "Climate and the orbital parameters of the Earth". *Compte Rendus Geoscience*. 342: 273–285.
- Paillard, D., L. D. Labeyrie, and P. Yiou. 1996. "Macintosh program performs time-series analysis". *Eos Transactions AGU*. 77: 379.
- Parrenin, F., J.-R. Barnola, J. Beer, T. Blunier, and E. E. Castellano. 2007. "The EDC3 chronology for the EPICA Dome C ice core". *Climate of the Past*. 3: 485–497.
- Parrenin, F., J.-R. Petit, V. Masson-Delmotte, and E. E. Wolff. 2012. "Volcanic synchronisation between the EPICA Dome C and Vostok ice cores (Antarctica) 0–145 kyr BP". *Climate of the Past*. 8: 1031–1045.
- Pfeiffer, A., R. Millar, C. Hepburn, and E. Beinhocker. 2016. "The '2°C capital stock' for electricity generation: Committed cumulative carbon emissions from the electricity generation sector and the transition to a green economy". *Applied Energy*. 179: 1395–1408.

- Pistone, K., I. Eisenman, and V. Ramanathan. 2019. "Radiative heating of an ice-free Arctic Ocean". *Geophysical Research Letters*. 46: 7474–7480.
- Pretis, F. 2017. "Exogeneity in Climate Econometrics". *Working paper*. Oxford University: Economics Department.
- Pretis, F. 2019. "Econometric Models of Climate Systems: The Equivalence of Two-Component Energy Balance Models and Cointegrated VARs". *Journal of Econometrics*: <https://doi.org/10.1016/j.jeconom.2019.05.013>.
- Pretis, F. and D. F. Hendry. 2013. "Comment on "Polynomial cointegration tests of anthropogenic impact on global warming" by Beenstock et al. (2012) – some hazards in econometric modelling of climate change". *Earth System Dynamics*. 4: 375–384.
- Pretis, F. and R. K. Kaufmann. 2018. "Out-of-Sample Paleo-Climate Simulations: Testing Hypotheses About the Mid-Brunhes Event, the Stage 11 Paradox, and Orbital Variations". *Discussion Paper*. Canada: University of Victoria.
- Pretis, F. and R. K. Kaufmann. 2020. "Managing Carbon Emissions to Avoid the Next Ice Age". *Discussion Paper*. Canada: University of Victoria.
- Pretis, F., M. L. Mann, and R. K. Kaufmann. 2015a. "Testing Competing Models of the Temperature Hiatus: Assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection". *Climatic Change*. 131: 705–718.
- Pretis, F., J. J. Reade, and G. Sucarrat. 2018a. "Automated General-to-Specific (GETS) Regression Modeling and Indicator Saturation for Outliers and Structural Breaks". *Journal of Statistical Software*. 68, 4: <https://www.jstatsoft.org/article/view/v086i03>.
- Pretis, F. and M. Roser. 2017a. "Carbon Dioxide Emission-Intensity in Climate Projections: Comparing the Observational Record to Socio-Economic Scenarios". *Energy*. 135: 718–725. <https://doi.org/10.1016/j.energy.2017.06.119>.
- Pretis, F., L. Schneider, J. E. Smerdon, and D. F. Hendry. 2016. "Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation". *Journal of Economic Surveys*. 30: 403–429.
- Pretis, F., M. Schwarz, K. Tang, K. Haustein, and M. R. Allen. 2018b. "Uncertain impacts on economic growth when stabilizing global temperatures at 1.5°C or 2°C warming". *Philosophical Transactions of the Royal Society*. A376: 20160460. <https://doi.org/10.1098/rsta.2016.0460>.

- Prothero, D. R. 2008. "Do Impacts Really Cause Most Mass Extinctions?" In: *From Fossils to Astrobiology*. Ed. by J. Seckbach and M. Walsh. Netherlands: Springer. 409–423.
- Rampino, M. and S.-Z. Shen. 2019. "The end-Guadalupian (259.8 Ma) biodiversity crisis: the sixth major mass extinction?" *Historical Biology*. <https://doi.org/10.1080/08912963.2019.1658096>.
- Ramsey, J. B. 1969. "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis". *Journal of the Royal Statistical Society. B*, 31: 350–371.
- Randerson, T., M. Thompson, T. Conway, I. Fung, and C. Field. 1997. "The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide". *Global Biogeochemical Cycles*. 11:4: 535–560.
- Ravishankara, A. R., J. S. Daniel, and R. W. Portmann. 2009. "Nitrous Oxide (N₂O): The Dominant Ozone-Depleting Substance Emitted in the 21st Century". *Science*. 326: 123–125.
- Riccardi, A., L. R. Kump, M. A. Arthur, and S. D'Hondt. 2007. "Carbon isotopic evidence for chemocline upward excursion during the end-Permian event". *Palaeogeography, Palaeoclimatology, Palaeoecology*. 248: 263–291.
- Richard, J.-F. 1980. "Models with Several Regimes and Changes in Exogeneity". *Review of Economic Studies*. 47: 1–20.
- Rigby, M. E. 2010. "History of atmospheric SF₆ from 1973 to 2008". *Atmospheric Physics and Chemistry*. 10: 10305–10320. doi:10.5194/acp-10-10305-2010.
- Rothenberg, T. J. 1971. "Identification in Parametric Models". *Econometrica*. 39: 577–592.
- Rothenberg, T. J. 1973. *Efficient Estimation with A Priori Information*. Cowles Foundation Monograph. No. 23. New Haven: Yale University Press.
- Rousseeuw, P. J. 1984. "Least median of squares regression". *Journal of the American Statistical Association*. 79: 871–880.
- Rowan, S. S. 2019. "Pitfalls in Comparing Paris Pledges". *Climatic Change*: <https://link.springer.com/article/10.1007/s10584-019-02494-bibbrangedash7>.
- Ruddiman, W., ed. 2005. *Plows, Plagues and Petroleum: How Humans took Control of Climate*. Princeton: Princeton University Press.

- Salkever, D. S. 1976. "The Use of Dummy Variables to Compute Predictions, Prediction Errors and Confidence Intervals". *Journal of Econometrics*. 4: 393–397.
- Sargan, J. D. 1964. "Wages and Prices in the United Kingdom: A Study in Econometric Methodology (with discussion)". In: *Econometric Analysis for National Economic Planning*. Ed. by P. E. Hart, G. Mills, and J. K. Whitaker. Vol. 16. *Colston Papers*. London: Butterworth Co. 25–63.
- Schneider, L., J. E. Smerdon, F. Pretis, C. Hartl-Meier, and J. Esper. 2017. "A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures". *Environmental Research Letters*. 12, 9: <http://iopscience.iop.org/article/10.1088/1748\bibrangedash9326/aa7a1b/meta>.
- Schwarz, G. 1978. "Estimating the Dimension of a Model". *Annals of Statistics*. 6: 461–464.
- Siddall, M., E. J. Rohling, A. Almogi-Labin, C. Hemleben, D. Meischner, and I. E. Schmelzer. 2003. "Sea-level fluctuations during the last glacial cycle". *Nature*. 423: 853–858.
- Snir, A., D. Nadel, I. Groman-Yaroslavski, Y. Melamed, M. Sternberg, and O. E. Bar-Yosef. 2015. "The Origin of Cultivation and Proto-Weeds, Long Before Neolithic Farming". *PLoS ONE*. 10(7): e0131422. <https://doi.org/10.1371/journal.pone.0131422>.
- Spanos, A. and J. J. Reade. 2015. "Heteroskedasticity/Autocorrelation Consistent Standard Errors and the Reliability of Inference". *Unpublished paper*. USA: Virginia Tech.
- Stein, K., A. Timmermann, E. Y. Kwon, and T. Friedrich. 2003. "Timing and magnitude of Southern Ocean sea ice/carbon cycle feedbacks". *PNAS*. 117 (9): 4498–4504. <https://www.pnas.org/content/117/9/4498>.
- Stern, D. I. 2004. "The Rise and Fall of the Environmental Kuznets Curve". *World Development*. 32: 1419–1439.
- Stern, N. 2006. *The Economics of Climate Change: The Stern Review*. Cambridge: Cambridge University Press.
- Stone, R. 2007. "A World Without Corals?" *Science*. 316: 678–681.

- Suess, H. E. 1953. "Natural Radiocarbon and the Rate of Exchange of Carbon Dioxide between the Atmosphere and the Sea". In: *Nuclear Processes in Geologic Settings*. Ed. by National Research Council Committee on Nuclear Science. Washington, D. C.: National Academy of Sciences. 52–56.
- Sundquist, E. T. and R. F. Keeling. 2009. "The Mauna Loa Carbon Dioxide Record: Lessons for Long-Term Earth Observations". *Geophysical Monograph Series*. 183: 27–35.
- Thomson, K. S. 1991. *Living Fossil: The Story of the Coelacanth*. London: Hutchinson Radius.
- Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society. B*, 58: 267–288.
- U.S. Energy Information Administration. 2009. "Emissions of Greenhouse Gases in the U.S." *Report DOE/EIA-0573(2008)*. https://www.eia.gov/environment/emissions/ghg_report/.
- Vaks, A., A. Mason, S. Breitenbach, A. Kononov, A. Osinzev, M. P. Rosensaft, A. Borshevsky, O. Gutareva, and G. Henderson. 2020. "Palaeoclimate evidence of vulnerable permafrost during times of low sea ice". *Nature*. 577: 221–225. <https://www.nature.com/articles/s41586-019-1880-1>.
- Víšek, J. A. 1999. "The Least Trimmed Squares – Random Carriers". *Bulletin of the Czech Econometric Society*. 6: 1–30.
- Vousdoukas, M. I., L. Mentaschi, E. Voukouvalas, M. Verlaan, S. Jevrejeva, L. P. Jackson, and L. Feyen. 2018. "Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard". *Nature Communications*. 9: 2360.
- Walker, A., F. Pretis, A. Powell-Smith, and B. Goldacre. 2019. "Variation in Responsiveness to Warranted Behaviour Change Among NHS Clinicians: A Novel Implementation of Change-Detection Methods in Longitudinal Prescribing Data". *British Medical Journal*. 367: I5205. <https://doi.org/10.1136/bmj.I5205>.
- Ward, P. D. 2006. "Impact from the Deep". *Scientific American*. 295: 64–71.
- Weart, S. 2010. "The Discovery of Global Warming". <http://www.aip.org/history/climate/co2.htm>.
- White, H. 1980. "A Heteroskedastic-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica*. 48: 817–838.

- Winchester, S. 2001. *The Map that Changed the World*. London: Harper Collins.
- Worrell, E., L. Price, N. Martin, C. Hendriks, and L. Meida. 2001. “Carbon Dioxide Emissions from the Global Cement Industry”. *Annual Review of Energy and the Environment*. 26: 303–329.
- Zanna, L., S. Khatiwala, J. M. Gregory, J. Ison, and P. Heimbach. 2019. “Global reconstruction of historical ocean heat storage and transport”. *PNAS*: <https://doi.org/10.1073/pnas.1808838115>.