Production networks and planetary boundaries: challenges and opportunities for Integrated Assessment Models

Andrea Bacilieri

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2023

St Anthony's College
School of Geography and the Environment
University of Oxford
Institute for New Economic Thinking at the Oxford Martin School
Smith School of Enterprise and the Environment

If a factory is torn down but the rationality which produced it is left standing, then that rationality will simply produce another factory. If a revolution destroys a government, but the systematic patterns of thought that produced that government are left intact, then those patterns will repeat themselves in the succeeding government. There's so much talk about the system. And so little understanding.

—Robert Pirsig, *Zen and the Art of Motorcycle Maintenance*

This thesis is dedicated to Mother Earth and to all the living beings on this beautiful planet.

Supervisors

Prof. Cameron Hepburn
Prof. Doyne Farmer
Dr. François Lafond

Examiners

Prof. Samuel Fankhauser (University of Oxford)
Prof. Alexandra Brintrup (University of Cambridge)

Acknowledgements

I am deeply grateful to Cameron Hepburn, Doyne Farmer and François Lafond. It has been an honour to learn from you and work with you. I am so thankful for the intellectual freedom you have given me – it has led me down a truly unexpected path. François, a special thank you to you; this DPhil would never have been possible without you. Thank you for being part of my (bumpy) academic journey and for guiding me along the way.

I would also like to express my sincere gratitude to Sam Fankhauser and Alexandra Brintrup for agreeing to serve as committee members. Your valuable insights have undeniably enhanced the quality of this thesis.

My intellectual journey, some parts of which are recounted in this thesis, has been influenced by many great minds and people. You have all made my DPhil an invaluable experience and I will be forever grateful to have crossed paths with all of you.

To my co-authors, François Lafond, András Borsos and Pablo Astudillo-Estévez; it has been an absolute pleasure to engage in stimulating conversions, make new discoveries and solve problems together.

To the whole group at INET Oxford and the people I encountered at the many conferences and summer schools, thank you for the breadth of perspectives, insights, advice, foosball games, concerts, late nights (either in the office or out and about), adventures, hikes and much, much more. I will hold these memories very dear. Thank you to Giancarlo Antoniucci, Nils Rochowicz, Fabian Dablander, Penny Mealy, Maria del Rio-Chanona, Anton Pichler, Jangho Yang, Marcon Pangallo, Ruper Way, Luca Mungo, José Moran, Deepanshu Singh, Torsten Heinrich, Julian Winkler, Blas Kolic, Matthew Ives, Peter Barbrook-Johnson, Vilhelm Verendel, Maarten Scholl, Joris Bücker, Juan Sabuco, Linus Mattauch, Stefania Innocenti, Kerstin Hötte, François Cohen, Valentina Semenova, Adrián Carro, Donovan Platt, Rupert Stuart-Smith, Aabir Abubaker Kar, Anshuman Swain, Bhargav Srinivasa Desikan, Brennan Klein, Alexander Bakus, Christian Diem, Tobias Reisch, Daniel Borrero, Ruggiero Lo Sardo, Douglas Guilbeault, Ernest Aigner, Elissa Cohen, Hunter Wapman, Mark Chu, Paula Parpart, Leto Peel, Oriol Artime and Riccardo Gallotti.

To the whole group of weird and absolutely wonderful people at the 2019 Santa Fe summer school. Our month in isolation in the middle of New Mexico's desert landscape has been one of the most amazing experiences of my life. A huge thank you also to the Institute of American Indian Arts for hosting us on their gorgeous campus.

Special thanks to Dorothy Nicholas and Susan Mousley for all your help and care over the years at INET.

I am also extremely grateful to have met all the people at the colive in Portugal while I was preparing for my VIVA. Thank you for our daily meditations, yoga classes, ecstatic dances and your radiant energy. Thomas, thank you for being such an amazing teacher and human being!

To my dearest friends, Gemma Callovini, Margherita Birri and Sumit Kothari, three beautiful souls I have had the great fortune to meet. You have been constant and invaluable companions on this journey. Marghez! Sinistra, destra... Sinistra, destra.

Expressing my gratitude to my family, especially my aunt Carla, is a task that words cannot truly capture. Your unwavering support from the very beginning of this challenging journey means more to me than I can ever convey. I couldn't have got here without you. To my sister Mila, whose support over the last few months has been a pillar of vitality and perseverance.

Abstract

Integrated Assessment Models (IAMs) are used to understand the complex interactions between the Earth system and socio-economic processes. These models help us understand possible futures associated with different levels of human impact on the climate system. As such, they wield significant influence on policy-making and society as a whole. At the heart of this thesis is a fundamental inquiry into advancing IAMs. This deep inquiry is rooted in the understanding that the economy is a complex system and in the acknowledgement of the intertwined nature of the climate and ecological crises; both are mostly overlooked in IAMs. This thesis provides insights and advances for IAMs with the goal of strengthening their ability to address the climate and ecological crises together.

First, this thesis finds that IAMs are not up to the task of addressing the climate and ecological crises together. Using the Planetary Boundaries (PBs) framework, it assesses the ecological feasibility of Pari-compliant mitigation pathways that were considered in the recent IPCC Sixth Assessment Report. Almost all scenarios transgress PBs. Even "low-demand" or "sustainability" pathways do not meet the ecological feasibility criteria set forth by the PBs framework. These findings highlight the need for a comprehensive and integrated approach.

Second, drawing from complexity economics and recent findings in the literature on supply chain networks, this thesis argues that IAMs overlook key micro-level mechanisms that are essential for understanding the evolution, stability and resilience of the economic system – and thus of society as a whole. It explores how we might have a more fine-grained macroeconomic model that takes into account supply chain interactions at the firm level. It finds that serious data limitations must be overcome before we can achieve this level of granularity in IAMs. Using methods from complexity science, it addresses the data limitations using two different approaches and makes two major contributions: (1) it provides the first comprehensive picture of the most fundamental statistics on production networks, thus providing a basis for generating synthetic data or calibrating macroeconomic models; and (2) it provides the first thorough evaluation of a maximum entropy reconstruction method applied to firm-level production networks. This thesis also contributes to an emerging agenda to develop standards for data collection, cleaning and matching for micro-level production network data around the world.

DPhil papers

1 **Evaluating the efficacy of Paris-compliant mitigation pathways in tackling the ecological crisis**
Andrea Bacilieri.
Submitted to *Nature Climate Change*.

2 **Firm-level production networks: what do we (really) know?**
Andrea Bacilieri, András Borsos, Pablo Astudillo-Estevez and François Lafond.
Submitted to the *Journal of Political Economy*.

3 **Reconstructing firm-level input-output networks from partial information**
Andrea Bacilieri and Pablo Astudillo-Estevez.
Submitted to the *Journal of Economic Dynamics and Control*.

Contents

Chapter 1

Introduction

This thesis, and my research inquiries, are motivated by two observations. First, we are in a climate and ecological emergency (IPCC, 2021, 2022a,b; IPBES, 2019; UNEP, 2019). And second, the socio-economic system is a complex system (Arthur, 2021; Meadows, 2008; Farmer, draft), but this is rarely taken into account in the models currently used to understand these crises and assess possible solutions (Farmer et al., 2015; Axtell, 2014; Mercure et al., 2016).

The challenges to avert climate and ecological breakdown are enormous, with UN Secretary-General António Guterres declaring that it is "code red for humanity" (Guterres, 2022b) following the release of the Working Group I (WGI) contribution to the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) on the physical science basis (IPCC, 2021). Both the recent WGI report and the Special Report on 1.5°C (IPCC, 2021, 2018) make clear that the differences between 1.5°C and 2°C of warming are non-negligible, with many interacting tipping elements very likely to be crossed at 1.5°C and some likely to have already been crossed (McKay et al., 2021; Wunderling et al., 2021). At the same time, natural ecosystems are being degraded at an unprecedented rate (IPBES, 2019; UNEP, 2019) and, in addition to climate change, we have crossed several other planetary boundaries, leaving the safe operating space for humanity (Steffen et al., 2015b; Rockström et al., 2023; Persson et al., 2022; Fanning et al., 2021).

The transformation that our society must undergo over the next few decades is massive and daunting. Many possible futures lie ahead, some considerably more desirable than others. Scientists and policy-makers have typically approached the problem of envisioning possible futures and assessing policy options to avoid climate breakdown by developing *what-if* scenarios using computer models that simulate Earth system and societal processes as well as their interactions. These models are called *Integrated Assessment Models* (IAMs). As useful as IAMs have been, they have also received criticism (Farmer et al., 2015; Anderson and Jewell, 2019; Mercure et al., 2016; Pauliuk et al., 2017). At the heart of this thesis is the question of how we can advance IAMs so that they are better suited to address the challenges posed by the climate and ecological crises.

This thesis joins the crowd of people calling for a new wave of IAMs that builds on complexity economics (e.g., Farmer, draft; Beinhocker et al., 2013; Mercure et al., 2016; Hafner et al., 2020);[1] complexity economics draws on complex systems science (see Miller and Page, 2007; Arthur, 2014; Thurner et al., 2018, for an introduction to complexity science and economics). This thesis also argues that IAMs are being developed with limited systems thinking, hindering our ability to address the climate and ecological crises together. Systems thinking and complexity theory are related. Indeed, systems thinking recognises that most systems, such as our bodies, trees, ecosystems and societies, are complex systems (Meadows, 2008).

Viewing socio-economic systems as *complex systems* would entail recognising the strong interdependencies between the components of the system, often involving non-linear relationships, feedbacks, learning and adaptation. Complex systems are path-dependent systems where macro-level outcomes emerge from the interaction of agents at the micro-level. In turn, the agents react to the aggregate outcomes and, as a result, the system may be in constant motion, never stabilising at an equilibrium, or it may reach one or multiple equilibria (Arthur, 2021; Thurner et al., 2018). In most IAMs, however, models of the economic system are based on neoclassical economic theory, which applies a reductionist approach to what is an inherently complex system. The reductionist scientific approach breaks down in complex systems (Miller and Page, 2007) because "the whole becomes not only more than but very different from the sum of its parts" (Anderson, 1972, pg. 395).

Adopting a systems thinking approach would further lead to a recognition of the interconnectedness of the climate and ecological crises, and an assessment of the synergies and trade-offs between them, which in turn would lead to very different solutions and policy prescriptions. Addressing these two crises with a systems thinking and complexity science lens would also require a departure from optimisation approaches (used in most IAMs) in favour of agent-based or system dynamics models, which are better able to capture characteristics of real socio-economic systems (Hafner et al., 2020).

Within this context, this thesis takes two different perspectives. The first perspective takes a holistic view and focuses on the interactions between human societies and the Earth System. The second perspective takes a granular view and focuses on the interactions between firms in the production system. With regard to the first perspective, this thesis shows that, on the one hand, some IAMs have what could be considered a fairly comprehensive and integrated approach, but, on the other hand, they do not exploit it for addressing the environmental impacts of mitigation scenarios. Most importantly, it shows that the limits of the proposed solutions imposed by the carrying capacity of the planet are not adequately addressed. Exceeding these limits could push the Earth system into new states with potentially disastrous consequences (Steffen et al., 2015b; Rockström et al., 2009).

[1] The call is actually much broader, encompassing all of economics.

With regard to the second perspective, this thesis draws on recent findings in the literature on production networks that highlight the key role played by supply chain networks in shaping macroeconomic outcomes and it thus argues that the correct level for modelling the economic system is the firm level. Most IAMs simply ignore interactions, while those that do not ignore interactions analyse them on a level that is too aggregated (i.e., the sectoral level). Modelling the production system at the sector level is problematic because it washes away firm heterogeneity, thereby underestimating the impacts of shocks (Diem et al., 2023; Inoue and Todo, 2019). It also fails to capture the environmental impacts that occur throughout the supply chain, minimizing current impacts and those of the proposed solutions (Hertwich et al., 2015; Heck et al., 2018a; Usubiaga et al., 2017) and it prevents us from identifying leverage points and bottlenecks in the system (Inoue and Todo, 2019; Henriet et al., 2012; Stangl et al., 2023).

1.1 Problem statement and research aims

This section explores in more detail the gaps that this thesis addresses by relating them to the two perspectives that this thesis takes on IAMs that were discussed before.

The first perspective notes that IAMs do not address the ecological consequences of climate mitigation scenarios (although some of them potentially could). While some have emphasised that mitigation pathways are not intended to serve as prescriptive guidelines or blueprints for policy-makers, it is important to recognise their influence on policy decisions and on society as a whole. Given their influence, it is paramount that climate mitigation scenarios do not exacerbate other ecological issues.

To date, only a few studies have addressed ecological issues, albeit to different degrees. Few studies have developed or (less than a handful) assessed scenarios using techniques and models of varying levels of sophistication and in relation to one or more ecological concerns (e.g., Heck et al., 2018b; Weidner and Guillén-Gosálbez, 2023; Conijn et al., 2018; Hertwich et al., 2015). Only one study has developed national pathways that stay within the carrying capacity of the planet (Allen et al., 2021).

To fill this gap in the literature, one could either develop scenarios that respect ecological constraints or assess the ecological feasibility of the scenario afterwards, once it has been developed. This thesis chooses the latter option. It focuses on assessing the ecological feasibility of existing scenarios that were considered in the recent IPCC AR6. These scenarios have been selected for their high relevance to policy-makers and other societal actors. The following reach question is addressed in *Paper 1*:

1. Do climate mitigation scenarios that stay within 1.5°C or 2°C of warming address climate change at the expense of ecological issues?

The second perspective this thesis takes on IAMs argues that their macroeconomic models fail to recognise that the economic system is a complex system and lack granularity.

This thesis calls for the need to develop a macroeconomic model that draws on complexity economics, captures supply chain interactions, and is empirically grounded and granular (i.e., at the firm level). However, there are serious data limitations with respect to firm-level supply chain networks that severely hamper our ability to develop such a model. Data are scarce and the few datasets that do exist are riddled with missing data. Missing data at the network level relate to firms, supply chain relationships and the value of transactions. Missing data at the firm level mainly concerns information on CO_2 emissions and other environmental impacts such as natural resource use and chemical pollutants released into the environment.

There are two approaches to dealing with these problems. One approach is to live with the fact that one may not have access to these data, but still wants to create synthetic data or calibrate a macroeconomic model. The second approach is to use the handful of supply chain datasets that are available at the global level (one would want a model covering the global economy) and reconstruct the missing data. This thesis explores both of these potential solutions in the context of the supply chain network, leaving for future work the question of how to reconstruct missing data on environmental impacts.

The first approach leads to the following research questions, which are addressed in *Paper 2*:

2. What data are available on firm-level production networks?

3. What do we know about the structure of firm-level production networks?

4. Are there generic properties of firm-level production networks that hold across different countries and over time?

5. Do differences between datasets arise from the methods used to collect and clean the data?

The second approach leads to the following research questions, which are addressed in *Paper 3*:

6. Can we reconstruct the monetary value of the transactions given that we have partial knowledge of the binary topology and information on firms' revenues and expenditures from their financial statements?

7. Can we capture the rest of the economic activity not covered in the dataset by merging the data with input-output data at the sector level?

8. Can national accounts be reconciled with firms' financial statements to build an input-output table at the firm level?

These research questions focus on missing data problems in global datasets, as the aim is to have a model that covers the global economy. Most global datasets only contain information on the existence of a trade relationship between two firms (binary topology), but not on the value of the monetary transaction (weighted topology).

This thesis leaves for future work the question of reconstructing missing links (i.e., trade relationships) and then predicting the monetary value of the transactions. Some work in this regard has been done (e.g., Reisch et al., 2021; Ialongo et al., 2022), although the quality of the reconstruction has only been assessed in rare cases and to a small extent. Other studies have focused solely on predicting supply chain relationships (e.g., Brintrup et al., 2018; Mungo et al., 2022; Kosasih and Brintrup, 2022).

1.2 Contributions, summary and limitations

Paper 1: Evaluating the efficacy of Paris-compliant mitigation pathways in tackling the ecological crisis

Paper 1 addresses research question 1 by assessing whether mitigation pathways that stay within 1.5°C or 2°C of warming, as agreed in the Paris Agreement, do so at the expense of other ecological problems. To answer this question, I use the mitigation pathways considered in the WGIII contribution to the IPCC AR6, which assesses options for mitigating climate change (IPCC, 2022b). To assess whether these mitigation pathways exacerbate the ecological crisis, I use the Planetary Boundaries (PBs) framework (Steffen et al., 2015b).

Although discussed in detail in Section 2, I briefly outline the PBs framework for a better understanding of the results discussed below. The PBs framework identifies nine fundamental Earth system processes that regulate the stability of the planet: climate change, ocean acidification, stratospheric ozone depletion, nitrogen and phosphorus cycle, land system change, freshwater use, biosphere integrity, atmospheric aerosol loading and chemical pollution. For most of these processes, boundaries can be identified that, if crossed, could destabilise the Earth system on a global scale.

Of the mitigation scenarios considered by WGIII, 700 stay within the warming levels agreed in Paris. I find that few scenarios consider the control variables that underpin the processes of the PBs framework, or other variables relevant to addressing the ecological crisis. Of those that do, most scenarios do not respect the safe limits set by the PBs framework. Of particular concern is the transgression of PBs by scenarios that are considered by most actors in society (e.g., policy-makers and scientists) to be "safe paths" that we should follow. More worryingly, many of these "ideal" scenarios do not even consider ecological dimensions in their analyses.

I then examine the phosphorus PB in more detail and use an Elastic Net regression technique to understand the drivers behind its transgression. An interesting finding is that scenarios that transgress the phosphorus PB depict a future with high penetration of renewables and trading of electricity via global connectors. Even so, these scenarios use more biomass (mostly used as a biofuel) than those that stay within the boundary. These scenarios also have meat-heavy diets. Consequently, they have higher land requirements and lower forest cover.

The results for the phosphorus PB confirm the need for demand-side policies to promote more plant-based diets and the limited role biomass can play in the transition to a zero-carbon economy. Overall, the findings highlight the need for a systems thinking approach when developing mitigation pathways so that the climate and ecological crises are addressed together.

Future research should extend the analysis of the drivers leading to the breaching of a PB to the other 3 PBs (nitrogen use, freshwater use and land-system change). Furthermore, models are initialised before the present date, so that the scenarios' predictions can be compared with actual outcomes. The predictions are often off. These discrepancies should be further investigated in future research.

***Paper 2*: Firm-level production networks: what do we (really) know?**

Paper 2 addresses research questions 2-5 with a comprehensive review of the literature and a detailed analysis of three important datasets: administrative VAT data from Ecuador and Hungary, and a leading commercial dataset covering large firms in the global supply chain network (FactSet). The paper provides a taxonomy of production network datasets by conducting an extensive review of the literature. It then examines a variety of binary and weighted network properties as well as several elasticities.

We provide benchmark results from the two administrative VAT datasets, which are exceptional in that there is no reporting threshold. We call these datasets "complete" because, in principle, we observe the population of firm-to-firm transactions. In most datasets, which we call "incomplete", firms have to report a supply chain transaction if it is above a certain amount, which our two administrative datasets do not have. We compare the complete networks with FactSet and published results on national firm-level production networks.

Complete networks have remarkably similar quantitative properties, so there are properties of production networks that we believe are robust enough to be considered "really known". When incomplete networks show different patterns and deviate from complete networks in a clear direction, our results allow us to put a sign on the bias resulting from incomplete reporting. In general, weighted quantities are usually both in very good agreement between complete datasets and not dramatically biased in incomplete datasets. In contrast, binary statistics are more affected by the reporting threshold since a lower threshold leads to the presence of more links but with low weights. Therefore, the use of incomplete datasets to calibrate models can lead to targeting biased moments.

We also establish a few facts of economic significance. For instance, the distribution of the number of customers exhibits a much heavier tail than the distribution of the number of suppliers. This is intuitive, as firms may tend to grow by acquiring new customers but tend to rely on their existing suppliers when scaling up. Second, we find that many large firms have very few customers or suppliers; this suggests the existence of very large firms with limited downstream and upstream diversification, with potential implications for systemic risk. Finally, we show that the distribution of firms' centrality (the influence vector) has a divergent second moment, a key property to establish the role of production networks in aggregate fluctuations.

Overall, *Paper 2* provides the first comprehensive picture of the most fundamental statistics on production networks and can serve as a reference point for those interested in these networks with two objectives. First, it makes key moments and statistics publicly available allowing researchers, who do not have access to these data, to calibrate their macroeconomic models or create synthetic datasets. It also provides a crucial benchmark for all researchers and statisticians putting together these data. Unlike other studies that interpret facts qualitatively, we go beyond "stylised" facts and systematically provide *quantitative* estimates. Second, it contributes to an emerging agenda to develop standards for data collection, cleaning and matching for micro-level production network data around the world.

There are several limitations to this work, which we regard as a fundamental first step in an important research agenda. First, future research needs to examine in more detail the data collection methods and the comparability of firm-level datasets with national accounts. Second, future research should investigate more sophisticated properties, perhaps driven by theoretical research. For example, we have not explored quantities that make use of industrial classification systems or geographical locations; this is a clear avenue for applications.

***Paper 3*: Reconstructing firm-level input-output tables from partial information**

Paper 3 addresses research questions 6-8. It provides the first rigorous assessment of a recently developed network reconstruction method (Parisi et al., 2020) using the administrative dataset of Ecuador. While Ecuador is not a global dataset, it is an ideal case study as it (a) collects information on supply chain relationships for all (in principle) firms in the formal economy and (b) firms have to report the value of the transaction so we can assess the quality of the reconstruction method. We engineer our test network so that it is comparable to global datasets, meaning that we artificially remove firms and links that are more likely to be missing in global datasets.

It was not possible to test the reconstruction method on global datasets because transaction values are only available in Bloomberg's dataset, which is prohibitively expensive. However, we were able to access FactSet's global dataset, so an additional contribution we make is to reconstruct the input-output table of globally listed firms using this dataset. We can only judge the quality of the reconstructed network as FactSet does not provide the monetary value of the transactions.

We evaluate the network reconstruction method on the weights (transaction values), the technical and allocation coefficients (microscale quantities) and two measures of firms' systemic importance (the output multipliers and the influence vector) that are prominent in general equilibrium macroeconomic input-output models of shock propagation. We then use a general equilibrium input-output model (Acemoglu et al., 2012) to assess how shocks to firms' total factor productivity propagate through the network, ultimately affecting aggregate GDP volatility.

For Ecuador, the method reconstructs the distribution of microscale quantities reasonably well, but shows diverging results for the measures of firms' systemic importance. For both the microscale quantities and the measures of systemic importance, quantities that are more dependent on the number of customers firms have are less well reconstructed because they are more affected by missing firms and links due to the structure of the network and the sampling process of firms and links. Aggregate volatility is overestimated by the reconstruction method we employ. We also find that including the proxy node, to represent the rest of the economic activity not captured by the network, improves the prediction of the microscale quantities and of the measures of systemic importance, but worsens the prediction of aggregate volatility due to the different role the proxy node plays in the network compared to that of the firms it is supposed to represent.

To reconstruct the input-output table of globally listed firms, we merge FactSet with a global input-output dataset at the sector level (the World Input-Output Database). Differences in accounting standards between national accounts and firms' financial statements prevent us from (a) merging the two datasets at the desired country-sector or even sector level and (b) carrying out an accurate quantification (at the firm level) of the key variables making up an input-output table. We then reconstruct and compute weights, coefficients and the measures of systemic importance. The inability to accurately quantify the variables of the input-output table dramatically reduces the quality of the final dataset and thus of the reconstruction.

There are several avenues for future research. First, as already suggested by the results in *Paper 2*, *Paper 3* also highlights the need for future work to focus on reconciling firms' financial accounts with national accounts, an essential step for making a more accurate reconstruction. Second, our study assumed knowledge of the binary topology to cover a use case that could help reconstruct global datasets. A natural next step would be to predict links and then reconstruct weights, possibly with varying degrees of knowledge of the production network. To enable the study of production networks for countries that do not collect these data, it would be valuable to assess the performance of reconstruction methods when the binary topology is unknown. Lastly, our assessment of macroscale quantities was limited to a standard general equilibrium input-output model. A clear avenue for future applications is the use of different macroeconomic models (and scenarios), especially agent-based models, where we believe that the accuracy of the reconstruction of microscale quantities matters more for the model's outcomes than for general equilibrium models.

Structure of the thesis

This thesis draws from different strands of research, ranging from climate and Earth system science to economics and complexity science. I introduce the main topics and concepts in those fields and briefly review key findings in the literature that are relevant to the three papers. Each paper also has its own literature review that is more specific to its particular context.

Section 2 sets the context for the thesis with a review of the physical science basis underpinning the climate and ecological crises. Section 3 more narrowly contextualises the thesis, providing the overarching motivation. In specific, Section 3.1 reviews and classifies IAMs. Section 3.2 discusses some of the drawbacks IAMs have, from two distinct perspectives. Section 3.2.1 focuses on the ecological feasibility of the scenarios developed using IAMs and on life-cycle environmental impacts. Section 3.2.2 looks at production networks, highlighting biases that can arise from using the wrong level of aggregation and discussing why the network structure matters for firms' resilience and aggregate outcomes. Section 4 concludes by weaving together the findings of this thesis and by providing insights and advances for IAMs with the goal of strengthening their ability to address the climate and ecological crises. It does so by discussing challenges and opportunities for IAMs (Section 4.1).

Chapter 2

Key concepts in Earth system science

Since the invention of the steam engine humans have progressively become a planetary-scale geological force, so much so that Earth system scientists say that we have left the Holocene and entered a new geological epoch, the *Anthropocene* (Crutzen, 2002). The Holocene is the stable geological epoch that enabled the flourishing of human life as we know it. Crutzen identifies the beginning of the Anthropocene in the late eighteenth century "when analyses of air trapped in polar ice showed the beginning of growing global concentrations of carbon dioxide and methane" (Crutzen, 2002, p. 23).

The *Great Acceleration*, starting after the 1950s, further identifies the period in which human activities began to have a major impact on the structure and functioning of the Earth system, pointing to dramatic changes in the socio-economic and biophysical spheres of the Earth system (Steffen et al., 2015a). Since the 1950s, human societies have experienced rapid, exponential growth in population, energy use, agricultural production, fertiliser use, urbanisation, GDP and many other socio-economic indicators. As much as these changes have improved human life, they have also adversely impacted the Earth system (with exponential trends in many "environmental" variables), pushing the Earth system out of the Holocene limits. If we continue along this path, we could "potentially lead [the Earth system] to conditions that resemble planetary states that were last seen several millions of years ago, conditions that would be inhospitable to current human societies and to many other contemporary species" (Steffen et al., 2018, pg. 8253). It is important to emphasise that the distribution of the benefits of the Great Acceleration was and is highly unequal, with most of the benefits being captured by OECD countries, while environmental degradation is "exported" abroad thanks to international trade (Steffen et al., 2015a; Wiedmann et al., 2015; Dorninger et al., 2021).

Earth system scientists have identified nine fundamental Earth system processes that regulate the stability of the planet (Rockström et al., 2009; Steffen et al., 2015b): climate change, ocean acidification, stratospheric ozone depletion, biogeochemical flows, land system change, freshwater use, biosphere integrity, atmospheric aerosol loading and new entities. These processes underpin the *Planetary Boundaries* (PBs) framework, which enables us to understand what may be safe limits within which human activities should remain in order to maintain the Holocene state. For most of the nine PBs, thresholds can be identified which, if crossed, could lead to potentially irreversible damage – at least on human timescales. As discussed in more detail below, we have already crossed five of the nine PBs – six if we introduce a new boundary.

2.1 Defining and assessing planetary boundaries

I discuss PBs that we have crossed at the global scale, recognising that some of the variables that quantify the PBs at the sub-global level may affect their behaviour at the global scale. To take this into account, a two-tiered approach to defining the PBs is used where appropriate (Steffen et al., 2015b). The PBs discussed below are taken from the most recent comprehensive study (Steffen et al., 2015b).[1] There is ongoing debate in the literature about their applicability and some proposed revised boundaries (e.g., Biermann and Kim, 2020; Downing et al., 2019) or new boundaries and control variables (e.g., Running, 2012; Wang-Erlandsson et al., 2022). In line with the precautionary principle, the limits are set at the lower end of the scientific uncertainty ranges, before the threshold. This not only takes uncertainty into account but should also allow sufficient time for society to act.

Climate change. The PB that is most in the spotlight is climate change. The PB on climate change is set at 350 parts per million (ppm) of atmospheric CO_2 concentrations; we are currently at 410 ppm CO_2. There is also a second control variable, argued by Steffen and co-authors to be more fundamental: the top-of-atmosphere radiative forcing (or Earth's energy imbalance, IPCC, 2021), which indicates "the change in the net, downward minus upward, radiative flux (expressed in W/m^2) due to a change in an external driver of climate change" (IPCC, 2021, pg. 2245). The energy imbalance should not exceed +1 W/m^2. It was 2.72 W/m^2 in 2019 relative to 1750 (IPCC, 2021). A more recent paper set the PB using temperature changes and the associated likelihood of crossing tipping points. They set the boundary at 1°C of warming above pre-industrial levels. We currently are at 1.2°C of warming (Rockström et al., 2023).

Steffen et al. (2015b) choose 350 ppm as a boundary to minimise the risk of crossing thresholds that could lead to irreversible, abrupt changes in the climate system (discussed in more detail in Rockström et al., 2009). Atmospheric CO_2 concentrations, clearly, do not account for other GHGs, which are instead accounted for in the radiative forcing. The PB is set on the lower bound of the uncertainty zone, which goes from 350 ppm to 450

[1] A new study (Rockström et al., 2023) was published towards the end of writing the thesis, which also integrates justice considerations and revises some boundaries. It examines 5 of the 9 PBs. I integrate their findings with regard to revised boundaries, where appropriate, but refrain from discussing justice considerations as these are outside of the scope of this thesis.

ppm. This uncertainty zone includes the (at the time) 2°C target set by policymakers (Rockström et al., 2009).

Climate policy has always set targets in terms of temperature goals.[2] Up until the IPCC Fourth Assessment Report, temperature goals were mainly informed by atmospheric CO_2 concentrations, whereas since the IPCC Fifth Assessment Report, these have been related to carbon budgets (Lahn, 2021), i.e., "the maximum amount of cumulative net global anthropogenic CO_2 emissions that would result in limiting global warming to a given level with a given probability, taking into account the effect of other anthropogenic climate forcers" (IPCC, 2021, pg. 2220). The carbon budget had already been discussed in the scientific literature for some time. However, it gained prominence and was able to infiltrate the political debate only once scientists rigorously established the near-linear relationship between cumulative emissions and temperature changes (Lahn, 2021) – making warming independent of the emission path (Allen et al., 2009).[3]

Because of the long-lived nature of CO_2 emissions, temperature changes are determined by the cumulative amount of CO_2 in the atmosphere. To stabilise temperatures, emissions need to reach net zero. Stable CO_2 concentrations do not necessarily translate into stable temperatures (Wigley, 2005; Meehl et al., 2005; Matthews and Caldeira, 2008; Millar et al., 2016). Although estimating the carbon budget associated with a specific warming level has its issues (IPCC, 2021), atmospheric concentrations, relying on the equilibrium climate sensitivity,[4] have much higher uncertainties, especially with regard to "worst-case" equilibrium temperature changes (Allen et al., 2006; Millar et al., 2016). Therefore, it has been argued that relying on the carbon budget provides a more sound policy target as carbon budgets are better constrained by empirical observations than equilibrium climate sensitivity and are thus more robust to scientific uncertainty (Allen et al., 2009; Millar et al., 2016; Lahn, 2021).

Importantly, carbon budgets are more pertinent to policy-making also because they (1) show the relationship between cumulative emissions and peak warming (the two most important variables to climate change) and (2) highlight the link between short-term (in)action and long-term goals, something that is lost in long-term stabilisation targets based on concentrations (Millar et al., 2016; Frame et al., 2014). The move from concentration targets to carbon budgets has shifted the policy debate from the optimal rate of emissions reductions to how much more emissions can be allowed before they have to come to a halt (Lahn, 2021). Nevertheless, debates around the usefulness of the carbon budget as a policy tool remain (Lahn, 2020).

Climate change, and hence greenhouse gas (GHG) emissions, are primarily driven by the

[2]Higher emissions lead to higher atmospheric concentrations, which change the Earth's energy balance, leading to rising temperatures.

[3]Several assumptions need to hold for the relationship to be nearly linear, see Allen et al. (2009) and IPCC (2021).

[4]Equilibrium climate sensitivity is defined as"[t]he equilibrium (steady state) change in the surface temperature following a doubling of the atmospheric carbon dioxide (CO2) concentration from pre-industrial conditions" (IPCC, 2021, pg. 2223).

use of fossil fuels (IPCC, 2021, 2022b). Current levels of GHGs have already dramatically altered the climate system, with severe negative impacts on human societies. The list is so long that UN Secretary-General António Guterres, speaking at the press conference to launch the WGII contribution to the IPCC AR6, said (Guterres, 2022a):

> I have seen many scientific reports in my time, but nothing like this. Today's IPCC report is an atlas of human suffering. With fact upon fact, this report reveals how people and the planet are getting clobbered by climate change. Nearly half of humanity is living in the danger zone – now. Many ecosystems are at the point of no return – now.

Among the impacts Guterres was implicitly referring to there is sea level rise, glaciers retreat, more frequent and extreme heat, more frequent and heavier rainfalls, increased droughts, more frequent and severe fires, climate zones and species moving polewards, and oxygen loss, warming and acidification of the oceans. All of these changes have already severely impacted ecosystems' functions, structure and resilience; the same can be said with regard to impacts on socio-economic systems (IPCC, 2021, 2022a).

Bioshpere integrity. The biosphere is defined as "the totality of all ecosystems (terrestrial, freshwater and marine) on Earth and their biota." (Steffen et al., 2015b, pg. 8). Two control variables are proposed to assess their integrity: genetic diversity and functional diversity. Genetic diversity is proxied by extinction rates, with the limit set at 10 extinctions per million species-years (E/MSY). Functional diversity is proxied by the Biodiversity Intact Index (BII), which quantifies "changes in population abundance as a result of human impacts". (Steffen et al., 2015b, pg. 7). The limit on the BII is set at 90%, but with large confidence bounds (90-30%) that reflect our lack of knowledge.

A conservative estimate suggests that we are seeing extinction rates that are 100 times higher than the background rate (Ceballos et al., 2015).[5] Current extinction rates are also found comparable to, or higher than, the extinction rates that have preceded the Big Five mass extinction events (Barnosky et al., 2011). A study for southern Africa finds that BBI was 84% in 2000 (Scholes and Biggs, 2005). A very comprehensive and more recent study finds that the global average of local abundance of originally present species is 85%, with 9 out of 14 biomes crossing this PB (Newbold et al., 2016). Agriculture is the main driver behind these changes (Campbell et al., 2017). The demand for commodities of high-income countries through international trade has also been found to play a substantial role (Lenzen et al., 2012).

Biogeochemical flows. Biogeochemical flows are represented by the nitrogen and phosphorus cycle (although more might need to be included as we gain more knowledge).

[5] Steffen et al. (2015b) use a background rate of 1 E/MSY, while Ceballos et al. (2015) suggest that the background rate should be almost twice as that given more recent data and analyses.

The PB on phosphorus is set at 11 Tg/yr from freshwater into the ocean, representing the safe limit to avoid a large-scale ocean anoxic event. The PB on nitrogen is set at 62 Tg/yr from industrial and intentional biological nitrogen fixation; it aims at avoiding eutrophication of aquatic ecosystems.

Steffen et al. (2015a) assess that, globally, we are using around 22 Tg/yr of phosphorus and 150 Tg/yr of nitrogen. Both of these PBs are transgressed because of fertiliser use in agriculture production (Campbell et al., 2017; Rockström et al., 2009), with disproportional contributions of some geographical regions (Steffen et al., 2015b).

Land-system change. The PB on land system change has forest cover as its control variable. The major forest biomes (tropical, temperate and boreal) are considered because of their strong influence on the climate system beyond their region. The global PB is a weighted average of the minimum cover required in the three biomes, with a minimum total forest cover of 75%, which is around 47.9 million km^2.

We have already crossed this PB, with 62% of forest cover remaining (Steffen et al., 2015b). Intensive agriculture is found to be one of the main drivers of transgressing this PB (Campbell et al., 2017; Bowles et al., 2019; Rockström et al., 2009; Pendrill et al., 2022) as well as human settlements, although to a lesser extent (Foley et al., 2005).

Novel entities. Novel entities are "new substances, new forms of existing substances, and modified life forms that have the potential for unwanted geophysical and/or biological effects" (Steffen et al., 2015b, p. 7). This PB is particularly difficult to quantify and to establish safe operating limits since (1) there is no "natural variability" against which to compare the human-induced change and (2) there is no biophysical background level for identifying thresholds. Persson et al. (2022) propose to use a set of control variables that capture (a) trends in the production of novel entities, i.e., the production volumes of chemicals and plastics, and the share of chemicals available on the market that have safety data or regulatory assessment; (b) trends in the release of novel entities, i.e., the emission quantities of hazardous chemicals and the release quantities of plastics into the environment; and (c) unwanted impact of novel entities on Earth system processes, i.e., the toxicity of chemical pollution and the disturbance to biosphere integrity by plastic pollution.

Persson et al. (2022) conclude that there is an ever-increasing production, diversity and release of novel entities into the Earth system. These novel entities are found in remote locations and are already and they are already contaminating several sites, despite efforts for remediation. Our assessment of the safety of these substances and the subsequently required introduction of regulations cannot keep up with the speed of their production and release. This PB is thus also considered to be transgressed.

Freshwater use. The control variable for freshwater use is consumptive blue water use, where blue water refers to rivers, lakes, reservoirs and renewable groundwater stores (Rost et al., 2008) and water consumption is defined as "the part of water use [(i.e., the total amount of water extracted from its source to be utilized)] that does not return to the first water source once taken out" (Senthil Kumar and Yaashikaa, 2019, p. 5). The global threshold is set at 4,000 km^3/yr; however, Gerten et al. (2013) suggest the boundary might be lower and around 2,800 km^3/yr. Steffen et al. (2015b) report that we are at about 2,600 km^3/yr. Agriculture is one of the main drivers behind the transgression of this PB (Campbell et al., 2017).

Wang-Erlandsson et al. (2022) have recently proposed setting a PB for green water (i.e. terrestrial precipitation, evaporation and soil moisture) because of its critical role in regulating Earth system processes. The control variable chosen is root zone soil moisture and the PB is set at "$\sim$10% of the land area in which root zone soil moisture is wetter or drier than the local variability bounds". (Wang-Erlandsson et al., 2022, pg. 387). 18% of the land area is currently outside the local variability limits.

Ocean acidification. The control variable for ocean acidification is the average global surface ocean saturation state with respect to aragonite. It should be sustained at least at, or above 80% of the pre-industrial average of the annual global saturation state of aragonite. Steffen et al. (2015b) report a value of aragonite saturation of 84% in the early 2000s. Anthropogenic CO_2 emissions are the main driver of ocean acidification (IPCC, 2021). Higher aqueous CO_2 concentrations result in higher water acidity and in a decrease in the saturation state of aragonite (IPCC, 2022a). Considering that the oceans absorb around 25% of yearly CO_2 emissions (Friedlingstein et al., 2022; IPCC, 2021), the aragonite saturation state has very likely decreased. Indeed, IPCC (2022a, 2021) report a consistent decline in aragonite saturation state, and an increase in water acidity over the past 2-3 decades. If we have not breached this PBs yet, we will surely do so in the future because of climate change.

Many marine organisms, such as corals and mollusks, rely on aragonite to build their skeletons. As the saturation of aragonite decreases, it becomes more difficult for these organisms to build and maintain their skeletons, which can lead to a decline in their populations, changing marine ecosystems with negative repercussions on the livelihoods of many other species relying on these ecosystems for their survival (Rockström et al., 2009). As we will see in Section 2.2, the rate of change can be more crucial than absolute changes. The rate of change is at least 100 times faster than what was experienced over the past 20 million years, leaving little time for organisms to adapt (IPCC, 2021, 2022a).

Atmospheric aerosol loading. The control variable for atmospheric aerosol loading is the aerosol optical depth (AOD). While aerosols have a variety of effects on the Earth system, Steffen et al. (2015b) focus on their regional effects on the ocean-atmospheric circulation. They present a case study for the South Asian monsoon, where aerosols have the potential to trigger a shift in the monsoon system towards a drier state. In addition, aerosols could affect the circulation of the Asian monsoon and accelerate the melting of the Himalayan glaciers. The PB is set for the Indian subcontinent: total anthropogenic AOD must be below 0.25. Steffen et al. (2015b) report an observed annual mean of 0.3 AOD and IPCC (2021) reports increasing trends for South Asia. This PB has therefore also been breached.

A recent paper (Rockström et al., 2023) extends this boundary to the global scale (while retaining the sub-regional one). They set the global PB for the annual mean interhemispheric AOD difference at 0.15. This choice is based on evidence that the AOD difference between the north and south hemispheres can trigger tipping points at the regional scale and significantly alter regional hydrological cycles. Rockström et al. (2023) report that the current value is about 0.05 on average, reaching about 0.1 in the boreal spring and summer.

In addition to destabilising the Earth system, atmospheric aerosol loading also has adverse effects on human health and ecosystems (IPCC, 2022a). Particulate matter is an aerosol and human exposure to it can cause premature death, heart problems and respiratory diseases. The number of deaths is estimated at 4.2 million per year, although other pollutants are also included (WHO, 2022).

Stratospheric ozone depletion. The control variable for stratospheric ozone depletion is the ozone concentration, which should not fall below 275 DU (Dobson units). This PB is only breached over Antarctica during the austral spring, when ozone concentration falls to around 200 DU (Steffen et al., 2015b). A decrease in stratospheric ozone concentration leads to a thinning of the ozone layer, allowing more incoming ultraviolet radiation from the sun, with adverse effects on human health and marine organisms (Rockström et al., 2009). This PB has been breached, mainly due to anthropogenic ozone-depleting substances such as chlorofluorocarbons. Thanks to the Montreal Protocol of 1987, ozone concentrations have increased and the ozone layer has partially recovered (Steffen et al., 2015a). This is a very nice example, and probably the only one, where coordinated global action has been able to halt human-induced changes and lead to a recovery (which is still unfolding) after a PB was breached.

As much as the PBs framework has received broad support, it has also received some criticism (see Biermann and Kim, 2020; Downing et al., 2019, for reviews). Earth system scientists have pointed out, for example, that the definition of some PBs may be inadequate and that the inclusion of some sub-scale processes that are not systemic and do not exhibit threshold behaviour should be excluded; others have simply suggested adding new processes.

According to development scholars, the PBs framework fails to address global inequality and equity,[6] with some scholars seeing an underlying, implicit need for the Global South to limit economic growth and thus potentially development. Social sciences, humanities and science and technology studies also criticise the PBs framework's relevance to governance and potential agency issues. They argue that the target-setting approach is expert-driven and top-down – with all the experts, in this case, from rich industrialised countries. Policymakers are then left with the sole function of ensuring that these limits are respected. Instead, critics argue, stakeholders and the public should be involved in a process of consultation and deliberation. In addition, the use of the precautionary principle in the PBs framework may not take into account the varying degrees of risk taking and risk aversion of actors in different societies. These factors should be taken into account when setting targets.

2.2 Interacting elements and tipping points

So far we have discussed elements of the Earth system in isolation; however, these can interact, possibly even leading to a cascade of domino effects. Many of the PBs processes do, in fact, interact, forming a dense network of complex interactions. These interactions reduce (possibly also to zero) the safe operating space because the boundary values and thresholds can change as a result of another one changing, so they are better thought of as *conditional* thresholds. These complex interactions can also induce domino effects, thereby amplifying human impacts (Lade et al., 2020; Anderies et al., 2013; Heitzig et al., 2016).

The PBs and elements in the climate system can exhibit critical thresholds – also known as *tipping points* – such that a small change in external conditions can cause the system to shift into a new regime (Lenton et al., 2008). In the Earth system, this could lead it to tip into an undesirable state, away from the Holocene stable state. Reversing the system back to its prior state is often difficult, particularly on a human timescale. In fact, some of these tipping elements have been deemed "too risky to bet against" (Lenton et al., 2019). A tipping point can also be crossed due to what is called *rate-induced* tipping, whereby the system might be far from the tipping point but the rate of change is fast enough to cause the system to tip (Ashwin et al., 2012; Lohmann and Ditlevsen, 2021; Lohmann et al., 2021).

Almost two decades ago, the IPCC recognised the existence of tipping elements and determined that they were likely to be crossed at 5°C of warming (Lenton et al., 2019). As we have gained more knowledge over time, these tipping points have been revised downwards to lower degrees of warming. A recent study finds that there are a total of 14 climate tipping elements (e.g., the West Antarctic and Greenland ice sheets, the Amazon rainforest, the Boreal permafrost and the Atlantic meridional overturning circulation;

[6] A recent paper (Rockström et al., 2023) has revised and expanded the PBs framework to encompass justice criteria. The forthcoming response to this update remains to be witnessed. Although subjective value judgements are inherently inevitable, the extent of disagreements can vary based on the specific situation.

AMOC). Five of these tipping elements may have already crossed their critical threshold at current levels of warming and another six are likely to cross it within the Paris agreed targets (i.e., 1.5–2°C of warming; McKay et al., 2021). Steffen et al. (2015b) asses that climate change and biosphere integrity are the two most important PBs, where crossing the threshold of either of the two could, on its own, drive the Earth system out of the Holocene stable state.

Tipping elements also interact and could trigger a *cascade of domino effects*. Domino effects are still poorly understood, but Wunderling et al. (2021) find that the "Greenland and West Antarctic [ice sheets] are oftentimes the initiators of tipping cascades, while the AMOC acts as a mediator transmitting cascades" (Wunderling et al., 2021, pg. 601). The West Antarctic ice sheet may have already crossed the tipping point, Greenland and the AMOC are showing signs of destabilisation that suggest they are approaching the tipping point. Each of these tipping elements could trigger further changes. For instance, the collapse of the AMOC would subsequently lead to a cooling of the North Atlantic and a warming of the Southern Hemisphere. It would also change monsoon patterns, which would then intensify or cause droughts in the Sahel and Amazonia, leading to a further degradation of ecosystems and carbon sinks, and further fuelling climate change (Lenton et al., 2019; McKay et al., 2021). The extent of the domino effects could be so vast that biogeophysical feedbacks in the Earth system could become the dominant process and lead to an irreversible "Hothouse Earth" pathway (some researchers suggest that this might already happen at 1.5–2°C, Steffen et al., 2018).

In this section, we have explored the complex relationships among the various components of the Earth system and how these interact, some playing a more significant role in causing irreversible changes than others. To tackle the challenges of climate change and ecological degradation, it is essential to adopt a systems-thinking approach that considers the interactions among the elements of the Earth system and human societies, as well as interactions among elements within each system. How have scientists and economists approached this so far? In the following section (Section 3), I will examine this question by delving into the quantitative models that have been developed to understand and address these crises.

Chapter 3

Integrated Assessment Models

To understand the Earth system and the socio-economic processes discussed in Section 2, scientists have developed computer models. These models have been instrumental in understanding possible futures associated with different degrees of human impact and some have also been influential in providing support to decision-makers regarding what policies might be best to adopt to stay within 1.5°C or 2°C of warming.

As highlighted before, climate change has received most of the attention among all the other ecological problems we are facing. This is also reflected in the modelling efforts, where considerable effort has been put towards the goal of having an integrated model representing climate and socio-economic processes, and how these two systems interact; these models are known as Integrated Assessment Models (IAMs). Roughly, IAMs combine an economic model with a physical climate model. They may also integrate a rather detailed description of the energy system and the nexus with water and land use.

In Section 3.1, I review IAMs by classifying them based on three criteria: model class (Section 3.1.1), modelling technique and economic framework (Section 3.1.2). In Section 3.2, I discuss the drawbacks of IAMs with a focus on environment-economy interactions (Section 3.2.1) and on interactions in the production system (Section 3.2.2). The first drawback is at the heart of *Paper 1*, while the second is the focus of *Paper 2* and *Paper 3*.

3.1 Typology of Integrated Assessment Models

Before delving into classifying IAMs, we should consider what features we would like these models to have. A useful classification has been proposed by Hourcade et al. (2006), who define three axes: technological explicitness, macroeconomic completeness and microeconomic realism. These axes do not exhaustively include all the features we would want IAMs to have, but they are a good starting point.

A *technology explicit* model has a detailed technological representation of the supply and demand side of the energy system, which allows us to explore various low-carbon technological futures and assess the impacts of technology policies on the socio-technical system, for example, by lowering financial costs and thus promoting the diffusion and adoption of technologies. Examples of technology explicit IAMs are MARKAL (Loulou, 2004) and TIAM-UCL (Anandarajah et al., 2011). *Macroeconomic completeness* implies that the model represents all the economic activities in the economy, i.e., the production sector, households, the financial sector, the government and international trade. Macroeconomic completeness is useful for understanding the impact of policies on, e.g., input substitution, structural change, competitiveness, employment and public finances. An example is GEM-E3 (Van Regemorter et al., 2013). *Microeconomic realism* describes the behaviour of firms and consumers, and allows us to understand how policies affect their choices, behaviour and practices. Examples are BLUE (Li and Strachan, 2017) and DSK (Lamperti et al., 2018).

IAMs can be categorised according to different criteria, none of which provides a hierarchical classification. A first, instructive approach is to consider three criteria: model class, modelling technique and economic framework. Looking at the model class, IAMs can be either cost-benefit or process-based. When organised by modelling technique, IAMs use either optimisation or simulation techniques. Optimisation models tend to use neoclassical and new-Keynesian economic frameworks, while simulation models also use other frameworks, such as post-Keynesian and complexity economics. In most cases, optimisation IAMs are process-based models, but the boundary with cost-benefit models is somewhat blurred, as one could add a damage function and perform a cost-benefit analysis (CBA). However, cost-benefit models tend to be aggregate models with little detail, something that process-based models tend to depart from. For example, IMACLIM (Waisman et al., 2012) is mostly process based, but has a damage function that can be easily turned on and off. The model space is thus very heterogeneous, without well-defined and clear boundaries. I first classify IAMs by model class (Section 3.1.1) and then by modelling technique and economic framework (Section 3.1.2).

3.1.1 Model classes

This section organises IAMs according to the model class. I follow Weyant (2017) and IPCC (2022b), and distinguish between cost-benefit and process-based IAMs.

Cost-benefit IAMs (CB-IAMs) represent the techno-economic system at a rather aggregate level, use a simplified climate model and can have one or more world regions, e.g., PAGE (Hope and Hope, 2013), RICE (Nordhaus, 2010) and FUND (Anthoff and Tol, 2013). For example, RICE considers aggregate GDP for different world regions and is based on neoclassical economic growth theory. Therefore, the model has one rational representative agent (for each region), i.e., a social planner, who maximises social welfare. While it scores decently on macroeconomic completeness, it scores poorly on microeconomic realism and technology explicitness (as it only includes a carbon-saving technology). It incorporates feedbacks between the economy and the climate using a simplified climate module.

CB-IAMs aim to assess the impact of rising climate temperatures on consumption and to explain how economic activity affects the climate through the emission of greenhouse gases (GHGs). CB-IAMs inform policy decisions by estimating the "optimal" level of the social cost of carbon through a CBA. The social cost of carbon is then used to assess the level of a potential carbon tax. CBA helps to appraisal of the costs and benefits of action and inaction on climate change, and therefore to compare and choose between different policy options. By comparing costs and benefits, CB-IAMs also derive the "optimal" level of warming. To estimate the costs associated with higher temperatures, these models use a damage function (known to be highly problematic, Pindyck, 2013; Farmer et al., 2015), which models the economic impacts of climate shocks associated with different levels of warming.

Process-based IAMs (PB-IAMs) couple a macroeconomic model with a technologically detailed model of the energy system. Most of these models use a climate model and some also use other physical models to capture land use change, food production and water use, each with varying levels of detail in different IAMs (Harfoot et al., 2014). The macroeconomic model may have a disaggregated sectoral representation and often includes several world regions. The social planner can have perfect or limited foresight, or be myopic, meaning that it has no knowledge about the future (IPCC, 2022b; Keppo et al., 2021). With the exception of a few (e.g., IMACLIM, Waisman et al., 2012), PB-IAMs cannot assess the economic damages associated with rising CO_2 emissions, hence they cannot estimate the social cost of carbon and an "optimal" level of warming, as CB-IAMs do. Instead, a temperature target is set and a number of model runs are carried out, each assuming a different carbon price. PB-IAMs perform what-if analyses aimed at finding the cost-effective mitigation pathway under predefined outcomes (e.g. staying within the carbon budget to limit temperature to 1.5°C or 2°C; Wilson et al., 2021; Weyant, 2017). Some examples are REMIND (Luderer et al., 2015), WITCH (Emmerling et al., 2016), Imaclim-R (Waisman et al., 2012), MESSAGEix-GLOBIOM (Krey et al., 2020) IMAGE (van Vuuren et al., 2015), MEDAS (Capellán-Pérez et al., 2020) and DSK (Lamperti et al., 2018).

3.1.2 Modelling techniques and economic frameworks

This section classifies IAMs according to the modelling technique they use, which can be either optimisation or simulation (Keppo et al., 2021; IPCC, 2022b). Within each modelling technique, I also discuss the different economic frameworks that IAMs use (IPCC, 2022b).

Optimisation models. In general terms, optimisation models view the economy as composed of rational, homogeneous agents with optimising behaviour that ultimately lead the economy to an efficient equilibrium. Optimisation models can be divided into macroeconomic growth models, computable general equilibrium models, dynamic stochastic general equilibrium models and partial equilibrium models (Hafner et al., 2020; Stanton et al., 2009).

Macroeconomic growth models (e.g., Nordhaus, 2010) depict the economy at an aggregate level, having only one sector represented by aggregate GDP. They have a single production function that integrates capital and labour to produce a generic product for consumption and investment. If the model has more than one world region, each region has its own production function. Usually, these models simply have a generic carbon-saving technology that reduces CO_2 emissions. Macroeconomic growth models are rooted in neoclassical growth theory, so they assume perfect competition and have a rational representative agent who maximises social welfare. They tend to score decently in terms of macroeconomic completeness, and poorly on microeconomic realism and technological explicitness.

Computable General Equilibrium (CGE) models (e.g., Waisman et al., 2012; Van Regemorter et al., 2013) are based on general equilibrium theory. They can either represent economic output at an aggregate level or differentiate between several industries linked by input-output relationships. Unlike macroeconomic growth models, CGE models with aggregate output include energy as a factor of production. The representative agent maximises social welfare and may have perfect or limited foresight, or be myopic. Equilibrium is found by the set of prices that clear markets, i.e., that equate supply and demand. If there are several world regions, they are connected by international trade. These models tend to score high on macroeconomic completeness, low on microeconomic realism and have varying degrees of technology explicitness, although it is usually quite high.

Dynamic Stochastic General Equilibrium (DSGE) models (e.g., Cai et al., 2012; Golosov et al., 2014; Lontzek et al., 2015), as the name suggests, are rooted in general equilibrium theory, but also on real business cycle and New Keynesian theory. These models also have an aggregate representation of the economy, similar to macroeconomic growth models. Compared to CGE modes, DSGE models are better able to account for shocks, uncertainties in economic variables, market imperfections and nominal rigidities (Farmer et al., 2015), but are otherwise much less sophisticated. They score relatively well on macroeconomic completeness, low on microeconomic realism and poorly on technology explicitness.

Unlike the others, *partial equilibrium* models (e.g., Anandarajah et al., 2011) focus on one sector or a subset of sectors only while taking the rest of the economy as given.

Simulation models. There are fewer simulation models compared to optimisation models. These include macroeconometric, agent-based and system dynamics models. Compared to optimisation models, simulation models can better account for interdependencies between components of the system, agent heterogeneity, feedback mechanisms and non-linear dynamics. Simulation models may also allow for increasing returns to scale, which equilibrium models cannot do due to numerical instability (Mercure et al., 2016).

Macroeconometric models are based on econometric theory and mostly post-Keynesian economic theory. They use econometric techniques to extrapolate historical trends to make projections into the future. An example is E3EM (Dwesar et al., 2022); it scores well on three dimensions (i.e., macroeconomic completeness, microeconomic realism and technology explicitness).

System Dynamics models use a system of partial differential equations to analyse cause and effect relationships within a system, usually visualised by causal-loop diagrams. Variables are divided into stocks and flows. Stocks represent values that accumulate over time, while flows are the rate of change of a stock (Forrester, 1961). The most famous system dynamics model in the sustainability context is the World3 model developed for the The Limits to Growth report (Meadows et al., 1972). IAMs using system dynamics have varying degrees of technological explicitness, macroeconomic completeness and microeconomic realism. For example, the EUROGREEN model developed by D'Alessandro et al. (2020) scores high on macroeconomic completeness but low on technological explicitness and microeconomic realism. Instead, the BLUE model (Li and Strachan, 2017) is rich in technological detail and scores well on microeconomic realism, but performs very poorly on macroeconomic completeness. The MEDAS model (Capellán-Pérez et al., 2020) is a technology explicit and macroeconomic complete model that performs poorly in terms of microeconomic realism.

Agent-based (ABMs) models are computer "simulations of the economy based on interactions of a large number of heterogeneous agents according to specified rules that aim to allow more flexible and realistic characterisations of socio-economic systems." (Farmer et al., 2015, pg. 330). There are few agent-based IAMs. For example, the Lagom regiO (Wolf et al., 2013) and ENGAGE (Gerst et al., 2013) models, which are macroeconomic complete and score well on microeconomic realism. While the Lagom regiO model falls short in terms of technological explicitness, the ENGAGE model is somewhat more sophisticated. The DSK model developed by Lamperti et al. (2018) is almost macroeconomic complete, missing only the government, scores quite well on microeconomic realism, but is rather sketchy on technological detail.

3.2 Drawbacks of Integrated Assessment Models

Although IAMs have provided a valuable starting point for understanding the relationship between climate change, GHG emissions and the socio-economic system, they suffer from several limitations (Farmer et al., 2015; Mercure et al., 2016; Hickel et al., 2021; Pedersen et al., 2022). The list of limitations is long, covering issues such as the ethical implications of selecting the value of the discount rate (Stern, 2007, Part I, Chap. 2) and Weitzmann's Dismal Theorem, which discusses the unsuitability of CBA in the context of a fat-tailed distribution of climate disasters (Weitzman, 2009). Other limitations include the modelling of the economic system (i.e., they do not account for its complexity, as discussed in Chapter 1) and technological progress (Mercure et al., 2016; Axtell, 2014; Farmer et al., 2015), the disregard of global justice issues (Kanitkar et al., 2022) and the often unrealistic scale of deployment of negative emissions technologies (Anderson and Peters, 2016; Hickel et al., 2021).

This thesis delves into three significant drawbacks, addressing them within the context of optimisation PB-IAMs. However, other typologies of IAMs are not necessarily exempt. I focus on optimisation PB-IAMs for three reasons: (a) they are the most widely used models; (b) they typically include modules representing Earth system processes – except for a few examples like Capellán-Pérez et al. (2020) and Allen et al. (2019); and (c) they constitute the bulk of the contributions to IPCC reports, making their scenarios highly influential in shaping societal outcomes. Throughout this section, I will refer to IAMs for short rather than optimisation PB-IAMs.

The three drawbacks I focus on are the neglect of

1. ecological issues
2. life-cycle environmental impacts, and
3. supply chain networks at the firm level.

Neglecting (1) and (2) leads to overly optimistic scenarios and problem shifting, thereby exacerbating the ecological crisis (Arvesen et al., 2011; Liu et al., 2015). This oversight further prevents the identification of synergies and trade-offs between different technology options (Hertwich et al., 2015; Hertwich and Wood, 2018). These two drawbacks are discussed in Section 3.2.1.

Neglecting (3) implies that IAMs are missing processes at the micro-level shown to be crucial in shaping macroeconomic outcomes. IAMs either do not account for supply-chain interactions or model them at a level that is too aggregated, resulting in substantial information losses and biased results (Diem et al., 2023; Inoue and Todo, 2019). In particular, when shocks propagate throughout the network, its structure, feedback mechanisms and non-linearities can lead to sizeable aggregate fluctuations, triggering major disruptions and possibly also a cascade of defaults (Moran and Bouchaud, 2019; Lamperti et al., 2018; Inoue and Todo, 2019). Therefore, overlooking supply chains prevents us from understanding the resilience of the production system (Henriet et al., 2012; König et al., 2022) and, for example, also from identifying possible bottlenecks in the system (Otto et al., 2017; Pichler and Farmer, 2022). Section 3.2.2 delves into these topics.

3.2.1 Neglecting ecological issues and life-cycle impacts

This section examines drawbacks (1) and (2). I first discuss ecological issues – i.e., drawback (1) – and then life-cycle environmental impacts – i.e., drawback (2). Ecological issues are assessed through the lens of the PBs framework.

Neglecting ecological issues. IAMs account for Earth's subsystems and processes that are relevant to climate change. They usually can include (with different levels of detail) a terrestrial vegetation model, a hydrological model, and a model of the ocean and atmosphere (Harfoot et al., 2014).

In IAMs, human societies affect the Earth System not only through the release of CO_2 into the atmosphere, but also through, for example, the production and consumption of food, wood products and bioenergy, thereby clearing vegetation and reducing the amount of land available for ecosystems. The drivers of human land use in IAMs are only prices and land rents (e.g. Fujimori et al., 2014; Krey et al., 2020; Waisman et al., 2012). Some models also account for the land requirements of human settlements (e.g., Rothman et al., 2016; Keramidas et al., 2018; Calvin et al., 2019). However, as *Paper 1* shows, land use is not constrained to a safe level that would prevent ecosystems from tipping into undesirable states. For instance, the Amazon rain forest could, after a certain threshold, tip into a savanna or grassland and become a net CO_2 source, something that may be already underway for easter Amazonia due to deforestation driven by primarily by pasture, but also crops (Gatti et al., 2021).

Food production also affects the Earth system by realising nitrogen and phosphorus (used as fertilisers) into the environment. Nitrogen and phosphorus use is accounted for in some IAMs, but, as *Paper 1* shows, these are not constrained to sustainable levels that would avoid dramatic changes in terrestrial, marine and aquatic ecosystems. To give an example, if constraints on the nitrogen, phosphorus and land-system change PBs (and also the freshwater PB, as I discuss below) were enforced in IAMs, the deployment of bioenergy with carbon capture and storage (BECCS) would be severely curtailed due to the costly land and fertiliser (and water) requirements. In fact, several studies find that an excessive reliance on BECCS would lead to a transgression of the nitrogen, phosphorus, land-system change, freshwater and biosphere integrity PBs (Heck et al., 2018a, 2016). BECCS also entails trade-offs among competing needs such as food production and other land uses (Sen and Dabi, 2021).

Many IAMs do not consider water use (e.g., Anandarajah et al., 2011; Emmerling et al., 2016), some use it for the energy sector only (e.g., Fricko et al., 2016) and others have a more extended treatment (e.g., Calvin et al., 2019; Hughes and Hedden, 2016). For instance, GCAM (Calvin et al., 2019) considers both water demand and supply for each region and sector (e.g., energy sector, manufacturing sector, livestock and municipal use), but leaves supply unconstrained. The drives of water consumption in GCAM are population, GDP per capita, industrial output, technologies and agriculture. Another example is IMAGE, which has an integrated assessment of the water cycle (an account of this is given in Schaphoff et al., 2018). IAMGE includes the water demand of the agricultural and electricity sector, and of industries and households. It considers water availability, withdrawals and water stress.

Using IMAGE, Fricko et al. (2016) show that nuclear and CCS are particularly water intensive – something highlighted also by Mouratiadou et al. (2018). Under the different 2°C decarbonisation pathways they consider, these technologies would lead to increased water consumption due to water-intensive power cooling technologies and would also adversely affect aquatic ecosystems because of thermal pollution. Increased irrigation for, in specific, BECCS would severely stress water systems (Mouratiadou et al., 2016; Bonsch et al., 2016; Beringer et al., 2011) and likely lead to a transgression of the freshwater PB (Heck et al., 2018a).

Neglecting life-cycle impacts. It is the amount of natural resources used and the cumulative environmental impacts that ultimately matter (Jackson and Victor, 2019; Parrique et al., 2019). In some IAMs, technology choice takes into account some of the indirect emissions. For instance, they consider emissions like methane leakages from fossil fuel production and emissions connected to land use changes related to biomass production. However, these models often overlook a range of other indirect emissions related to infrastructure, manufacturing, construction and transport (Arvesen et al., 2018; Daly et al., 2015). More generally, IAMs do not account for life-cycle environmental impacts (i.e., the impacts occurring throughout the supply chain), such as chemical pollution and eutrophication, and neither account for natural resource use (Pauliuk et al., 2017; Budzinski et al., 2023). Life-Cycle Analysis (LCA) studies account not only for the life-cycle environmental impacts, but also for the natural resource requirements.

LCA studies point out that renewable technologies shift the impacts from the operation to the construction phase, and partly also to the decommissioning phase (Hertwich et al., 2015; Naegler et al., 2022; Junne et al., 2020; Luderer et al., 2019). These studies also find that while renewable technologies have lower environmental impacts (e.g., eutrophication, ecotoxicity and particulate matter) than fossil fuel-based technologies, renewables increase the demand for land, and certain minerals and materials, in particular steel and cement (Hertwich et al., 2015; Naegler et al., 2022; Xu et al., 2020; IEA, 2021). Material extraction and use put pressure on ecosystems and are linked to most of the PBs (Hoekstra and Wiedmann, 2014), while the mining industry is one of the most polluting industries and it is also difficult to decarbonise (Lee et al., 2020). Whether the environmental impacts should be of concern depends on the scenario. Therefore, the trade-offs of renewable energy technologies and their life-cycle environmental impacts need to be accounted for when developing mitigation scenarios.

This Section has shown that IAMs fail to assess the climate and ecological crises in a holistic manner (Arvesen et al., 2011; Harfoot et al., 2014; Liu et al., 2015) by neglecting the life-cycle environmental impacts of the proposed solutions and the ecological limits posed by the carrying capacity of the planet. Therefore, IAMs may underestimate the environmental impacts of mitigation pathways, which could lead to problem shifting and to overly optimistic scenarios about our ability to address climate change (Arvesen et al., 2011; Liu et al., 2015; Kouloumpis et al., 2015). Taking into account the material flows and environmental impacts that occur throughout the supply chain would allow us to consider for the synergies and trade-offs between climate and environmental concerns for each technology option, resulting in scenarios that address both the climate and the ecological crisis. The level of granularity at which we would be able to carry out this analysis depends, of course, on the availability of data. As I discuss in Section 4.1, it is currently almost impossible to carry out such an analysis at the firm level.

3.2.2 Neglecting firm-level supply-chain networks

Interactions in the economy take different forms and involve different actors. In my thesis, I focus on supplier-customer interactions in the production system; that is, on supply-chain networks, also known as production networks. Production networks have been shown, for example, to influence aggregate fluctuations (e.g., Acemoglu et al., 2012; Mandel et al., 2015; Di Giovanni and Levchenko, 2012; Contreras and Fagiolo, 2014), economic growth (e.g., McNerney et al., 2022; Acemoglu and Azar, 2020) and innovation activity (e.g., Carvalho and Draca, 2018; Carnovale and Yeniyurt, 2015).

CGE models, but also other economic models used in IAMs based on input-output tables (e.g., Capellán-Pérez et al., 2020), represent production networks at the sectoral level, whereas others ignore them. Below, I summarise findings that highlight why the sector level is not the appropriate aggregation level to study economies and why it is important to know the network structure at the firm level.

The correct level of analysis. Recent results show that analysing the economy at the sector level rather than at the firm level can lead to significant information losses and biased results (Inoue and Todo, 2019; Diem et al., 2023, 2022). Estimating economy-wide losses from micro-level shocks using sector-level data could underestimate losses by 18.7% on average and sometimes by up to 37% (Diem et al., 2023). Sector-level data lead to more optimistic results because aggregation washes away firm heterogeneity and assumes that all firms within the same sector are equally affected by the shock (as also note by Henriet et al., 2012). However, firms within a sector are highly heterogeneous and show little similarity in terms of the suppliers they rely on and the customers they sell to (Diem et al., 2023). Moreover, the macro-level consequences of micro-level shocks cannot be deduced by simple aggregation due to non-linearities, feedback loops and heterogeneity. As shocks propagate through a network of heterogeneous agents, they are amplified, leading to larger aggregate effects (Lamperti et al., 2018).

Importance of the network structure. A *scale-free* network is a network that has a power-law degree distribution (for a production network it corresponds to the distribution of the number of suppliers/customers firms have), where most nodes have a very low degree and few nodes have a very high degree. Such a network structure can make the economic system more sensitive to shocks, leading to large aggregate fluctuations and possibly also a cascade of failures (Moran and Bouchaud, 2019; Bonart et al., 2014; Inoue and Todo, 2019;

Magerman et al., 2016; Henriet et al., 2012; Brintrup et al., 2015b). As shocks propagate through the network these are amplified leading to indirect losses higher than direct ones (Inoue and Todo, 2019; Lamperti et al., 2018).

Scale-free networks are characterised by the presence of *hubs*, i.e., firms with a very large number of partners, which shorten the distance between any two firms. Hub firms drive the economic system towards instability because when they are hit by a shock, a large number of firms in the economy are immediately affected. In fact, this relates to a well-known result in network theory: scale-free networks are robust to random attacks, but the presence of hubs makes them extremely vulnerable to targeted attacks. On the contrary, random networks, in which the degree distribution follows a Poisson with tails that decay exponentially (i.e., most firms have the same number of connections), have a low degree of tolerance to random attacks (Albert et al., 2000).

Thadakamaila et al. (2004) shows that a network structure that lies somewhere between a random and a scale-free network is more robust to targeted attacks than a scale-free network, while the robustness to random attacks remains comparable to that of a scale-free network. Brintrup et al. (2015b) show that the supply chain network of the automotive industry may indeed lie somewhere in between a random and a scale-free network. A key feature to achieve the in-between network structure suggested by Thadakamaila et al. (2004) is to limit the number of connections that firms can form, leading to a degree distribution with a tail that decays much faster than a power-law, but slower than an exponential.

Studies in the Supply Chain Management (SCM) literature, which focus on the supply chain network of a single firm or industry, have debated about the functional form of the degree distribution as it implies different degrees of robustness of the network. As we explain in *Paper 2*, we refrain from such a debate while still providing useful results to inform robustness. Nevertheless, I briefly summarise the findings in the SCM literature. Some empirical studies suggest a power-law form (Orenstein, 2016, 2021), others an exponential (Brintrup et al., 2015b, 2011a,b) and Kito et al. (2014) suggest a lognormal, for the distribution of the number of suppliers, and a stretched exponential, for the number of customers. Brintrup et al. (2017) can neither refute nor confirm the power-law hypothesis but note that real-world networks do not follow a pure power-law as, usually, after a cut-off point, the degree distribution decays exponentially (as also discussed in Hearnshaw and Wilson, 2013, and as we discuss in *Paper 2*, Appendix B.1). This can be due to (1) the finite size of the economy, (2) constraints on the ability of firms to form connections (e.g., "size" of final demand and transaction costs), (3) the ageing and exit of firms, and (4) finite-size effects induced by the data collection process (Hearnshaw and Wilson, 2013; Brintrup et al., 2017).

Regardless of the exact functional form of the distributions, very broad degree distributions imply the presence of hubs, which we confirm in *Paper 2*. We also find a fairly short average *path length* of around 3 – similar to findings in the SCM literature (Brintrup et al., 2017, 2015b; Hearnshaw and Wilson, 2013; Orenstein, 2016, 2021). A short path length implies that when a (hub) firm is affected by a shock, the shock can quickly reach a (very) large number of firms and potentially spread to the whole network in just a few steps. At the same time, a short path length could increase the resilience of the network, as more upstream firms would be less affected by changes in demand (Brintrup et al., 2011b) that are amplified as they propagate through the network – a phenomenon known as the bullwhip effect (Lee et al., 1997). It has also been argued that shorter path lengths lead to a more efficient flow of goods and services, as fewer steps and less time are required to complete a given task (Kim et al., 2011; Hearnshaw and Wilson, 2013).

Firm-level determinants of resilience. The structure of the production network is affected by the local choices of firms (Choi and Hong, 2002; Choi et al., 2001) and how their choices then combine to give rise to the production network (Choi et al., 2001). It is impossible for a firm to have knowledge of the entirety of its supply chain, so a firm acts based on local knowledge of its supply chain. The broader the knowledge it has of its supply chain (beyond its immediate partners), the better the firm can respond to changes in the environment (Choi et al., 2001; Christopher and Peck, 2004). So, the structure of the supply chain also affects a firm's behaviour (Choi and Kim, 2008).

The resilience of the network deteriorates when there are few *redundancies* in the system (Henriet et al., 2012; Moran and Bouchaud, 2019; Otto et al., 2017; Brintrup et al., 2015a; Hearnshaw and Wilson, 2013; Christopher and Peck, 2004), meaning that firms rely on few suppliers making it hard for them to *substitute* their inputs and absorb the shock. Substitution becomes even more relevant when the supplier provides specialised, rare inputs (Brintrup et al., 2015a; Barrot and Sauvagnat, 2016). Constraints in input substitution increase both the impact and persistence of shocks (Inoue and Todo, 2019). Firms might be driven to have a limited number of suppliers, for instance, because of high searching and matching costs (Moran and Bouchaud, 2019). Indeed, one of the major factors that contribute to shaping a supply chain has been found to be cost considerations (Choi and Hong, 2002).

The degree of *monopolistic power* also plays a critical role in shock transmission. For example, firms with a higher market power might be able to absorb the shock by increasing their prices, but this comes at the detriment of their customers, who might be able, or not, to transmit the price increase further down the chain (Wu, 2016).

However, sufficient *inventories* can buffer disruptions, thereby increasing the robustness of the network (Henriet et al., 2012; Otto et al., 2017; Colon et al., 2021; Hearnshaw and Wilson, 2013; Christopher and Peck, 2004). Shocks can be critically amplified also because of supply chain *bottlenecks* and firms' *rationing* strategies (Otto et al., 2017; Pichler and Farmer, 2022). Kim et al. (2011) suggest that firms with high betweenness centrality may be potential bottlenecks in the system – they may be "gatekeepers" – as a high betweenness centrality means that a firm lies on many shortest paths that connect any two firms in the network, hence these firms are operationally critical since they have a very high degree of control over the flow of goods and services. However, betweenness has its limitations since there may be other paths that are still viable and in use, but that are not the shortest path between two firms (Brintrup et al., 2017).

These findings suggest that specialised, cost-efficient global supply chains geared for just-in-time production decrease the resilience of the global economy by increasing vulnerability at the local level of the firm (as also note, for instance, by Henriet et al., 2012 and Christopher and Peck, 2004).

To have a better chance of withstanding shocks, firms also need to be able and willing to adapt their, e.g., goals and infrastructure in response to a changing environment (Choi et al., 2001; Hearnshaw and Wilson, 2013). Response time is also key to the resilience not only of the firm but of the whole supply chain (Choi et al., 2001; Christopher and Peck, 2004; Hearnshaw and Wilson, 2013), as "[r]esilience implies agility" (Christopher and Peck, 2004, pg. 13). Similarly, by having shared norms and procedures (development and manufacturing processes, work norms, etc.) with its partners a firm can enhance its resilience as well as that of the other firms in the supply chain. Shared norms and procedures also lead to lower transaction costs and increased communication efficiency (Choi et al., 2001). Communication among firms in a supply chain, especially regarding risks and uncertainties, and collaborative planning also lead to higher resilience, once again, not only of the firm but of the whole supply chain (Christopher and Peck, 2004).

A few studies focusing on climate change. Stangl et al. (2023) show the importance of firm-level supply chain networks in the context of climate change mitigation. Having information on the national production network enables one to identify strategic firms (i.e., leverage points) in the network that would yield the highest emission reductions while reducing adverse repercussions on the labour market.

Lamperti et al. (2018) develop an ABM IAMs with capital and consumption good firms, a monopolist energy firm, a stylised financial sector, the government and households. The model considers feedbacks between the economy and the climate system through a stylised stochastic damage function. As the economy emits CO_2, the temperature rises, leading to more extreme and frequent climate events. Shocks hit firms and workers heterogeneously and can impact labour productivity, firms' capital stocks, inventories and energy efficiency, eventually affecting the economy at the aggregate level. However, as shocks propagate through the network of heterogeneous agents, they are amplified, resulting in larger aggregate effects. The capital stock and inventory levels are found to be the main drivers of the economy's instability, while shocks to labour productivity affect economic growth and unemployment.

Interactions between heterogeneous agents shape the diffusion and adoption of technologies through imitation and learning (Mercure et al., 2016). In the context of electricity markets, Barazza and Strachan (2020) develop an ABM in which heterogeneous agents make investment decisions with limited foresight. Market participants learn from past failures or successes, and may imitate fruitful investment strategies of their peers. Their actions are also influenced by the carbon price, set by the government. Similarly, actors' strategies co-evolve with the decisions of the regulator, who can enforce capacity requirements if he forecasts gaps between electricity supply and demand. These co-evolutionary dynamics create path dependencies through virtuous or vicious feedback loops. A successful transition can only be achieved if the policy mix consists of renewable energy subsidies and high carbon prices that lead to successful investments in renewables, which are reinforced by learning and imitation.

The results of the literature reviewed above strongly suggest the need to analyse the economy at the firm level in order to assess mitigation pathways, identify levers and bottlenecks, and more accurately quantify the risks and impacts of climate shocks, as well as other risks associated with the transition to a sustainable economy (i.e., financial and policy risks). However, as *Paper 2* and *Paper 3* show, data availability severely limits our ability to have an empirically ground model at this level of granularity that can inform policy-making.

Chapter 4

Conclusions

The central theme of this thesis is a fundamental inquiry into advancing IAMs. It does so by taking a holistic and a granular approach. Understanding and finding solutions to climate and ecological challenges requires a holistic approach and a systems thinking mindset due to their interconnected nature (Lade et al., 2020; Anderies et al., 2013; Heitzig et al., 2016). Nevertheless, these two challenges are mostly treated in isolation, not only in IAMs but also (unsurprisingly) in society at large (as there is some reflexivity between the two). As we have seen in Section 3.2.1, while some IAMs have a fairly comprehensive and integrated approach, they rarely – if ever – use it to generate climate mitigation scenarios that would also address ecological issues. This leaves an enormous gap in the literature that has very direct repercussions on society, especially through the IPCC reports.

Paper 1 put climate mitigation scenarios, developed by IAMs, to a test. Using the mitigation scenarios considered in the recent IPCC AR6, *Paper 1* assessed the efficacy of Paris-compliant mitigation pathways in addressing the ecological crisis using the PBs framework. It showed that IAMs fall short in accounting for ecological problems (as also discussed in more depth in Section 3.2.1) and that only a handful of mitigation scenarios stay within PBs. Even scenarios that are developed from storylines depicting sustainable futures, seen as paths that we should embark on, do not meet the ecological feasibility criteria put forward in the PBs framework. While these scenarios rein in GHG emissions, we would likely still overshoot other PBs, potentially tipping the Earth system into an undesirable stable state with irreversible consequences on human timescales. These conclusions are reinforced by the fact that IAMs do not consider the life-cycle environmental impacts of the different solutions (Section 3.2.1). Therefore, inquiring how to advance society-Earth system interactions in IAMs so that ecological issues are addressed is paramount. This thesis examined IAMs to assess how they could be advanced to better integrate relevant ecological dimensions. Section 4.1 suggests avenues for future research.

The granular approach taken in this thesis is primarily based on complex systems science and complexity economics, but it also draws insights from SCM and recent discoveries in macroeconomics. It focused on the interactions between firms in the production system. Section 3.1 and 3.2.2 showed that the macroeconomic models used in IAMs are disproportionately developed with a reductionist approach and lack granularity. Feedback mechanisms, non-linearities, interactions, learning and adaptation give rise to aggregate outcomes that defy straightforward deduction through the isolated study and mere aggregation of the system's components. Our ability to understand the evolution of the economy is further contingent upon the level of granularity at which we analyse the system. As Section 3.2.2 showed, the correct level of analysis is the firm level. Choosing an excessively aggregated level of analysis, and neglecting complex features of the production system and other mechanisms at the firm level prevent us from, e.g., identifying bottlenecks and leverage points in the system, and from understanding the resilience of the economy – and thus of society as a whole.

Calls for new approaches to economics are nothing new (e.g., Anderson et al., 1988; Farmer and Geanakoplos, 2009; Farmer and Foley, 2009; Vines and Wills, 2018). At present, data limitations prevent us from taking this agenda forward and building a more realistic model that is better suited to address the challenges posed by the climate and ecological crises. *Paper 2* and *Paper 3* explored the data gap by focusing on the production network and taking two different approaches.

Paper 2 addressed the lack of data by providing the first comprehensive picture of the most fundamental statistics on production networks. Our results show that there are some properties of these networks that we can consider "really known", thus providing a basis for generating synthetic data or calibrating macroeconomic models. However, there are several limitations to this work, so it is only the first step of an important research agenda. Subsequent research should look at more sophisticated properties, for example using the industrial classification and geographical location of firms.

Paper 3 took a different approach to overcome the data limitations that prevent us from modelling the economy from the bottom up. It applied methods from statistical physics to reconstruct missing data in production networks. It focused on reconstructing the transaction values when the binary topology is partially known, using aggregate information about firms available in their financial statements. It showed that some properties can be reconstructed fairly well even when many firms and supply chain relationships are missing. It was also found that the quality of the reconstruction depends on the network structure and the sampling process of firms and links.

Arguably, our test network still represents a somewhat more favourable case than data on global supply chains since when we reconstructed the production network of globally listed firms, results showed that much more work needs to be done. The main problem is reconciling national accounts with firms' financial statements; as *Paper 2* also suggests, more work is needed on this front. Similarly to *Paper 2*, *Paper 3* is also only a first step in what is an emerging and important research agenda. Another necessary line of future research is to test the quality of the reconstruction method using ABMs (a step in this direction was taken by Reisch et al., 2021), where transaction values are likely to play a more influential role in the model's outcomes than in general equilibrium models. It is also crucial to attempt to reconstruct firms' environmental impacts, which would be a key ingredient of the new, granular model that I (and others) envision. This thesis did not address this data gap but it is touched upon in Section 4.1, where I propose future research directions for advancing IAMs.

Section 4.1 weaves together the findings of this thesis to draw insights for advancing IAMs. It does so by identifying challenges and opportunities for future research.

4.1 Challenges and opportunities for IAMs

There are two overarching avenues for future research: one is to refine existing models, while the other is to develop entirely new ones. These directions are not mutually exclusive and can both bring their insights. The former, mainly offering opportunities to improve IAMs, should not pose particular challenges, although some may still exist, especially for optimisation models. The pursuit of the novel model envisaged in this thesis, however, represents a completely different endeavour, markedly representing a challenge for IAMs. Therefore, I organise the discussion of future research into challenges, i.e., build a new model, and opportunities, i.e., refine existing models.

In summary, the ideal model envisioned in this thesis would

A. capture features of complex systems (see Section 1 and 3.2.2, but also Hafner et al., 2020, provide an excellent list)

B. be at the firm level

C. account for natural resource use, and life-cycle emissions and environmental impacts

D. be calibrated using empirical data

E. account for shocks (e.g., extreme weather events or other natural disasters); and

F. constrain the scenarios to meet climate and ecological goals.

Challenges: build new models. As discussed throughout the thesis the biggest challenge that prevents us from having the envisioned model is data availability on the production network, firms' emissions and other environmental impacts, so the discussion centres around this. Section 3.2.2 provides many references that develop macroeconomic models under a complexity economic lens from which interested readers can draw upon; these should be integrated with insights from hybrid LCA and industrial ecology models (e.g., Hertwich et al., 2015; Luderer et al., 2019; Pauliuk et al., 2017).

Some datasets on firms' emissions are available, although they provide only a partial picture. For example, the Carbon Disclosure Project and the EU Emissions Trading Scheme track the emissions of some companies, while the US Environmental Protection Agency for power plants. The Carbon Disclosure Project has also information on biodiversity, water use and deforestation. So it would be a matter of collecting, collating and reconciling all the data. These data could then be merged with data on global supply chains. Of course, the final dataset, which joins together all these datasets, would be riddled with missing data.

How useful this endeavour would be is not entirely clear, but the results of *Paper 3* raise some scepticism and suggest that more work needs to be done on reconstructing the missing data before these can be of any practical use. As already discussed, these would entail reconstructing missing supply-chain relationships, the value of the transactions, and firms' emissions and environmental impacts. There has been some recent work on predicting supply chains that looks promising (Mungo et al., 2022; Brintrup et al., 2018; Kosasih and Brintrup, 2022), but predicting money flows, as we have seen, is still an under-researched area that requires much more work, as is the reconstruction of firms' emissions and environmental impacts (some work on firms' emissions has been done by Serafeim and Velez Caicedo, 2022; Huidobro et al., 2023).

Most importantly, it seems unlikely that the endeavours discussed above would yield fruits within the timescales required to solve the climate crisis (we need to turn the tide this decade, IPCC, 2023). Two possible routes lie ahead: (1) settle for a model at the sector level that uses the most fine-grained datasets and (2) restrict the analysis at the country level, where, as we saw in *Paper 2*, there are more data on production networks. (1) is discussed below, where I discuss opportunities for IAMs. Regarding (2), Stangl et al. (2023) take a step in this direction by merging VAT data for Hungarian firms (the dataset we also used in *Paper 2*) with emission data from the EU Emission Trading Scheme. Other countries that have access to datasets similar to those that Stangl et al. (2023) use, could do something analogous.

If emission or other environmental indicators are not available, but the country collects data on firms' supply chains, supply chain risk could still be assessed in a vacuum, i.e., without assessing emissions and environmental impacts. For example, they could assess risks related to shocks, such as natural disasters, to think about how to avoid or minimise supply chain disruptions, and assess the resilience of the supply chains. This analysis would inform about possible measures to manage and mitigate risks, making the supply chains more resilient. Then these countries could pursue option (1) to address climate and ecological issues, which I examine in what follows after concluding this part on challenges with a discussion on introducing shocks into the model.

Regardless of data availability and whether the analysis is at the global or country level, the model should simulate shocks. However, there are significant challenges in simulating the impacts of future climate shocks. As the climate continues to change, past weather data tell us very little about the future (Pindyck, 2013). To address this, a scenario analysis, which simulates a range of climate shocks and impacts can provide insight into worst-case outcomes. This information can then be used to develop contingency plans that can help manage risks, minimise potential damage and ensure essential needs are met

Opportunities: improve current IAMs. Improving current models implies that one needs to give up requirement (B) and settle for a model at the sector level that uses the most fine-grained datasets that are available. Before exploring this route, there are still some other options that are easier to implement in current models. Opportunities lie for both optimisation and simulation IAMs (Section 3.1.2), but optimisation models suffer a more fundamental flaw that cannot be overcomed until a new modelling technique is employed: optimisation models will never be able to satisfy requirement (A) as they stand. Nevertheless, some upgrades could still be done to optimisation IAMs that would allow them to meet requirements (F), (C) and (E). To meet requirement (A), agentisation (Guerrero and Axtell, 2011) could be a route to follow for optimisation IAMs, but it would entail substantial effort as it basically implies a complete remake of the model.

A low-hanging fruit for both optimisation and simulation IAMs that already have a comprehensive treatment of the PBs indicators is to generate pathways that stay within PBs. PBs that are essential to account for are climate change, nitrogen and phosphorous use, land-system change, novel entities and freshwater use. For IAMs that do not have such a comprehensive treatment but already have enough level of detail, it may be relatively easy to extend their model to include (other) PBs indicators and generate pathways that meet both climate and ecological goals. These enhancements, do not necessarily imply that the resulting model would meet all the required features, but it is a first step that can bring it insights.

Models should include natural resource use (other than those considered in the PBs framework, but still fundamental, e.g., mineral use), and life-cycle emissions and environmental impacts, since these cannot ignored, as highlighted in Section 3.2.1. Another low-hanging fruit in this regard, which might be feasible only for some models, is to include LCA coefficients directly in the model, as proposed by Arvesen et al. (2018) and applied to life-cycle emissions in Pehl et al. (2017).

A more involved approach to incorporate these dimensions (requirement (C)) would entail, from a modelling perspective, integrating insights from hybrid LCA and industrial ecology models (e.g., Hertwich et al., 2015; Luderer et al., 2019; Pauliuk et al., 2017) with insights from complexity economics (see Section 3.2.2). Implementing this approach would require modifying IAMs, especially those that do not use a macroeconomic input-output model. This modification should not pose particular challenges apart from increasing the model's complexity, which should be sufficiently constrained by using sector-level input-output tables. From a data perspective, including natural resource use and life-cycle impacts would involeve collating and reconciling LCA data with environmentally extended, multi-country input-output tables at the sector level (as done, for example, in Hertwich and Wood, 2018, other methods could be those reviewed in Crawford et al., 2018). Some interesting datasets are GaBi, for LCA data, and Exiobase and the IELab, for sector-level data.

Most IAMs do not take shocks into account (Section 3.1.1), so this is also another required feature to include. A potential avenue to address this issue was previously explored when discussing challenges for IAMs.

Bibliography

Acemoglu, D. and Azar, P. D. (2020). Endogenous production networks. *Econometrica*, 88(1):33–82.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *nature*, 406(6794):378–382.

Allen, C., Metternicht, G., Wiedmann, T., and Pedercini, M. (2019). Greater gains for Australia by tackling all SDGs but the last steps will be the most challenging. *Nature Sustainability*, 2(11):1041–1050. Number: 11 Publisher: Nature Publishing Group.

Allen, C., Metternicht, G., Wiedmann, T., and Pedercini, M. (2021). Modelling national transformations to achieve the SDGs within planetary boundaries in small island developing states. *Global Sustainability*, 4. Publisher: Cambridge University Press.

Allen, M., Andronova, N., Booth, B., Dessai, S., Frame, D., Forest, C., Gregory, J., Hegerl, G., Knutti, R., Piani, C., Sexton, D., and Stainforth, D. (2006). Observational constraints on climate sensitivity. In *Avoiding dangerous climate change*, pages 281–290. Cambridge University Press.

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., and Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458(7242):1163–1166. Number: 7242 Publisher: Nature Publishing Group.

Anandarajah, G., Pye, S., Usher, W., Kesicki, F., and Mcglade, C. (2011). TIAM-UCL global model documentation. Working Paper REF UKERC/WP/ESY/2011/001, UK ERC, UK Energy Research Center. Available at https://www.researchgate.net/publication/239917973_TIAM-UCL_Global_Model_Documentation [Accessed on 6 April 2023]. Documentation of the model can also be found at https://www.ucl.ac.uk/energy-models/models/tiam-ucl.

Anderies, J. M., Carpenter, S. R., Steffen, W., and Rockström, J. (2013). The topology of non-linear global carbon dynamics: from tipping points to planetary boundaries. *Environmental Research Letters*, 8(4):044048. Publisher: IOP Publishing.

Anderson, K. and Jewell, J. (2019). Debating the bedrock of climate-change mitigation scenarios. *Nature*, 573(7774).

Anderson, K. and Peters, G. (2016). The trouble with negative emissions. *Science*, 354(6309):182–183.

Anderson, P. W. (1972). More Is Different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396.

Anderson, P. W., Arrow, K., and Pines, D. (1988). *The Economy As An Evolving Complex System*. Addison-Wesley, Redwood City, California.

Anthoff, D. and Tol, R. S. (2013). The uncertainty about the social cost of carbon: A decomposition analysis using fund. *Climatic change*, 117:515–530.

Arthur, W. B. (2014). *Complexity economics*. Oxford University Press.

Arthur, W. B. (2021). Foundations of complexity economics. *Nature Reviews Physics*, 3(2):136–145.

Arvesen, A., Bright, R. M., and Hertwich, E. G. (2011). Considering only first-order effects? How simplifications lead to unrealistic technology optimism in climate change mitigation. *Energy Policy*, 39(11):7448–7454.

Arvesen, A., Luderer, G., Pehl, M., Bodirsky, B. L., and Hertwich, E. G. (2018). Deriving life cycle assessment coefficients for application in integrated assessment modelling. *Environmental Modelling & Software*, 99:111–125.

Ashwin, P., Wieczorek, S., Vitolo, R., and Cox, P. (2012). Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Publisher: The Royal Society Publishing.

Axtell, R. (2014). Pathologies of 'Integrated Assessments' of Climate Change. Available at: https://css1.gmu.edu/~axtell/Rob/Research/Pages/Environment_files/Axtell%20on%20IAMs.pdf [Accessed 7 April 2023].

Barazza, E. and Strachan, N. (2020). The co-evolution of climate policy and investments in electricity markets: Simulating agent dynamics in UK, German and Italian electricity sectors. *Energy Research & Social Science*, 65:101458.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., et al. (2011). Has the earth's sixth mass extinction already arrived? *Nature*, 471(7336):51–57.

Barrot, J.-N. and Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592.

Beinhocker, E. D., Farmer, J. D., and Hepburn, C. (2013). Next generation economy, energy and climate modeling. *Global Commission on Economy and Climate*, 11.

Beringer, T., Lucht, W., and Schaphoff, S. (2011). Bioenergy production potential of global biomass plantations under environmental and agricultural constraints. *GCB Bioenergy*, 3(4):299–312. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1757-1707.2010.01088.x.

Biermann, F. and Kim, R. E. (2020). The Boundaries of the Planetary Boundary Framework: A Critical Appraisal of Approaches to Define a "Safe Operating Space" for Humanity. *Annual Review of Environment and Resources*, 45(1):497–521.

Bonart, J., Bouchaud, J.-P., Landier, A., and Thesmar, D. (2014). Instabilities in large economies: aggregate volatility without idiosyncratic shocks. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(10):P10040. Publisher: IOP Publishing and SISSA.

Bonsch, M., Humpenöder, F., Popp, A., Bodirsky, B., Dietrich, J. P., Rolinski, S., Biewald, A., Lotze-Campen, H., Weindl, I., Gerten, D., and Stevanovic, M. (2016). Trade-offs between land and water requirements for large-scale bioenergy production. *GCB Bioenergy*, 8(1):11–24. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcbb.12226.

Bowles, N., Alexander, S., and Hadjikakou, M. (2019). The livestock sector and planetary boundaries: A 'limits to growth' perspective with dietary implications. *Ecological Economics*, 160:128–136.

Brintrup, A., Barros, J., and Tiwari, A. (2015a). The nested structure of emergent supply networks. *IEEE Systems Journal*, 12(2):1803–1812.

Brintrup, A., Kito, T., López, E., New, S., and Reed, F. (2011a). The structure of the Toyota supply network: The emergence of resilience. CADByN Working Paper [2011-05-012], University of Oxford, Oxford. https://www.cabdyn.ox.ac.uk/complexity_PDFs/Working%20Papers%202011/ToyotaRobustness.pdf.

Brintrup, A., Kito, T., New, S., and Reed-Tsochas, F. (2011b). From transaction cost economics to food webs: A multi-disciplinary discussion on the length of supply chains: 18th International Annual EurOMA Conference. *Proceedings of the 18th International Annual EurOMA Conference*. Publisher: European Operations Management Association (EurOMA).

Brintrup, A., Ledwoch, A., and Barros, J. (2015b). Topological robustness of the global automotive industry. *Logistics Research*, 9(1):1.

Brintrup, A., Wang, Y., and Tiwari, A. (2017). Supply Networks as Complex Systems: A Network-Science-Based Characterization. *IEEE Systems Journal*, 11(4):2170–2181. Conference Name: IEEE Systems Journal.

Brintrup, A., Wichmann, P., Woodall, P., McFarlane, D., Nicks, E., and Krechel, W. (2018). Predicting hidden links in supply networks. *Complexity*, 2018:1–12.

Budzinski, M., Wood, R., Zakeri, B., Krey, V., and Strømman, A. H. (2023). Coupling energy system models with multi-regional input-output models based on the make and use framework–insights from MESSAGEix and EXIOBASE. *Economic Systems Research*, pages 1–19.

Cai, Y., Judd, K. L., and Lontzek, T. S. (2012). DSICE: A Dynamic Stochastic Integrated Model of Climate and Economy.

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., Waldhoff, S., and Wise, M. (2019). GCAM v5.1: representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, 12(2):677–698. Publisher: Copernicus GmbH.

Campbell, B. M., Beare, D. J., Bennett, E. M., Hall-Spencer, J. M., Ingram, J. S. I., Jaramillo, F., Ortiz, R., Ramankutty, N., Sayer, J. A., and Shindell, D. (2017). Agriculture production as a major driver of the Earth system exceeding planetary boundaries. *Ecology and Society*, 22(4):art8.

Capellán-Pérez, I., de Blas, I., Nieto, J., de Castro, C., Miguel, L. J., Carpintero, O., Mediavilla, M., Lobejón, L. F., Ferreras-Alonso, N., and Rodrigo, P. (2020). MEDEAS: A new modeling framework integrating global biophysical and socioeconomic constraints. *Energy & environmental science*, 13(3):986–1017. Publisher: Royal Society of Chemistry.

Carnovale, S. and Yeniyurt, S. (2015). The Role of Ego Network Structure in Facilitating Ego Network Innovations. *Journal of Supply Chain Management*, 51(2):22–46. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jscm.12075.

Carvalho, V. and Draca, M. (2018). Cascading innovation. 2018 Meeting Papers, No 1322, Society for Economic Dynamics.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., and Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science advances*, 1(5):e1400253.

Choi, T. Y., Dooley, K. J., and Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: control versus emergence. *Journal of Operations Management*, 19(3):351–366.

Choi, T. Y. and Hong, Y. (2002). Unveiling the structure of supply networks: case studies in Honda, Acura, and DaimlerChrysler. *Journal of Operations Management*, 20(5):469–493.

Choi, T. Y. and Kim, Y. (2008). Structural Embeddedness and Supplier Management: A Network Perspective*. *Journal of Supply Chain Management*, 44(4):5–13. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-493X.2008.00069.x.

Christopher, M. and Peck, H. (2004). Building the Resilient Supply Chain. *International Journal of Logistics Management*, 15:1–13.

Colon, C., Hallegatte, S., and Rozenberg, J. (2021). Criticality analysis of a country's transport network via an agent-based supply chain model. *Nature Sustainability*, 4(3):209–215.

Conijn, J., Bindraban, P., Schröder, J., and Jongschaap, R. (2018). Can our global food system meet food demand within planetary boundaries? *Agriculture, Ecosystems & Environment*, 251:244–256.

Contreras, M. G. A. and Fagiolo, G. (2014). Propagation of economic shocks in input-output networks: A cross-country analysis. *Physical Review E*, 90(6):062812.

Crawford, R. H., Bontinck, P.-A., Stephan, A., Wiedmann, T., and Yu, M. (2018). Hybrid life cycle inventory methods – A review. *Journal of Cleaner Production*, 172:1273–1288.

Crutzen, P. J. (2002). Geology of mankind. *Nature*, 415(6867):23–23.

Daly, H. E., Scott, K., Strachan, N., and Barrett, J. (2015). Indirect CO_2 Emission Implications of Energy System Pathways: Linking IO and TIMES Models for the UK. *Environmental Science & Technology*, 49(17):10701–10709.

Di Giovanni, J. and Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations in granular economies. *Journal of Political Economy*, 120(6):1083–1132.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2022). Quantifying firm-level economic systemic risk from nation-wide supply networks. *Scientific Reports*, 12(1):7719.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2023). Estimating the loss of economic predictability from aggregating firm-level production networks. arXiv preprint arXiv:2302.11451.

Dorninger, C., Hornborg, A., Abson, D. J., Von Wehrden, H., Schaffartzik, A., Giljum, S., Engler, J.-O., Feller, R. L., Hubacek, K., and Wieland, H. (2021). Global patterns of ecologically unequal exchange: Implications for sustainability in the 21st century. *Ecological economics*, 179:106824. Publisher: Elsevier.

Downing, A. S., Bhowmik, A., Collste, D., Cornell, S. E., Donges, J., Fetzer, I., Häyhä, T., Hinton, J., Lade, S., and Mooij, W. M. (2019). Matching scope, purpose and uses of planetary boundaries science. *Environmental Research Letters*, 14(7):073005.

Dwesar, I., Kömüves, Z., McGovern, M., Vu, A., Arsenio, F., Heald, S., Chewpreecha, U., and Pollitt, H. (2022). E3ME Model Manual v9. 0. Cambridge Econometrics, Cambridge, UK.

D'Alessandro, S., Cieplinski, A., Distefano, T., and Dittmer, K. (2020). Feasible alternatives to green growth. *Nature Sustainability*, 3(4):329–335.

Emmerling, J., Drouet, L., Reis, L., Bevione, M., Berger, L., Bosetti, V., Carrara, S., De Cian, E., de Maere d'Aertrycke, G., Longden, T., Malpede, M., Marangoni, G., Sferra, F., Tavoni, M., Witajewski-Baltvilks, J., and Havlík, P. (2016). The WITCH 2016 Model - Documentation and Implementation of the Shared Socioeconomic Pathways.

Fanning, A. L., O'Neill, D. W., Hickel, J., and Roux, N. (2021). The social shortfall and ecological overshoot of nations. *Nature Sustainability*, pages 1–11.

Farmer, J. D. (draft). *A Revolution in Economics: How complex-systems science can help solve our biggest problems*. Penguin.

Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.

Farmer, J. D. and Geanakoplos, J. (2009). The virtues and vices of equilibrium and the future of financial economics. *Complexity*, 14(3):11–38.

Farmer, J. D., Hepburn, C., Mealy, P., and Teytelboym, A. (2015). A third wave in the economics of climate change. *Environmental and Resource Economics*, 62(2):329–357.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., and Snyder, P. K. (2005). Global Consequences of Land Use. *Science*, 309(5734):570–574. Publisher: American Association for the Advancement of Science.

Forrester, J. W. (1961). *Industrial Dynamics*. MIT Press, Cambridge, Massachusetts.

Frame, D. J., Macey, A. H., and Allen, M. R. (2014). Cumulative emissions and climate policy. *Nature Geoscience*, 7(10):692–693. Number: 10 Publisher: Nature Publishing Group.

Fricko, O., Parkinson, S. C., Johnson, N., Strubegger, M., Vliet, M. T. v., and Riahi, K. (2016). Energy sector water use implications of a 2°C climate policy. *Environmental Research Letters*, 11(3):034011. Publisher: IOP Publishing.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, 14(11):4811–4900.

Fujimori, S., Hasegawa, T., Masui, T., and Takahashi, K. (2014). Land use representation in a global CGE model for long-term simulation: CET vs. logit functions. *Food Security*, 6(5):685–699.

Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Gatti Domingues, L., Cassol, H. L. G., Tejada, G., Aragão, L. E. O. C., Nobre, C., Peters, W., Marani, L., Arai, E., Sanches, A. H., Corrêa, S. M., Anderson, L., Von Randow, C., Correia, C. S. C., Crispim, S. P., and Neves, R. A. L. (2021). Amazonia as a carbon source linked to deforestation and climate change. *Nature*, 595(7867):388–393. Number: 7867 Publisher: Nature Publishing Group.

Gerst, M. D., Wang, P., Roventini, A., Fagiolo, G., Dosi, G., Howarth, R. B., and Borsuk, M. E. (2013). Agent-based modeling of climate policy: An introduction to the ENGAGE multi-level model framework. *Environmental Modelling & Software*, 44:62–75.

Gerten, D., Hoff, H., Rockström, J., Jägermeyr, J., Kummu, M., and Pastor, A. V. (2013). Towards a revised planetary boundary for consumptive freshwater use: role of environmental flow requirements. *Current Opinion in Environmental Sustainability*, 5(6):551–558.

Golosov, M., Hassler, J., Krusell, P., and Tsyvinski, A. (2014). Optimal Taxes on Fossil Fuel in General Equilibrium. *Econometrica*, 82(1).

Guerrero, O. A. and Axtell, R. L. (2011). Using Agentization for Exploring Firm and Labor Dynamics. In Osinga, S., Hofstede, G. J., and Verwaart, T., editors, *Emergent Results of Artificial Economics*, Lecture Notes in Economics and Mathematical Systems, pages 139–150. Springer, Berlin, Heidelberg.

Guterres, A. (2022a). Press conference launch of IPCC AR6 report by WG II [Speech transcript]. Available at https://media.un.org/en/asset/k1x/k1xcijxjhp [Accessed on 3 April 2023].

Guterres, A. (2022b). Secretary-General's address to the 76th Session of the UN General Assembly [Speech transcript]. Available at https://www.un.org/sg/en/content/sg/speeches/2021-09-21/address-the-76th-session-of-general-assembly [Accessed on 18 April 2023].

Hafner, S., Anger-Kraavi, A., Monasterolo, I., and Jones, A. (2020). Emergence of New Economics Energy Transition Models: A Review. *Ecological Economics*, 177:106779.

Harfoot, M., Tittensor, D. P., Newbold, T., McInerny, G., Smith, M. J., and Scharlemann, J. P. W. (2014). Integrated assessment models for ecologists: the present and the future. *Global Ecology and Biogeography*, 23(2):124–143.

Hearnshaw, E. and Wilson, M. (2013). A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33.

Heck, V., Donges, J. F., and Lucht, W. (2016). Collateral transgression of planetary boundaries due to climate engineering by terrestrial carbon dioxide removal. *Earth System Dynamics*, 7(4):783–796. Publisher: Copernicus GmbH.

Heck, V., Gerten, D., Lucht, W., and Popp, A. (2018a). Biomass-based negative emissions difficult to reconcile with planetary boundaries. *Nature Climate Change*, 8(2):151–155. Number: 2 Publisher: Nature Publishing Group.

Heck, V., Gerten, D., Lucht, W., and Popp, A. (2018b). Biomass-based negative emissions difficult to reconcile with planetary boundaries. *Nature climate change*, 8(2):151–155.

Heitzig, J., Kittel, T., Donges, J. F., and Molkenthin, N. (2016). Topology of sustainable management of dynamical systems with desirable states: from defining planetary boundaries to safe operating spaces in the Earth system. *Earth System Dynamics*, 7(1):21–50. Publisher: Copernicus GmbH.

Henriet, F., Hallegatte, S., and Tabourier, L. (2012). Firm-network characteristics and economic robustness to natural disasters. *Journal of Economic Dynamics and Control*, 36(1):150–167.

Hertwich, E. G., Gibon, T., Bouman, E. A., Arvesen, A., Suh, S., Heath, G. A., Bergesen, J. D., Ramirez, A., Vega, M. I., and Shi, L. (2015). Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. *Proceedings of the National Academy of Sciences*, 112(20):6277–6282.

Hertwich, E. G. and Wood, R. (2018). The growing importance of scope 3 greenhouse gas emissions from industry. *Environmental Research Letters*, 13(10):104013.

Hickel, J., Brockway, P., Kallis, G., Keyßer, L., Lenzen, M., Slameršak, A., Steinberger, J., and Ürge-Vorsatz, D. (2021). Urgent need for post-growth climate mitigation scenarios. *Nature Energy*, 6(8):766–768.

Hoekstra, A. Y. and Wiedmann, T. O. (2014). Humanity's unsustainable environmental footprint. *Science*, 344(6188):1114–1117. Publisher: American Association for the Advancement of Science.

Hope, C. and Hope, M. (2013). The social cost of CO_2 in a low-growth world. *Nature Climate Change*, 3(8):722–724.

Hourcade, J.-C., Jaccard, M., Bataille, C., and Ghersi, F. (2006). Hybrid modeling: New answers to old challenges introduction to the special issue of" the energy journal". *The Energy Journal*, pages 1–11.

Hughes, B. B. and Hedden, S. (2016). IFs Agriculture Model Documentation. Working Paper Research 2016.03.21, Pardee Center for International Futures, Josef Korbel School of International Studies, University of Denver, Denver, CO.

Huidobro, J. O., Tamarít, I., and Lipari, F. (2023). Data quality in the spotlight: a Hybrid-LCA approach to evaluating reported corporate carbon footprints. PREPRINT (Version 1) available at Research Square https://doi.org/10.21203/rs.3.rs-2672722/v1.

Ialongo, L. N., de Valk, C., Marchese, E., Jansen, F., Zmarrou, H., Squartini, T., and Garlaschelli, D. (2022). Reconstructing firm-level interactions in the dutch input–output network from production constraints. *Scientific reports*, 12(1):1–12.

IEA (2021). Net Zero by 2050: A Roadmap for the Global Energy Sector. Special Report, International Energy Agency, Paris. License: CC BY 4.0. Available at: https://www.iea.org/reports/net-zero-by-2050.

Inoue, H. and Todo, Y. (2019). Firm-level propagation of shocks through supply-chain networks. *Nature Sustainability*, 2(9):841–847.

IPBES (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages. https://doi.org/10.5281/zenodo.3831673.

IPCC (2018). Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland, 32 pp.

IPCC (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2391 pp., doi: 10.1017/9781009157896.

IPCC (2022a). Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi: 10.1017/9781009325844.

IPCC (2022b). Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926.

IPCC (2023). Summary for Policymakers. In: Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, (in press).

Jackson, T. and Victor, P. A. (2019). Unraveling the claims for (and against) green growth. *Science*, 366(6468):950–951.

Junne, T., Simon, S., Buchgeister, J., Saiger, M., Baumann, M., Haase, M., Wulf, C., and Naegler, T. (2020). Environmental Sustainability Assessment of Multi-Sectoral Energy Transformation Pathways: Methodological Approach and Case Study for Germany. *Sustainability*, 12(19):8225.

Kanitkar, T., Mythri, A., and Jayaraman, T. (2022). Equity Assessment of Global Mitigation Pathways in the IPCC Sixth Assessment Report. Available at https://doi.org/10.31219/osf.io/p46ty.

Keppo, I., Butnar, I., Bauer, N., Caspani, M., Edelenbosch, O., Emmerling, J., Fragkos, P., Guivarch, C., Harmsen, M., Lefèvre, J., Gallic, T. L., Leimbach, M., McDowall, W., Mercure, J.-F., Schaeffer, R., Trutnevyte, E., and Wagner, F. (2021). Exploring the possibility space: taking stock of the diverse capabilities and gaps in integrated assessment models. *Environmental Research Letters*, 16(5):053006. Publisher: IOP Publishing.

Keramidas, K., Kitous, A., Schade, B., Després, J., and Schmitz, A. (2018). POLES-JRC model documentation: 2018 update. European Commission and Joint Research Centre. Luxembourg: Publications Office of the European Union. Available at https://op.europa.eu/en/publication-detail/-/publication/2a03544b-f1f6-11e8-9982-01aa75ed71a1/language-en Accessed on 8 April 2023.

Kim, Y., Choi, T. Y., Yan, T., and Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, 29(3):194–211.

Kito, T., Brintrup, A., New, S., and Reed-Tsochas, F. (2014). The Structure of the Toyota Supply Network: An Empirical Analysis. Saïd Business School WP 2014-3, Available at SSRN: https://ssrn.com/abstract=2412512 or http://dx.doi.org/10.2139/ssrn.2412512.

König, M. D., Levchenko, A., Rogers, T., and Zilibotti, F. (2022). Aggregate fluctuations in adaptive production networks. *Proceedings of the National Academy of Sciences*, 119(38):e2203730119. Publisher: Proceedings of the National Academy of Sciences.

Kosasih, E. E. and Brintrup, A. (2022). A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research*, 60(17):5380–5393.

Kouloumpis, V., Stamford, L., and Azapagic, A. (2015). Decarbonising electricity supply: Is climate change mitigation going to be carried out at the expense of other environmental impacts? *Sustainable Production and Consumption*, 1:1–21.

Krey, V., Havlik, P., Kishimoto, P., Fricko, O., Zilliacus, J., Gidden, M., Strubegger, M., Kartasasmita, G., Ermolieva, T., Forsell, N., et al. (2020). MESSAGEix-GLOBIOM Documentation - 2020 release.

Lade, S. J., Steffen, W., de Vries, W., Carpenter, S. R., Donges, J. F., Gerten, D., Hoff, H., Newbold, T., Richardson, K., and Rockström, J. (2020). Human impacts on planetary boundaries amplified by Earth system interactions. *Nature Sustainability*, 3(2):119–128.

Lahn, B. (2020). A history of the global carbon budget. *WIREs Climate Change*, 11(3):e636. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.636.

Lahn, B. (2021). Changing climate change: The carbon budget and the modifying-work of the IPCC. *Social Studies of Science*, 51(1):3–27. Publisher: SAGE Publications Ltd.

Lamperti, F., Dosi, G., Napoletano, M., Roventini, A., and Sapio, A. (2018). Faraway, So Close: Coupled Climate and Economic Dynamics in an Agent-based Integrated Assessment Model. *Ecological Economics*, 150:315–339.

Lee, H. L., Padmanabhan, V., and Whang, S. (1997). The bullwhip effect in supply chains1. *Sloan Management Review*, 38(3):93–102.

Lee, J., Bazilian, M., Sovacool, B., Hund, K., Jowitt, S. M., Nguyen, T. P., Månberger, A., Kah, M., Greene, S., and Galeazzi, C. (2020). Reviewing the material and metal security of low-carbon energy transitions. *Renewable and Sustainable Energy Reviews*, 124:109789. Publisher: Elsevier.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J. (2008). Tipping elements in the earth's climate system. *Proceedings of the national Academy of Sciences*, 105(6):1786–1793.

Lenton, T. M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., and Schellnhuber, H. J. (2019). Climate tipping points — too risky to bet against. *Nature*, 575(7784).

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., and Geschke, A. (2012). International trade drives biodiversity threats in developing nations. *Nature*, 486(7401):109–112. Number: 7401 Publisher: Nature Publishing Group.

Li, F. G. N. and Strachan, N. (2017). Modelling energy transitions for climate targets under landscape and actor inertia. *Environmental Innovation and Societal Transitions*, 24:106–129.

Liu, J., Mooney, H., Hull, V., Davis, S. J., Gaskell, J., Hertel, T., Lubchenco, J., Seto, K. C., Gleick, P., Kremen, C., and Li, S. (2015). Systems integration for global sustainability. *Science*, 347(6225):1258832. Publisher: American Association for the Advancement of Science.

Lohmann, J., Castellana, D., Ditlevsen, P. D., and Dijkstra, H. A. (2021). Abrupt climate change as a rate-dependent cascading tipping point. *Earth System Dynamics*, 12(3):819–835.

Lohmann, J. and Ditlevsen, P. D. (2021). Risk of tipping the overturning circulation due to increasing rates of ice melt. *Proceedings of the National Academy of Sciences*, 118(9):e2017989118.

Lontzek, T. S., Cai, Y., Judd, K. L., and Lenton, T. M. (2015). Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy. *Nature Climate Change*, 5(5).

Loulou, R. (2004). Documentation for the MARKAL Family of Models. Energy Technology Systems Analysis Programme (ETSAP), International Energy Agency (IEA), Paris, France. Available at https://unfccc.int/resource/cd_roms/na1/mitigation/Module_5/Module_5_1/b_tools/MARKAL/MARKAL_Manual.pdf.

Luderer, G., Gunnar Luderer, Gunnar Luderer, Luderer, G., Pehl, M., Arvesen, A., Gibon, T., Bodirsky, B. L., de Boer, H. S., Fricko, O., Hejazi, M., Humpenöder, F., Iyer, G., Mima, S., Silvana Mima, Mouratiadou, I., Pietzcker, R. C., Popp, A., van den Berg, M. P., van den Berg, M., van Vuuren, D. P., and Hertwich, E. G. (2019). Environmental co-benefits and adverse side-effects of alternative power sector decarbonization strategies. *Nature Communications*, 10(1):5229–5229. MAG ID: 2990509501.

Luderer, G., Leimbach, M., Bauer, N., Kriegler, E., Baumstark, L., Bertram, C., Giannousakis, A., Hilaire, J., Klein, D., Levesque, A., Mouratiadou, I., Pehl, M., Pietzcker, R., Piontek, F., Roming, N., Schultes, A., Schwanitz, V. J., and Strefler, J. (2015). Description of the REMIND Model (Version 1.6).

Magerman, G., De Bruyne, K., Dhyne, E., and Van Hove, J. (2016). Heterogeneous firms and the micro origins of aggregate fluctuations. NBB Working Paper No. 312, National Bank of Belgium, Brussels.

Mandel, A., Landini, S., Gallegati, M., and Gintis, H. (2015). Price dynamics, financial fragility and aggregate volatility. *Journal of Economic Dynamics and Control*, 51:257–277.

Matthews, H. D. and Caldeira, K. (2008). Stabilizing climate requires near-zero emissions. *Geophysical Research Letters*, 35(4). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL032388.

McKay, D. A., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton, T. M. (2021). Updated assessment suggests > 1.5 C global warming could trigger multiple climate tipping points. Archive Location: world Publisher: Earth and Space Science Open Archive Section: Climatology (Global Change).

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V. M., and Farmer, J. D. (2022). How production networks amplify economic growth. *Proceedings of the National Academy of Sciences*, 119(1):e2106031118.

Meadows, D. H. (2008). *Thinking in systems:A primer*. Chelsea Green Publishing, White River Junction, VT, US.

Meadows, D. H., Meadows, D. L., Randers, J., and Behrens, W. (1972). *The limits to growth: a report for the club of Rome's project on the predicament of mankind*. Universe Books.

Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A., Buja, L. E., Strand, W. G., and Teng, H. (2005). How Much More Global Warming and Sea

Level Rise? *Science*, 307(5716):1769–1772. Publisher: American Association for the Advancement of Science.

Mercure, J.-F., Pollitt, H., Bassi, A. M., Viñuales, J. E., and Edwards, N. R. (2016). Modelling complex systems of heterogeneous agents to better design sustainability transitions policy. *Global Environmental Change*, 37:102–115.

Millar, R., Allen, M., Rogelj, J., and Friedlingstein, P. (2016). The cumulative carbon budget and its implications. *Oxford Review of Economic Policy*, 32(2):323–342.

Miller, J. H. and Page, S. E. (2007). *Complex adaptive systems: an introduction to computational models of social life*. Princeton University Press Princeton (NJ).

Moran, J. and Bouchaud, J.-P. (2019). May's instability in large economies. *Physical Review E*, 100(3):032307. Publisher: American Physical Society.

Mouratiadou, I., Bevione, M., Bijl, D. L., Drouet, L., Hejazi, M., Mima, S., Pehl, M., and Luderer, G. (2018). Water demand for electricity in deep decarbonisation scenarios: a multi-model assessment. *Climatic Change*, 147(1):91–106.

Mouratiadou, I., Biewald, A., Pehl, M., Bonsch, M., Baumstark, L., Klein, D., Popp, A., Luderer, G., and Kriegler, E. (2016). The impact of climate change mitigation on water demand for energy and food: An integrated analysis based on the Shared Socioeconomic Pathways. *Environmental Science & Policy*, 64:48–58.

Mungo, L., Lafond, F., Astudillo-Estévez, P., and Farmer, J. D. (2022). Reconstructing production networks using machine learning. INET Oxford Working Paper No. 2022-02, Institute for New Economic Thinking at the Oxford Martin School. Available at: https://www.inet.ox.ac.uk/publications/no-2022-02-reconstructing-production-networks-using-machine-learning/.

Naegler, T., Buchgeister, J., Hottenroth, H., Simon, S., Tietze, I., Viere, T., and Junne, T. (2022). Life cycle-based environmental impacts of energy system transformation strategies for Germany: Are climate and environmental protection conflicting goals? *Energy Reports*, 8:4763–4775.

Newbold, T., Hudson, L. N., Arnell, A. P., Contu, S., De Palma, A., Ferrier, S., Hill, S. L. L., Hoskins, A. J., Lysenko, I., Phillips, H. R. P., Burton, V. J., Chng, C. W. T., Emerson, S., Gao, D., Pask-Hale, G., Hutton, J., Jung, M., Sanchez-Ortiz, K., Simmons, B. I., Whitmee, S., Zhang, H., Scharlemann, J. P. W., and Purvis, A. (2016). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, 353(6296):288–291. Publisher: American Association for the Advancement of Science.

Nordhaus, W. D. (2010). Economic aspects of global warming in a post-Copenhagen environment. *Proceedings of the National Academy of Sciences*, 107(26):11721–11726. Publisher: Proceedings of the National Academy of Sciences.

Orenstein, P. (2016). How Does Supply Network Evolution and Its Topological Structure Impact Supply Chain Performance? In *2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO)*, pages 562–569.

Orenstein, P. (2021). The changing landscape of supply chain networks: an empirical analysis of topological structure. *INFOR: Information Systems and Operational Research*, 59(1):53–73. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/03155986.2020.1785263.

Otto, C., Willner, S. N., Wenz, L., Frieler, K., and Levermann, A. (2017). Modeling loss-propagation in the global supply network: The dynamic agent-based model acclimate. *Journal of Economic Dynamics and Control*, 83:232–269.

Parisi, F., Squartini, T., and Garlaschelli, D. (2020). A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks. *New Journal of Physics*, 22(5):053053.

Parrique, T., Barth, J., Briens, F., Kuokkanen, A., and Spangenberg, J. (2019). Evidence and arguments against green growth as a sole strategy for sustainability. *European Environmental Bureau*.

Pauliuk, S., Arvesen, A., Stadler, K., and Hertwich, E. G. (2017). Industrial ecology in integrated assessment models. *Nature Climate Change*, 7(1):13–20.

Pedersen, J. T. S., van Vuuren, D., Gupta, J., Santos, F. D., Edmonds, J., and Swart, R. (2022). IPCC emission scenarios: How did critiques affect their quality and relevance 1990–2022? *Global Environmental Change*, 75:102538.

Pehl, M., Arvesen, A., Humpenöder, F., Popp, A., Hertwich, E. G., and Luderer, G. (2017). Understanding future emissions from low-carbon power systems by integration of life-cycle assessment and integrated energy modelling. *Nature Energy*, 2(12):939–945. Number: 12 Publisher: Nature Publishing Group.

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., Bastos Lima, M. G., Baumann, M., Curtis, P. G., De Sy, V., Garrett, R., Godar, J., Goldman, E. D., Hansen, M. C., Heilmayr, R., Herold, M., Kuemmerle, T., Lathuillière, M. J., Ribeiro, V., Tyukavina, A., Weisse, M. J., and West, C. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, 377(6611):eabm9267. Publisher: American Association for the Advancement of Science.

Persson, L., Carney Almroth, B. M., Collins, C. D., Cornell, S., de Wit, C. A., Diamond, M. L., Fantke, P., Hassellöv, M., MacLeod, M., Ryberg, M. W., Søgaard Jørgensen, P., Villarrubia-Gómez, P., Wang, Z., and Hauschild, M. Z. (2022). Outside the Safe Operating Space of the Planetary Boundary for Novel Entities. *Environmental Science & Technology*, 56(3):1510–1521.

Pichler, A. and Farmer, J. D. (2022). Simultaneous supply and demand constraints in input–output networks: the case of Covid-19 in germany, italy, and spain. *Economic Systems Research*, 34(3):273–293.

Pindyck, R. S. (2013). Climate Change Policy: What Do the Models Tell Us? *Journal of Economic Literature*, 51(3):860–872.

Reisch, T., Heiler, G., Diem, C., and Thurner, S. (2021). Inferring supply networks from mobile phone data to estimate the resilience of a national economy. *arXiv preprint arXiv:2110.05625*.

Rockström, J., Steffen, W., Noone, K., Persson, r., Chapin, F. S. I., Lambin, E., Lenton, T., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R., Fabry, V., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. (2009). Planetary Boundaries: Exploring the Safe Operating Space for Humanity. *Ecology and Society*, 14(2). Publisher: The Resilience Alliance.

Rockström, J., Gupta, J., Qin, D., Lade, S. J., Abrams, J. F., Andersen, L. S., Armstrong McKay, D. I., Bai, X., Bala, G., Bunn, S. E., Ciobanu, D., DeClerck, F., Ebi, K., Gifford, L., Gordon, C., Hasan, S., Kanie, N., Lenton, T. M., Loriani, S., Liverman, D. M., Mohamed, A., Nakicenovic, N., Obura, D., Ospina, D., Prodani, K., Rammelt, C., Sakschewski, B., Scholtens, J., Stewart-Koster, B., Tharammal, T., van Vuuren, D., Verburg, P. H., Winkelmann, R., Zimm, C., Bennett, E. M., Bringezu, S., Broadgate, W., Green, P. A., Huang, L., Jacobson, L., Ndehedehe, C., Pedde, S., Rocha, J., Scheffer, M., Schulte-Uebbing, L., de Vries, W., Xiao, C., Xu, C., Xu, X., Zafra-Calvo, N., and Zhang, X. (2023). Safe and just Earth system boundaries. *Nature*, pages 1–10. Publisher: Nature Publishing Group.

Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J., and Schaphoff, S. (2008). Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, 44(9).

Rothman, D. S., Hughes, B. B., and Narayan, K. (2016). IFs Agriculture Model Documentation. Working Paper Research 2017.07.04, Pardee Center for International Futures, Josef Korbel School of International Studies, University of Denver, Denver, CO.

Running, S. W. (2012). A Measurable Planetary Boundary for the Biosphere. *Science*, 337(6101):1458–1459. Publisher: American Association for the Advancement of Science.

Schaphoff, S., von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., Gerten, D., Heinke, J., Jägermeyr, J., Knauer, J., Langerwisch, F., Lucht, W., Müller, C., Rolinski, S., and Waha, K. (2018). LPJmL4 – a dynamic global vegetation model with managed land – Part 1: Model description. *Geoscientific Model Development*, 11(4):1343–1375. Publisher: Copernicus GmbH.

Scholes, R. J. and Biggs, R. (2005). A biodiversity intactness index. *Nature*, 434(7029):45–49. Number: 7029 Publisher: Nature Publishing Group.

Sen, A. and Dabi, N. (2021). Tightening the Net: Net zero climate targets – implications for land and food equity. Oxfam. Available at https://oxfamilibrary.openrepository.com/handle/10546/621205.

Senthil Kumar, P. and Yaashikaa, P. R. (2019). 1 - Introduction—Water. In Muthu, S. S., editor, *Water in Textiles and Fashion*, pages 1–20. Woodhead Publishing.

Serafeim, G. and Velez Caicedo, G. (2022). Machine Learning Models for Prediction of Scope 3 Carbon Emissions. Harvard Business School Accounting & Management Unit Working Paper No. 22-080, Available at SSRN: https://ssrn.com/abstract=4149874 or http://dx.doi.org/10.2139/ssrn.4149874.

Stangl, J., Borsos, A., Diem, C., Reisch, T., and Thurner, S. (2023). Using firm-level production networks to identify decarbonization strategies that minimize social stress.

Stanton, E. A., Ackerman, F., and Kartha, S. (2009). Inside the integrated assessment models: Four issues in climate economics. *Climate and Development*, 1(2):166–184.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., and Ludwig, C. (2015a). The trajectory of the Anthropocene: the great acceleration. *The Anthropocene Review*, 2(1):81–98. Publisher: SAGE Publications Sage UK: London, England.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., De Vries, W., De Wit, C. A., et al. (2015b). Planetary boundaries: Guiding human development on a changing planet. *Science*, 347(6223).

Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., Summerhayes, C. P., Barnosky, A. D., Cornell, S. E., Crucifix, M., et al. (2018). Trajectories of the earth system in the anthropocene. *Proceedings of the National Academy of Sciences*, 115(33):8252–8259.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge University Press.

Thadakamaila, H. P., Raghavan, U., Kumara, S., and Albert, R. (2004). Survivability of multiagent-based supply networks: a topological perspect. *IEEE Intelligent Systems*, 19(5):24–31. Conference Name: IEEE Intelligent Systems.

Thurner, S., Hanel, R., and Klimek, P. (2018). *Introduction to the theory of complex systems*. Oxford University Press, Oxford, UK.

UNEP (2019). *Global Environment Outlook – GEO-6: Healthy Planet, Healthy People*. Cambridge University Press. UNEP - UN Environment Programme.

Usubiaga, A., Acosta-Fernández, J., McDowall, W., and Li, F. G. N. (2017). Exploring the macro-scale CO_2 mitigation potential of photovoltaics and wind energy in Europe's energy transition. *Energy Policy*, 104:203–213.

Van Regemorter, D., Perry, M., Capros, P., Ciscar, J., Paroussos, L., Pycroft, J., Karkatsoulis, P., Abrell, J., and Saveyn, B. (2013). GEM-E3 model documentation. Perry, M. (editor), Ciscar, J. (editor), Pycroft, J. (editor), Abrell, J. (editor), Saveyn, B. (editor), Publications Office of the European Union. doi: doi/10.2788/47872.

van Vuuren, D. P., Kok, M., Lucas, P. L., Prins, A. G., Alkemade, R., van den Berg, M., Bouwman, L., van der Esch, S., Jeuken, M., Kram, T., and Stehfest, E. (2015). Pathways to achieve a set of ambitious global sustainability objectives by 2050: Explorations using the IMAGE integrated assessment model. *Technological Forecasting and Social Change*, 98:303–323.

Vines, D. and Wills, S. (2018). The rebuilding macroeconomic theory project: an analytical assessment. *Oxford Review of Economic Policy*, 34(1-2):1–42.

Waisman, H., Guivarch, C., Grazi, F., and Hourcade, J. C. (2012). The Imaclim-R model: infrastructures, technical inertia and the costs of low carbon futures under imperfect foresight. *Climatic Change*, 114(1):101–120.

Wang-Erlandsson, L., Tobian, A., van der Ent, R. J., Fetzer, I., te Wierik, S., Porkka, M., Staal, A., Jaramillo, F., Dahlmann, H., Singh, C., Greve, P., Gerten, D., Keys, P. W., Gleeson, T., Cornell, S. E., Steffen, W., Bai, X., and Rockström, J. (2022). A planetary boundary for green water. *Nature Reviews Earth & Environment*, 3(6):380–392.

Weidner, T. and Guillén-Gosálbez, G. (2023). Planetary boundaries assessment of deep decarbonisation options for building heating in the European Union. *Energy Conversion and Management*, 278:116602.

Weitzman, M. L. (2009). On Modeling and Interpreting the Economics of Catastrophic Climate Change. *The Review of Economics and Statistics*, 91(1):1–19. MAG ID: 2097259059.

Weyant, J. (2017). Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy*, 11(1):115–137. Publisher: The University of Chicago Press.

WHO, W. (2022). Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

Wiedmann, T. O., Schandl, H., Lenzen, M., Moran, D., Suh, S., West, J., and Kanemoto, K. (2015). The material footprint of nations. *Proceedings of the national academy of sciences*, 112(20):6271–6276.

Wigley, T. M. L. (2005). The Climate Change Commitment. *Science*, 307(5716):1766–1769. Publisher: American Association for the Advancement of Science.

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J., and Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change*, 166(1):3.

Wolf, S., Fürst, S., Mandel, A., Lass, W., Lincke, D., Pablo-Martí, F., and Jaeger, C. (2013). A multi-agent model of several economic regions. *Environmental Modelling & Software*, 44:25–43.

Wu, D. (2016). Shock spillover and financial response in supply chain networks: Evidence from firm-level data. *Unpublished working paper.*

Wunderling, N., Donges, J. F., Kurths, J., and Winkelmann, R. (2021). Interacting tipping elements increase risk of climate domino effects under global warming. *Earth System Dynamics*, 12(2):601–619. Publisher: Copernicus GmbH.

Xu, L., Fuss, M., Maryegli Fuss, Poganietz, W.-R., Jochem, P., Schreiber, S., Zoephel, C., and Brown, N. (2020). An Environmental Assessment Framework for Energy System Analysis (EAFESA): The method and its application to the European energy system transformation. *Journal of Cleaner Production*, 243:118614.

Annex A

DPhil papers

Evaluating the efficacy of Paris-compliant mitigation pathways in tackling the ecological crisis

Andrea Bacilieri[1,2*]

[1]Institute for New Economic Thinking, University of Oxford, Oxford, UK.
[2]Smith School of Enterprise and Environment, University of Oxford, Oxford, UK.

Corresponding author(s). E-mail(s): andrea.bacilieri@sant.ox.ac.uk;

Abstract

Greater, unified efforts are needed to address the climate and ecological crises. While Integrated Assessment Models (IAMs) provide a basis for global policy decisions on climate change, they are not used to address ecological concerns. The Planetary Boundaries (PBs) framework sets a safe operating space for humanity that can be used in IAMs to provide a basis for addressing climate and ecological issues. We assess whether the Paris-compatible mitigation pathways considered by Working Group III in the latest IPCC report address ecological problems using the PBs framework. We find that few consider the variables underlying the PBs framework, and of those that do, very few remain within the PBs. We examine the phosphorus PB in more detail and discuss the drivers behind its transgression. Our findings highlight the need for a systems-thinking approach in developing mitigation pathways so that the climate and ecological crises are addressed together.

Keywords: Integrated Assessment Model, IAM, planetary boundary, IPCC, mitigation, scenario, pathway

1 Main

The current state of our planet is characterised by interconnected environmental challenges that threaten its stability and the survival of life on it. The challenges are many and varied, ranging from climate change to biodiversity loss, pollution and ecosystem degradation. The Planetary Boundaries (PBs) framework [1] establishes nine fundamental Earth system processes that regulate the stability of the planet: climate change, ocean acidification, stratospheric ozone depletion, biogeochemical flows (i.e., the nitrogen and phosphorus cycles), land-system change, freshwater use, biosphere integrity (divided into genetic diversity and functional diversity), atmospheric aerosol loading and chemical pollution. For most of these processes, boundaries can be identified that if crossed could drive the Earth system away from the Holocene stable state, the stable geological epoch that enabled the flourishing of human life as we know it. Crossing such boundaries could lead to potentially irreversible damages – at least in human timescales. We have already crossed five of these boundaries [1–3].

The PBs framework has received some criticisms [4–6], but it has also received broad support. In fact, it is arguably the most robust framework we currently have for assessing our trajectory and possible viable futures [7]. The PBs framework can provide a scientific basis for developing policies and practices that met the challenges of maintaining the Holocene state while enabling the flourishing of human societies. By monitoring these boundaries, policymakers can ensure that human activities remain within the safe operating space of the planet. Solutions to these challenges exist, but they often involve trade-offs and synergies between different economic, social and environmental factors [8]. These trade-offs and synergies require a comprehensive approach that takes into account the interactions between different human and natural systems.

Scientists have developed computer models to understand Earth system and socio-economic processes, and their interactions. These models are called Integrated Assessment Models (IAMs) and include an economic and an energy system model, which are typically integrated with a terrestrial vegetation model, a hydrological model, and an ocean and atmospheric model, each at varying levels of detail in different IAMs [9]. Besides the economic and energy system model, not all IAMs include each of the other modules. IAMs conduct 'what-if' analyses under different storylines that describe a variety of possible socio-economic developments; as such, they explore a wide range of futures that pose different challenges for mitigation and adaptation [10]; exercises of this kind provide important insights.

For policy purposes, however, it is important that these mitigation pathways are feasible, but this remains an open question. While there are many dimensions of feasibility that should be considered [11, 12], IAMs mainly, if not exclusively, consider economic costs and incentives [13]. This article focuses on ecological feasibility for two reasons. First, as discussed above, we are transgressing several PBs that could tip the Earth system into an undesirable state with potentially catastrophic consequences for the survival of the planet. Second, socio-technical transitions unfold over long periods of time with path dependencies that can lead to 'lock-in' technologies and infrastructure [14]. Therefore, the choices we make today require careful consideration along ecological and climate dimensions. Otherwise, we may solve the climate crisis while exacerbating others.

To fill this gap, the latest IPCC Sixth Assessment Report (AR6) makes groundbreaking progress in recognising the need to address the climate and ecological crises together. The contributions of Working Group II (WGII) and WGIII provide an in-depth discussion of the current and potential ecological consequences of the observed impacts and projected risks associated with climate change, as well as possible adaptation options and mitigation scenarios [8, 15]. However, to successfully address these challenges, ecological considerations require a more rigorous, and integrated approach and assessment, moving beyond discussing synergies and trade-offs only and instead integrate, for example, the PBs framework in the assessment of mitigation pathways.

Several studies take a step in this direction, many of which couple an energy model with a Life Cycle Assessment (LCA) model, and sometimes also link these to a macroeconomic input-output model, to assess the impacts of different mitigation pathways of the energy system on ecosystems and human health [16–20]. Although the integration of LCA models is a promising future avenue for IAMs, so far LCA studies have only identified trade-offs and synergies. Another set of models assesses one or more PBs, for example, on the adoption of bioenergy with carbon capture and storage (BECCS) [21–23], different building heating technologies [24], food demand [25, 26] or land use change [27]. A comprehensive example is [28], who use a system dynamics model to consider future development paths for Fiji that remain within the PBs.

In this paper, we analyse the vetted mitigation scenarios by WGIII that stay within Paris-compatible levels of warming under a PBs lens to assess their ecological feasibility. The report of WGIII assesses the feasibility of the vetted scenarios in terms of geophysical, economic, technological, institutional and socio-cultural dimensions. However, geophysical feasibility, which is the only dimension relevant to the ecological crisis, only considers primary energy production from biomass and wind, and secondary energy production from solar, establishing thresholds to identify low, medium and high levels of concern [8]. Our analysis is therefore (1) broader in scope because it considers all the technologies and lifestyle changes that have repercussions on the Earth system, (2) more detailed since it breaks down different ecological dimensions and (3) more rigorous since it relies on the PBs framework. Having assessed whether mitigation pathways remain within the PBs, we explore the drivers that lead to the PBs being breached, with a focus on phosphorus use.

2 Results

2.1 PBs in mitigation pathways

For almost all of the 9 Earth-system processes identified to be fundamental for the stability of the planet in the PBs framework, control variables with associated boundaries can be identified, which allow us to monitor anthropogenic environmental change on a global scale. We use these control variables and the established boundaries to assess the ecological feasibility of Paris-compliant mitigation scenarios. We use the mitigation scenarios available in the Scenario Explorer database (see Methods) and select the vetted scenarios by WGIII that stay within the levels of warming agreed upon by countries in the 2015 Paris Agreement (see Fig. 1 for more detail); there are a total of 700 scenarios.

Among the 9 control variables identified in the PBs framework, we can analyse only 4, namely phosphorus (P) and nitrogen (N) use, water consumption and land use change (see Fig. 1 for a definition of the control variables and associated boundary values we use). We cannot analyse the other control variables, and associated processes, because these are not included in the different IAMs associated with the selected scenarios. As discussed before, while all IAMs include an economic and energy model, they differ with regard to the modelling and inclusions of aspects related to, e.g., water and land use. Therefore, not all the scenarios we analyse include all 4 PB control variables.

We analyse how many of the scenarios report a PB control variable and how many stay within the boundary. Of the 700 vetted scenarios, only 120 consider P, 332 N, 90 water use and 560 land system change (Fig. 1).[1] In 2100, of the 120 scenarios considering P, 12% transgress the PB, while for water consumption, 30% of the scenarios exceed the PB. Only 13% of the scenarios return to safe levels of N use in 2100. The minimum forest cover for the land-system change PB is exceeded by 86% of the scenarios (Fig. 1). 14 scenarios report all four PBs; the only PB that is respected by all these 14 scenarios is water use.

[1]We do not consider water withdrawal as it is used as a control variable for freshwater use at the river basin scale [1].

It can be noticed that the scenarios' time series in Fig. 1 can have different starting years. This is because models are initialised in different years, often earlier than the present day. This means their predictions for the early years can be compared to actual outcomes. The predictions are often off (Fig. B8), highlighting that models are not constantly updated as new data come in. For example, the scenarios that stay within the P PB consistently underpredict its use between 2000 and 2020 by 2 orders of magnitudes, while those that do not stay within the P PB consistently overpredict its use by almost a factor of 2. For land-system change, a few scenarios match the empirical data, but most of the scenarios have a much lower forest cover. For N, some of the scenarios' predictions are consistent with the empirical data, while others overpredict its use by up to a factor of about 2 (see SI B for why we excluded water consumption). These discrepancies suggest that scenarios' projections should be treated with caution and point to the need for more in-depth analyses on this issue.

We focus on the Illustrative Mitigation Pathways (IMPs) defined by WGIII. The IMPs comprise five overarching mitigation-strategy themes across the scenarios considered by WGIII. Each IMP includes a storyline, defining socio-economic features of the pathway, and a reference quantitative scenario from the literature. IMPs differ with regard to the level of mitigation ambitions and strategies, for example, some exclude certain technologies while others focus on the Sustainable Development Goals or have delayed action.

There are 5 IMPs that stay with the Paris-agreed warming levels: GS, Neg, Ren, LD and SP. Two sensitivity cases, Ren-2.0 and Neg-2.0, are also considered to explore different levels of warming than the corresponding reference IMP. We describe some of them in more detail below and give a high-level description here. GS considers a scenario where current policies are gradually strengthened up to 2030, while Neg assumes more mitigation efforts than GS and relies on carbon dioxide removal (CDR) to bring temperatures back to 1.5° after a high overshoot. Ren describes a future characterised by high levels of renewables, LD a scenario with low energy demand and SP a world that embarks on a sustainable development pathway.

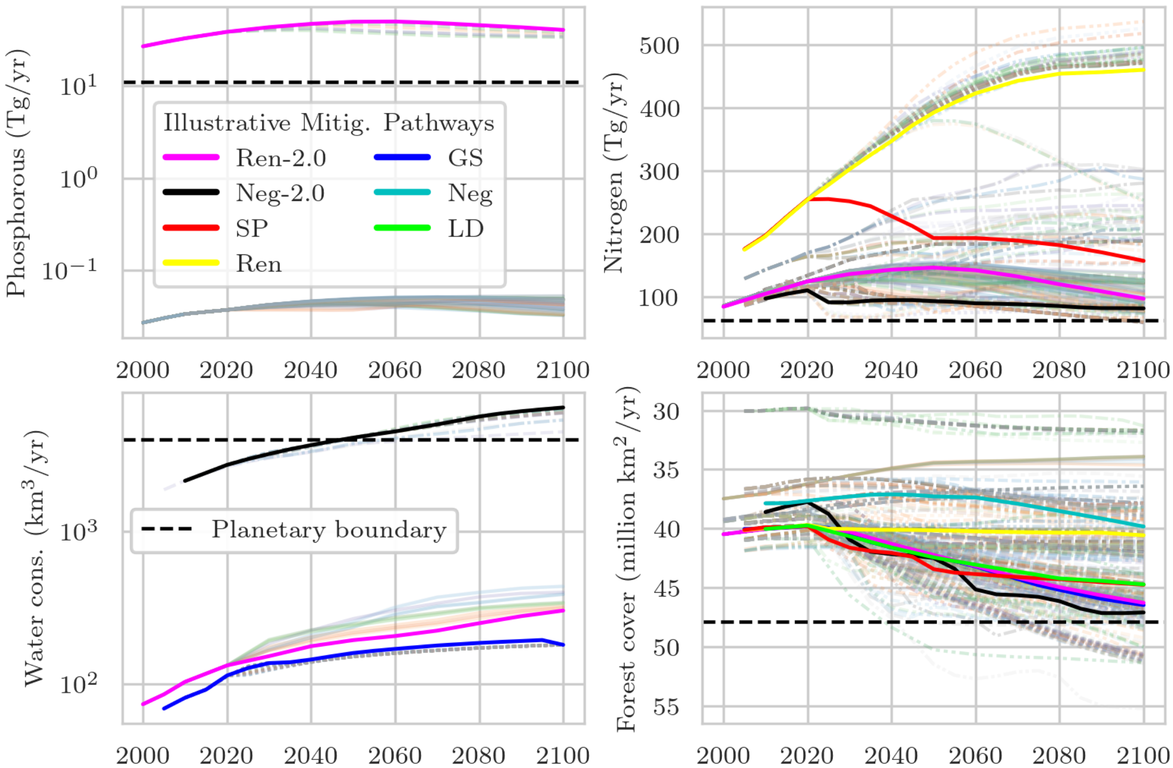

Fig. 1 Planetary boundaries in the 700 mitigation pathway. We highlight Illustrative Mitigation Pathways (Table 2, the Summary for Policymakers, WGIII; 29) that report a PB: Ren-2 (pink), Neg-2 (black), SP (red), Ren (yellow), GS (blue), Neg (light blue) and LD (green). For all other pathways, colours correspond to scenarios and line styles correspond to model families; the legend is not shown for parsimony. Except for the IMPs, we show vetted scenarios in categories C1, C2, C3 and C4; see Table 2 in the Summary for Policymakers of the WGIII report [29]. The black dashed line marks the lower limit of each PB; "PB" stands for Planetary Boundary. The PB definition is based on [1]. Global P use limits (11 Tg/yr) are related to the amount that flows from freshwater systems into the ocean. Global N use limits (62 Tg/yr) refer to industrial and intentional biological fixation of N. Freshwater use (4,000 km^3/yr) is consumptive blue water use; however, we observe water consumption without any additional information. Land-system change is defined as the area of forested land as a percentage of the original forest cover and is a weighted average of the boundaries on the three major biomes (tropical, temperate and boreal). At least 75% of potential forest cover should be maintained; this corresponds to 47.9 million km^2 of the Earth's ice-free land surface [1, SI, p. 7]. The variable we use for land system change is "Land Cover–Forest", which includes undisturbed forest, afforestation and reforestation, regrown secondary forest, forests providing wood for timber and energy, and harvested forest; it is in million ha [30], so we convert it to million km^2. This variable is an overestimate of the PBs for land system change, so the results are more optimistic and need to be interpreted with some caution.

In Fig. 1, we highlight the trajectory of the PB control variables of the IMPs reference quantitative scenarios that stay with the levels of warming agreed in Paris [Table 10, Annex III, 8]. All of them report at least one of the PB control variables, but only a few stay within PB. The LD and SP IMPs describe futures that would be considered more desirable to follow. These represent sustainable futures in which resources are used more efficiently, energy demand is reduced and lifestyles change towards more sustainable habits and practices, basic needs are met (LD) and possibly also inequality is reduced and the SDGs are achieved (SP). The LD scenario reports only forest cover, which increases over time and reaches 44.7 million km^2 in 2100, almost the PB level of 47.9 million km^2. SP has a lower forest cover of 44.8 million km^2 in 2100 and is one of the pathways with the lowest N use but the highest water use. Neg is the best-performing pathway with 47.1 million km^2 of forest cover in 2100, probably due to its heavy reliance on BECCS, afforestation and reforestation (A/R); this result is likely a consequence of how we define the land-use change variable (see Fig. 1).

GS and Ren are the only two pathways that remain within the freshwater PB. The Ren scenario is characterised by successful international coordination that is able to implement climate policies. They envisage a future with high electrification and rapid deployment of renewables, combined with high levels of storage and Power-to-X technologies, as well as better interconnections. Accordingly, demand adapts to the high level of renewable energy supply. However, there is no overarching narrative regarding land and water use, food and biodiversity. The GS pathway assumes that nationally determined contributions will be implemented by 2030, after which climate policies and international commitments and cooperation will be strengthened. However, due to the delay in action, it relies heavily on CDR, such as BECCS and A/R, leading to increased competition for land.

For the IMPs, we evaluate the geophysical feasibility measure used by WGIII. Geophysical feasibility considers physical and resource constraints associated with implementation. It is defined and assessed for primary energy from biomass and wind, and secondary energy from solar. Three levels of concern are defined: low, when implementation is feasible and akin to the past; medium, when challenges are mild and can be overcome; and high, when implementation entails unprecedented levels of transformation that pose considerable challenges. The thresholds associated with each concern level are taken from the literature.

Biomass is of medium concern if the primary energy is above 100 EJ/yr and of high concern if it is above 245 EJ/yr; for wind, these values are 830 EJ/yr and 2,000 EJ/yr, respectively; and for solar, the threshold is 1,600 EJ/yr for medium concern and 50,000 EJ/yr for high concern [Table 8, Annex III, 8]. Although feasibility concerns can be time and context dependent [11], we examine them only in 2100. Concerns are raised only for biomass, with high concerns for Neg and Neg-2.0, and medium concerns for Ren-2.0 and SP, one of the most positive scenarios; Ren is just below the threshold (99.54 EJ/yr). For Neg, there is a high level of concern about the feasibility of biomass because of the strong reliance on BECCS.

2.2 Transgressing the phosphorus PB

We examine in more detail the 120 scenarios that report P use, which fall into two distinct clusters (Fig. 1 and SI Fig. A1). To have a systematic procedure that can be used across many scenarios developed using different IAMs, we use an Elastic Net regression technique to investigate what are the likely drivers behind the crossing or not of the P PB. We run separate regressions for primary, secondary and final energy; the other variables are used in all three regressions (see Methods).

14 scenarios overshoot the P PB, of which 11 return warming to 1.5°C after a high overshoot (C2) and 3 limit warming to 2°C with a probability greater than 67% (C3; SI Tab. A1). Instead, P-PB-compliant scenarios may also limit warming to 1.5°C after no or limited overshoot (C1; 8% of P-PB-compliant scenarios) or limit warming to 2°C with a probability greater than 50% (C4; 26% of P-PB-compliant scenarios). The majority of the scenarios that respect the P PB limit warming to 2°C with a probability greater than 67% (42%).

The regression results (see Methods) show that the most influential variables for our models' predictions are related to land use change, the food system, water use and biomass use (Fig. A3, A4 and A5); this is emphasised by all three regressions for primary, secondary and final energy. Scenarios that exceed the P PB are associated with higher demand for round wood, agricultural products and livestock feed, and produce more energy from biomass. These scenarios therefore require more land for crops, especially energy crops, thereby leading to higher P use as a fertiliser. These factors in turn reduce natural forest cover and require more water for irrigation (Fig. A6).

Among the 120 scenarios, we select 3 representative scenarios using k-means clustering (see SI Fig. A1) and discuss a subset of variables among those previously selected as most significant (Fig. 2 and SI Fig. A6). For convenience, we refer to scenario `SSP2_openres_lc_120` as SSP2, scenario `EN_INDCi2030_900f_NDCp` as INDC and scenario `EN_NPi2020_900` as NP. SSP2 transgresses the P PB, while INDC and NP do not. Apart from a few differences in PV costs (taken from the LED scenario), hydro costs (taken from SSP3) and costs of all other electricity generation technologies (taken from SPP1), SSP2 follows the SSP2 storyline and quantitative assumptions [see 31, for details on the SSPs]. The NP scenario implements existing national policies until 2030, while INDC implements the Nationally Determined Contributions; after 2030, both scenarios continue with the same mitigation effort compared to a baseline scenario. Otherwise, they both follow the SSP2 scenarios [32].

SSP2 uses 988% and 751% more P than INDC and NP, respectively and none of the three scenarios is within the N PB, with SSP2 using 3% and 6% less than NP and INDC, respectively. Agricultural demand is 72% higher and livestock food demand is 27% higher in SSP2 than in NP, while these are 35% higher and 49% higher, respectively, than in INDC. SSP2 also uses more than twice as much round wood and primary energy from biomass than NP, but only 52% more round wood and 58% more primary energy from biomass than INDC. As a result, the land required for crops, food, feed and biomass in SSP2 is more than twice as much as in NP and 44% more than in INDC. In turn, SSP2 withdraws 44% and 77% more water than NP and INDC, respectively. While none of the 3 scenarios is within the PB for land system change, natural forest cover is significantly reduced in SSP2, with 45% and 32% less forest cover than in NP and INDC, respectively.

Biomass is not only used for electricity production but also as a biofuel. In fact, in SSP2, the consumption of refined liquids is more than double the amount consumed in NP and almost twice the amount consumed in INDC. SSP2 produces almost all liquid fuels from bioenergy, as does INDC, while NP also produces half of the liquid fuels from oil. Some of the biomass is also used to produce gases, which are then used, for example, in the residential and commercial sectors or in the transport sector. However, biogas is less important and most gases are produced from natural gas in all three scenarios.

As before, we check the geophysical feasibility used by WGIII and only for 2100. In line with our previous findings, SSP2 raises high feasibility concerns for biomass, while NP and INDC raise medium concerns. Feasibility concerns for solar and wind are not raised by any of the three scenarios.

3 Discussion

In this paper, we assessed the ecological feasibility of IPCC mitigation scenarios that stay within 1.5°C or 2°C of warming. To assess ecological feasibility, we use the PBs framework [1]. While some have emphasised that mitigation pathways are not intended to serve as prescriptive guidelines or blueprints for policymakers, it is important to recognise their influence on policy decisions and the need for careful consideration of their impacts on society as a whole. Therefore, pathways that are considered to represent more desirable futures, such as LD pathways, an example of which is the LED scenario [36], or the SP pathways, an example of which is [37], need to be assessed in the context of environmental challenges. The PB framework can help with this assessment.

We have shown that the SP reference scenario is of medium concern in terms of biomass use and that both the SP and LD reference scenarios do not stay within the PB for land system change. While forest cover is increasing over time and is close to the PB threshold, the forest cover variable we used includes areas of forest used for timber and energy production as well as harvested forest, and therefore paints a more optimistic picture. SP also exceeds the freshwater and N PB, although it has one of the lowest N uses. However, because other PB control variables were not explored in the SP and LD scenarios, we were unable to assess the other PBs, which would have been of particular interest given the positive futures they depict. The PB framework could be used in IAMs to explore pathways that provide climate solutions and meet other environmental goals. It could also be used in conjunction with other feasibility analyses, such as that used by WGIII [8, 11], which are able to address, for example, political and societal axes of feasibility.

We analysed the P PB in more detail and assessed the drivers behind crossing this PB using an Elastic Net regression. We find that land-use change, meat-intensive lifestyles and biomass use are key drivers of P use transgressing the PB threshold. These results highlight that, first, biomass should play a very limited role in the energy transition, as pointed out in other studies [21–23, 38–40]. Second, even in scenarios that assume a high use of renewables, possibly including also the use of global energy interconnections,

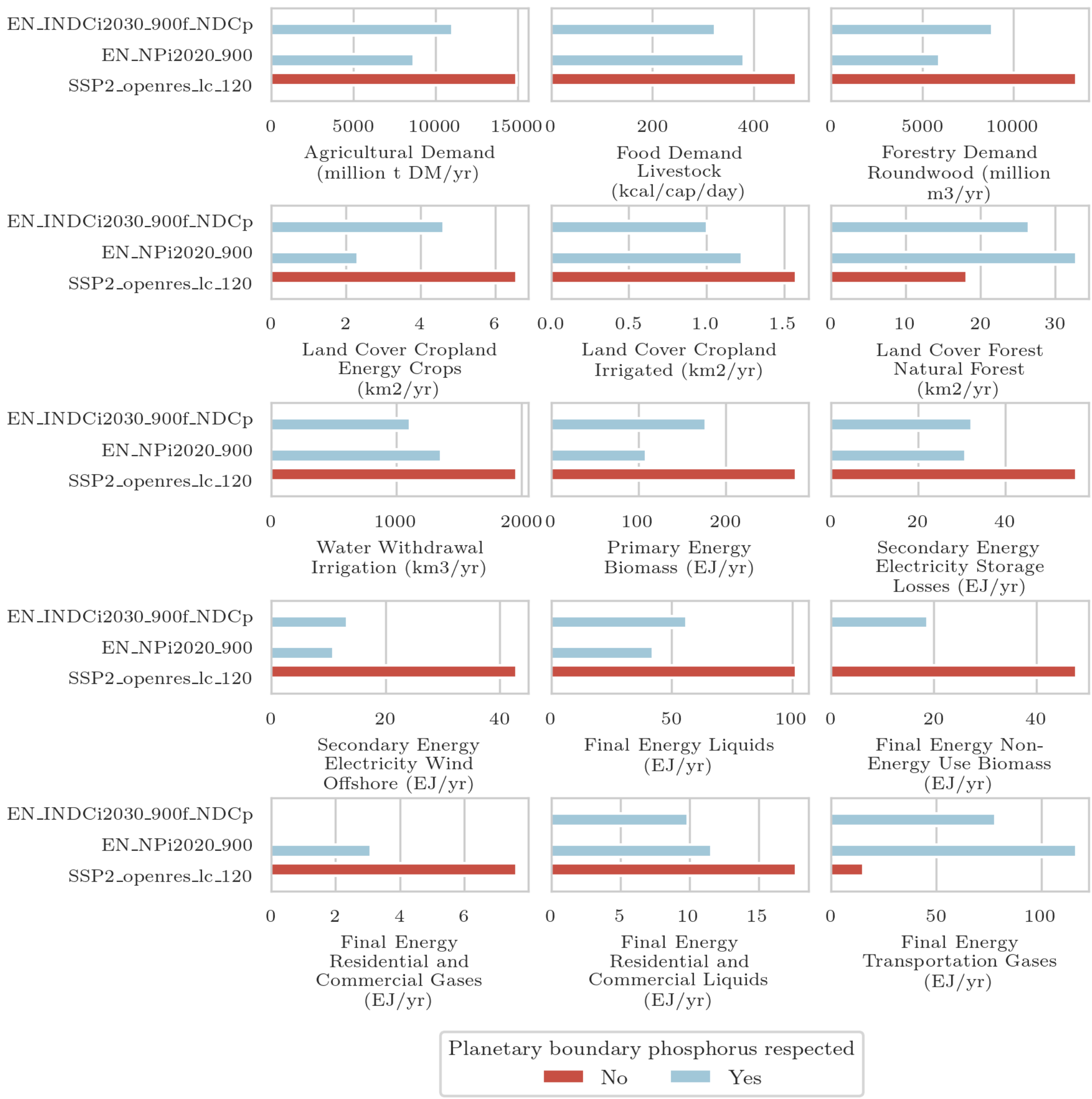

Fig. 2 Most representative features for three representative scenarios for 2100, for scenarios that stay within the PB level of phosphorus use (light blue) and scenarios that do not (red). We select the representative scenarios using the k-means clustering algorithm (see SI Fig. A1). The `EN_INDCi2030_900f_NDCp` and `EN_NPi2020_900` scenarios use MESSAGEix-GLOBIOM 1.1 [32–34]; the `SSP2_openres_lc_120` scenario uses MESSAGEix-GLOBIOM GEI 1.0 [35]. For selecting the most influential variables in predicting phosphorus use above the PB, we use Elastic Net regressions (see Methods); we show a subset of those variables that most represent differences across scenarios.

a high use of biomass as a fuel still has a negative impact on ecosystems, putting them under further stress. These findings underline the importance of demand-side policies aimed at changing current mobility patterns [8].

Increased agricultural production and meat-based lifestyles are also contributing to the observed decline in forest cover and could put further pressure on ecosystems [41]. Again, these findings highlight the urgency and need for demand-side policies and lifestyle changes, calling for a shift in our diets to be more plant-based and less meat-based [8, 26]. Increased agricultural production and demand for livestock feed also lead to high water use for irrigation. Although within safe limits in the three reference scenarios we analysed in more detail, more water for irrigation can lead to more phosphorus run-off into nearby water bodies such as lakes and rivers, increasing pressure on aquatic ecosystems and possibly also leading to eutrophication.

The control variables we used to assess the PBs are likely to give a more optimistic picture because, for example, forest cover includes harvested forests and forests used for timber and energy production, while water, P and N use do not take into account other production and consumption processes in the economy. A full accounting might therefore give a different picture, especially if the environmental impacts generated throughout the whole supply chain are taken into account [16, 18, 20, 42, 43].

Analysing mitigation scenarios through a PBs lens highlights the limitations of some of the Paris-compliant scenarios, and underlines the need for a comprehensive and integrated approach that seeks to solve the climate and ecological crises together. Mitigation scenarios need to be evaluated not only on their ability to reduce carbon emissions, but also on the ecological impacts of the proposed solutions. This is a necessary step forward that IAMs must take, especially as these models are an influential policy tool and have implications for society as a whole. Their current narrow focus on climate change alone can lead to solutions that inadvertently exacerbate other ecological problems.

4 Methods

Statistical analysis

To understand the drivers leading to the overshoot of the P PB, we use a penalised regression. There are three possible options of penalisation to use: L1 norm, L2 norm or a combination of the two. The L1 norm, also called Lasso regression, acts as a penalty term that encourages sparsity in the model, so it helps in the selection of important variables. The L2 norm, also called Ridge regression, shrinks the values of the estimated coefficients but not to zero. Most importantly, the Ridge regression helps alleviate problems related to multicollinearity by imposing a size constraint on the coefficients [44]. While Lasso keeps only one variable among a group of variables that are highly correlated, Ridge selects the whole group – i.e., it does group selection. The Elastic Net regression, which uses both the L1 and L2 norm, achieves the best of both words. It does variable selections like the Lasso while selecting groups of important, correlated variables [45]. We select the Elastic Net regression because we have a large number of independent variables that are highly correlated with each other. We use `scikit-learn`, a machine learning library in Python https://scikit-learn.org/stable/.

We conduct separate regressions for primary, secondary and final energy because of multicollinearity problems. All the other variables are common across all three regressions. We convert phosphorus use into a binary variable that equals 1 if a scenario stays within the P PB and 0 otherwise. We standardise all regressors.

To tune the hyperparameters, we perform a 2-fold cross-validation (CV) and use the accuracy to tune the hyperparameters. For the inverse of the regularisation strength, we explore values in $[0.01, 10]$, with values that are log-spaced. For the L1 ratio, we explore values that are in $[0, 0.99]$ with linear increments of 0.01. We do two tests: (1) we leave out 30% of the data and train the model on the rest of the data, doing the 2-fold CV, and (2) we do the 2-fold CV on the whole dataset. The results do not change, so we use (2) given the small size of the dataset (120 observations) and its unbalanced nature (14 observations exceeded the P PB). We therefore perform a stratified CV.

After training, we are left with a plane of possible hyperparameter values that maximise accuracy (SI Fig. A2). We fit all these models and then select the variables that were considered by at least 80% of the models.

Data cleaning

We use data about mitigation pathways available in the AR6 Scenario Explorer (see Data). We describe common procedures applied to all three regression analyses. First, we keep only scenarios in categories C1, C2, C3 and C4 as reported in Table 2 in the Summary for Policymakers of the report of WGIII [8]. Second, since pathways either always or never transgress the P PB threshold, for each variable, we investigate the 2100 values only. Third, we drop all variables related to the "AR6 climate diagnostics" and emissions. In the Scenario Database that is a total of 1,656 variables; dropping those two sets of variables leaves us with 1,583 variables, of which 775 are not covered by models that look at P use, so we are left with 808. We then drop variables that have missing values, that are zero across all scenarios and that have low variance, meaning that the coefficient of variation needs to be above 0.02. We further excluded certain variables manually because they exhibit very low variation between scenarios that transgressed the P PB and those that do not; we learned this through exploration of the data. We keep GDP in PPP and exclude GDP in MER.

Data. The raw data used in this study are available through the AR6 Scenario Explorer at https://data.ene.iiasa.ac.at/ar6/#/login [30]. We used file `AR6_Scenarios_Database_World_v1.1` and `AR6_Scenarios_Database_metadata_indicators_v1.1.xlsx`.

Supplementary information. This article has supplementary information.

Ethics declarations

Competing interests. The authors declare no competing interests.

References

[1] Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A., et al.: Planetary boundaries: Guiding human development on a changing planet. Science **347**(6223) (2015)

[2] Persson, L., Carney Almroth, B.M., Collins, C.D., Cornell, S., de Wit, C.A., Diamond, M.L., Fantke, P., Hassellöv, M., MacLeod, M., Ryberg, M.W., Søgaard Jørgensen, P., Villarrubia-Gómez, P., Wang, Z., Hauschild, M.Z.: Outside the Safe Operating Space of the Planetary Boundary for Novel Entities. Environmental Science & Technology **56**(3), 1510–1521 (2022). https://doi.org/10.1021/acs.est.1c04158. Accessed 2022-02-02

[3] Fanning, A.L., O'Neill, D.W., Hickel, J., Roux, N.: The social shortfall and ecological overshoot of nations. Nature Sustainability, 1–11 (2021)

[4] Biermann, F., Kim, R.E.: The Boundaries of the Planetary Boundary Framework: A Critical Appraisal of Approaches to Define a "Safe Operating Space" for Humanity. Annual Review of Environment and Resources **45**(1), 497–521 (2020). https://doi.org/10.1146/annurev-environ-012320-08033

[5] Downing, A.S., Bhowmik, A., Collste, D., Cornell, S.E., Donges, J., Fetzer, I., Häyhä, T., Hinton, J., Lade, S., Mooij, W.M.: Matching scope, purpose and uses of planetary boundaries science. Environmental Research Letters **14**(7), 073005 (2019). https://doi.org/10.1088/1748-9326/ab22c

[6] Montoya, J.M., Donohue, I., Pimm, S.L.: Why a Planetary Boundary, If It Is Not Planetary, and the Boundary Is Undefined? A Reply to Rockström et al. Trends in Ecology & Evolution **33**(4), 234 (2018). https://doi.org/10.1016/j.tree.2018.01.00. Accessed 2023-03-25

[7] Rockström, J., Steffen, W., Noone, K., Persson, r., Chapin, F.S.I., Lambin, E., Lenton, T., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R., Fabry, V., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.: Planetary Boundaries: Exploring the Safe Operating Space for Humanity. Ecology and Society **14**(2) (2009). https://doi.org/10.5751/ES-03180-14023. Publisher: The Resilience Alliance. Accessed 2023-03-12

[8] IPCC: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M.

Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926 (2022)

[9] Harfoot, M., Tittensor, D.P., Newbold, T., McInerny, G., Smith, M.J., Scharlemann, J.P.W.: Integrated assessment models for ecologists: the present and the future. Global Ecology and Biogeography **23**(2), 124–143 (2014). https://doi.org/10.1111/geb.1210. Accessed 2023-04-08

[10] Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change **42**, 153–168 (2017). https://doi.org/10.1016/j.gloenvcha.2016.05.009

[11] Brutschin, E., Pianta, S., Tavoni, M., Riahi, K., Bosetti, V., Marangoni, G., van Ruijven, B.J.: A multidimensional feasibility evaluation of low-carbon scenarios. Environmental Research Letters **16**(6), 064069 (2021). https://doi.org/10.1088/1748-9326/abf0ce

[12] Warszawski, L., Kriegler, E., Lenton, T.M., Gaffney, O., Jacob, D., Klingenfeld, D., Koide, R., Costa, M.M., Messner, D., Nakicenovic, N., Schellnhuber, H.J., Schlosser, P., Takeuchi, K., Leeuw, S.V.D., Whiteman, G., Rockström, J.: All options, not silver bullets, needed to limit global warming to 1.5°C: a scenario appraisal. Environmental Research Letters **16**(6), 064037 (2021). https://doi.org/10.1088/1748-9326/abfeec. Publisher: IOP Publishing

[13] Anderson, K., Jewell, J.: Debating the bedrock of climate-change mitigation scenarios. Nature Publishing Group (2019)

[14] Li, F.G.N., Trutnevyte, E., Strachan, N.: A review of socio-technical energy transition (STET) models. Technological Forecasting and Social Change **100**, 290–305 (2015). https://doi.org/10.1016/j.techfore.2015.07.01

[15] IPCC: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi: 10.1017/9781009325844 (2022)

[16] Hertwich, E.G., Gibon, T., Bouman, E.A., Arvesen, A., Suh, S., Heath, G.A., Bergesen, J.D., Ramirez, A., Vega, M.I., Shi, L.: Integrated life-cycle assessment of electricity-supply scenarios confirms global environmental benefit of low-carbon technologies. Proceedings of the National Academy of Sciences **112**(20), 6277–6282 (2015)

[17] Gibon, T., Arvesen, A., Hertwich, E.G.: Life cycle assessment demonstrates environmental co-benefits and trade-offs of low-carbon electricity supply options. Renewable and Sustainable Energy Reviews **76**, 1283–1290 (2017)

[18] Naegler, T., Buchgeister, J., Hottenroth, H., Simon, S., Tietze, I., Viere, T., Junne, T.: Life cycle-based environmental impacts of energy system transformation strategies for Germany: Are climate and environmental protection conflicting goals? Energy Reports **8**, 4763–4775 (2022). https://doi.org/10.1016/j.egyr.2022.03.14. Accessed 2023-01-16

[19] Junne, T., Simon, S., Buchgeister, J.,

Saiger, M., Baumann, M., Haase, M., Wulf, C., Naegler, T.: Environmental Sustainability Assessment of Multi-Sectoral Energy Transformation Pathways: Methodological Approach and Case Study for Germany. Sustainability **12**(19), 8225 (2020). https://doi.org/10.3390/su1219822

[20] Luderer, G., Gunnar Luderer, Gunnar Luderer, Luderer, G., Pehl, M., Arvesen, A., Gibon, T., Bodirsky, B.L., de Boer, H.S., Fricko, O., Hejazi, M., Humpenöder, F., Iyer, G., Mima, S., Silvana Mima, Mouratiadou, I., Pietzcker, R.C., Popp, A., van den Berg, M.P., van den Berg, M., van Vuuren, D.P., Hertwich, E.G.: Environmental co-benefits and adverse side-effects of alternative power sector decarbonization strategies. Nature Communications **10**(1), 5229–5229 (2019). https://doi.org/10.1038/s41467-019-13067-. MAG ID: 2990509501

[21] Heck, V., Donges, J.F., Lucht, W.: Collateral transgression of planetary boundaries due to climate engineering by terrestrial carbon dioxide removal. Earth System Dynamics **7**(4), 783–796 (2016). https://doi.org/10.5194/esd-7-783-2016. Publisher: Copernicus GmbH

[22] Heck, V., Hoff, H., Wirsenius, S., Meyer, C., Kreft, H.: Land use options for staying within the Planetary Boundaries – Synergies and trade-offs between global and local sustainability goals. Global Environmental Change **49**, 73–84 (2018). https://doi.org/10.1016/j.gloenvcha.2018.02.004

[23] Heck, V., Gerten, D., Lucht, W., Popp, A.: Biomass-based negative emissions difficult to reconcile with planetary boundaries. Nature climate change **8**(2), 151–155 (2018)

[24] Weidner, T., Guillén-Gosálbez, G.: Planetary boundaries assessment of deep decarbonisation options for building heating in the European Union. Energy Conversion and Management **278**, 116602 (2023)

[25] Conijn, J., Bindraban, P., Schröder, J., Jongschaap, R.: Can our global food system meet food demand within planetary boundaries? Agriculture, Ecosystems & Environment **251**, 244–256 (2018)

[26] Bowles, N., Alexander, S., Hadjikakou, M.: The livestock sector and planetary boundaries: A 'limits to growth' perspective with dietary implications. Ecological Economics **160**, 128–136 (2019). https://doi.org/10.1016/j.ecolecon.2019.01.03

[27] Hill, S.L.L., Gonzalez, R., Sanchez-Ortiz, K., Caton, E., Espinoza, F., Newbold, T., Tylianakis, J., Scharlemann, J.P.W., Palma, A.D., Purvis, A.: Worldwide impacts of past and projected future land-use change on local species richness and the Biodiversity Intactness Index. bioRxiv (2018). https://doi.org/10.1101/311787

[28] Allen, C., Metternicht, G., Wiedmann, T., Pedercini, M.: Modelling national transformations to achieve the SDGs within planetary boundaries in small island developing states. Global Sustainability **4** (2021). https://doi.org/10.1017/sus.2021.13. Publisher: Cambridge University Press. Accessed 2022-02-20

[29] IPCC: Summary for Policymakers. In: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA. doi: 10.1017/9781009157926.001 (2022)

[30] Byers, E., Krey, V., Kriegler, E., Riahi, K., Schaeffer, R., Kikstra, J., Lamboll, R., Nicholls, Z., Sandstad, M., Smith, C., van der Wijst, K., Al Khourdajie, A., Lecocq, F., Portugal-Pereira, J., Saheb, Y., Stromann, A., Winkler, H., Auer, C., Brutschin, E., Gidden, M., Hackstock, P., Harmsen, M., Huppmann, D., Kolp, P., Lepault, C., Lewis,

J., Marangoni, G., Müller-Casseres, E., Skeie, R., Werning, M., Calvin, K., Forster, P., Guivarch, C., Hasegawa, T., Meinshausen, M., Peters, G., Rogelj, J., Samset, B., Steinberger, J., Tavoni, M., van Vuuren, D.: *AR6 Scenarios Database hosted by IIASA*. International Institute for Applied Systems Analysis. doi: 10.5281/zenodo.5886911, url: data.ece.iiasa.ac.at/ar6/ (2022)

[31] Fricko, O., Havlik, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., Ermolieva, T., Forsell, N., Herrero, M., Heyes, C., Kindermann, G., Krey, V., McCollum, D.L., Obersteiner, M., Pachauri, S., Rao, S., Schmid, E., Schoepp, W., Riahi, K.: The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. Global Environmental Change **42**, 251–267 (2017). https://doi.org/10.1016/j.gloenvcha.2016.06.004

[32] Riahi, K., Bertram, C., Huppmann, D., Rogelj, J., Bosetti, V., Cabardos, A.-M., Deppermann, A., Drouet, L., Frank, S., Fricko, O., Fujimori, S., Harmsen, M., Hasegawa, T., Krey, V., Luderer, G., Paroussos, L., Schaeffer, R., Weitzel, M., van der Zwaan, B., Vrontisi, Z., Longa, F.D., Després, J., Fosse, F., Fragkiadakis, K., Gusti, M., Humpenöder, F., Keramidas, K., Kishimoto, P., Kriegler, E., Meinshausen, M., Nogueira, L.P., Oshiro, K., Popp, A., Rochedo, P.R.R., Ünlü, G., van Ruijven, B., Takakura, J., Tavoni, M., van Vuuren, D., Zakeri, B.: Cost and attainability of meeting stringent climate targets without overshoot. Nature Climate Change **11**(12), 1063–1069 (2021). https://doi.org/10.1038/s41558-021-01215-2

[33] Hasegawa, T., Fujimori, S., Frank, S., Humpenöder, F., Bertram, C., Després, J., Drouet, L., Emmerling, J., Gusti, M., Harmsen, M., Keramidas, K., Ochi, Y., Oshiro, K., Rochedo, P., van Ruijven, B., Cabardos, A.-M., Deppermann, A., Fosse, F., Havlik, P., Krey, V., Popp, A., Schaeffer, R., van Vuuren, D., Riahi, K.: Land-based implications of early climate actions without global net-negative emissions. Nature Sustainability **4**(12), 1052–1059 (2021). https://doi.org/10.1038/s41893-021-00772-w

[34] Bertram, C., Riahi, K., Hilaire, J., Bosetti, V., Drouet, L., Fricko, O., Malik, A., Nogueira, L.P., Zwaan, B.v.d., Ruijven, B.v., Vuuren, D.v., Weitzel, M., Longa, F.D., Boer, H.-S.d., Emmerling, J., Fosse, F., Fragkiadakis, K., Harmsen, M., Keramidas, K., Kishimoto, P.N., Kriegler, E., Krey, V., Paroussos, L., Saygin, D., Vrontisi, Z., Luderer, G.: Energy system developments and investments in the decisive decade for the Paris Agreement goals. Environmental Research Letters **16**(7), 074020 (2021). https://doi.org/10.1088/1748-9326/ac09ae

[35] Guo, F., van Ruijven, B., Zakeri, B., Krey, V., Riahi, K.: Global Energy Interconnection: A scenario analysis based on the MESSAGEix-GLOBIOM Model. Report, IIASA, Laxenburg, Austria (2021). Available at https://iiasa.dev.local/.

[36] Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D.L., Rao, N.D., Riahi, K., Rogelj, J., De Stercke, S., *et al.*: A low energy demand scenario for meeting the 1.5 c target and sustainable development goals without negative emission technologies. Nature energy **3**(6), 515–527 (2018)

[37] Soergel, B., Kriegler, E., Weindl, I., Rauner, S., Dirnaichner, A., Ruhe, C., Hofmann, M., Bauer, N., Bertram, C., Bodirsky, B.L., Leimbach, M., Leininger, J., Levesque, A., Luderer, G., Pehl, M., Wingens, C., Baumstark, L., Beier, F., Dietrich, J.P., Humpenöder, F., von Jeetze, P., Klein, D., Koch, J., Pietzcker, R., Strefler, J., Lotze-Campen, H., Popp, A.: A sustainable development pathway for climate action within the UN 2030 Agenda. Nature Climate Change **11**(8), 656–664 (2021). https://doi.org/10.1038/s41558-021-01098-3

[38] Anderson, K., Peters, G.: The trouble with negative emissions. Science **354**(6309), 182–183 (2016)

[39] Sen, A., Dabi, N.: Tightening the Net: Net

zero climate targets–implications for land and food equity. Oxfam (2021)

[40] Sanchez, D.L., Nelson, J.H., Johnston, J., Mileva, A., Kammen, D.M.: Biomass enables the transition to a carbon-negative power system across western north america. Nature Climate Change **5**(3), 230–234 (2015)

[41] Campbell, B.M., Beare, D.J., Bennett, E.M., Hall-Spencer, J.M., Ingram, J.S., Jaramillo, F., Ortiz, R., Ramankutty, N., Sayer, J.A., Shindell, D.: Agriculture production as a major driver of the earth system exceeding planetary boundaries. Ecology and society **22**(4) (2017)

[42] Lee, J., Bazilian, M., Sovacool, B., Hund, K., Jowitt, S.M., Nguyen, T.P., Månberger, A., Kah, M., Greene, S., Galeazzi, C.: Reviewing the material and metal security of low-carbon energy transitions. Renewable and Sustainable Energy Reviews **124**, 109789 (2020). Publisher: Elsevier

[43] Xu, L., Fuss, M., Maryegli Fuss, Poganietz, W.-R., Jochem, P., Schreiber, S., Zoephel, C., Brown, N.: An Environmental Assessment Framework for Energy System Analysis (EAFESA): The method and its application to the European energy system transformation. Journal of Cleaner Production **243**, 118614 (2020). https://doi.org/10.1016/j.jclepro.2019.11861

[44] Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction vol. 2. Springer, ??? (2009)

[45] Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(2), 301–320 (2005). https://doi.org/10.1111/j.1467-9868.2005.00503.

[46] Lu, C., Tian, H.: Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. Earth System Science Data **9**(1), 181–192 (2017). https://doi.org/10.5194/essd-9-181-2017. Publisher: Copernicus GmbH. Accessed 2023-08-28

Acknowledgments. I thank Sumit Kothari, François Lafond and Peter Barbrook-Johnson for useful comments. I also thank Cameron Hepburn and Doyne J. Farmer for inspiring discussions. This work was supported by the Oxford Martin Programme on the Post-Carbon Transition, the Institute for New Economic Thinking at the Oxford Martin School and Baillie Gifford. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract no. 2019-1902010003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorised to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

Appendix A Additional results for phosphorus use

Category	N. scenarios within P PB	N. scenarios not within P PB
C1: limit warming to 1.5°C (>50%) with no or limited overshoot	9	0
C2: return warming to 1.5°C (>50%) after a high overshoot	24	11
C3: limit warming to 2°C (>67%)	45	3
C4: limit warming to 2°C (>50%)	28	0

Table A1 Number of scenarios in each category classification for scenarios reporting P use. The second columns shows the number of scenarios that respect the P PB, while the last columns the number of scenarios that transgress it.

A.1 Principal component analysis and clustering

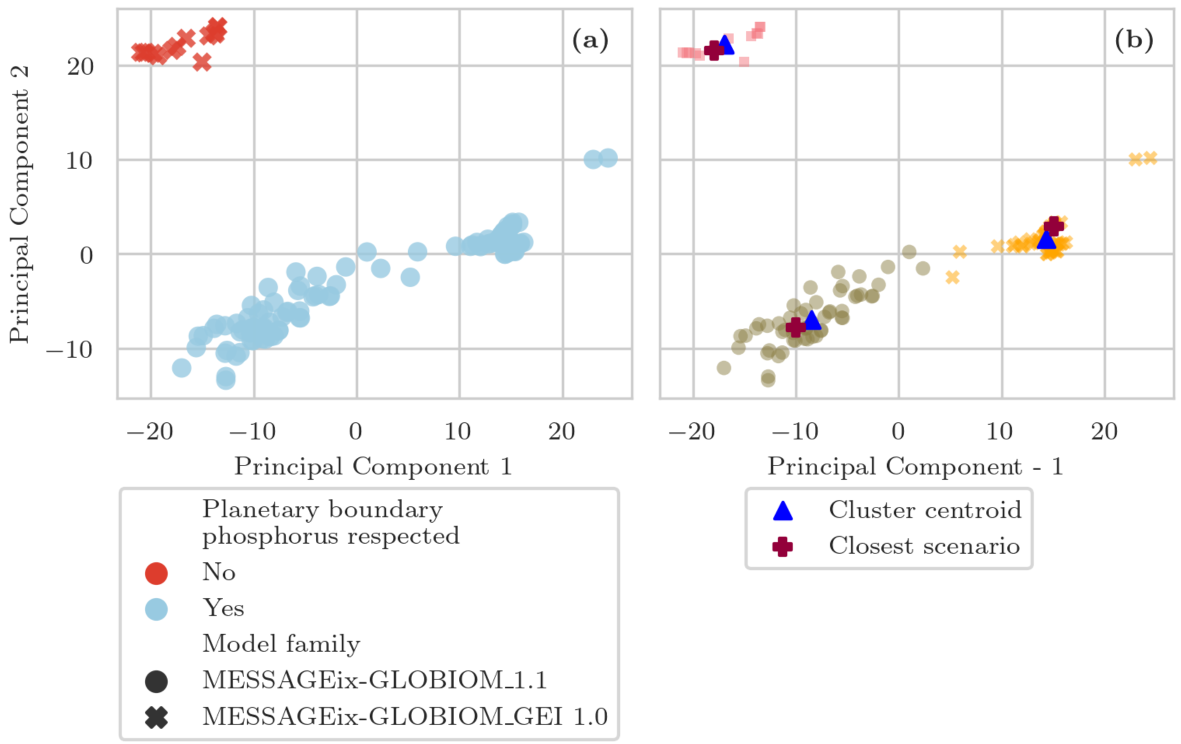

Fig. A1 **(a)** Visualisation of the scenarios that respect the phosphorus planetary boundary (light blue) and those that do not (red) in a low-dimensional space. We use Principal Component Analysis (PCA) and show the first principal component (PC, x-axis) and the second PC (y-axis). For the PCA we exclude all variables related to the "'AR6 climate diagnostics" and standardise the rest. The first PC explains 0.47 of the variation in the data, while the second PC explains 0.25. The scenarios that respect the P PB are part of a distinct cluster. **(b)** Scenarios that report P use coloured by cluster (pink, sand and yellow). Scenarios are clustered using k-means clustering, which we apply to the PCs; we set the number of clusters to 3. We show the centroid of each cluster (blue triangle) and the scenario closest to each centroid (red cross).

A.2 Elastic Net regressions

A.2.1 Hyperparameter tuning

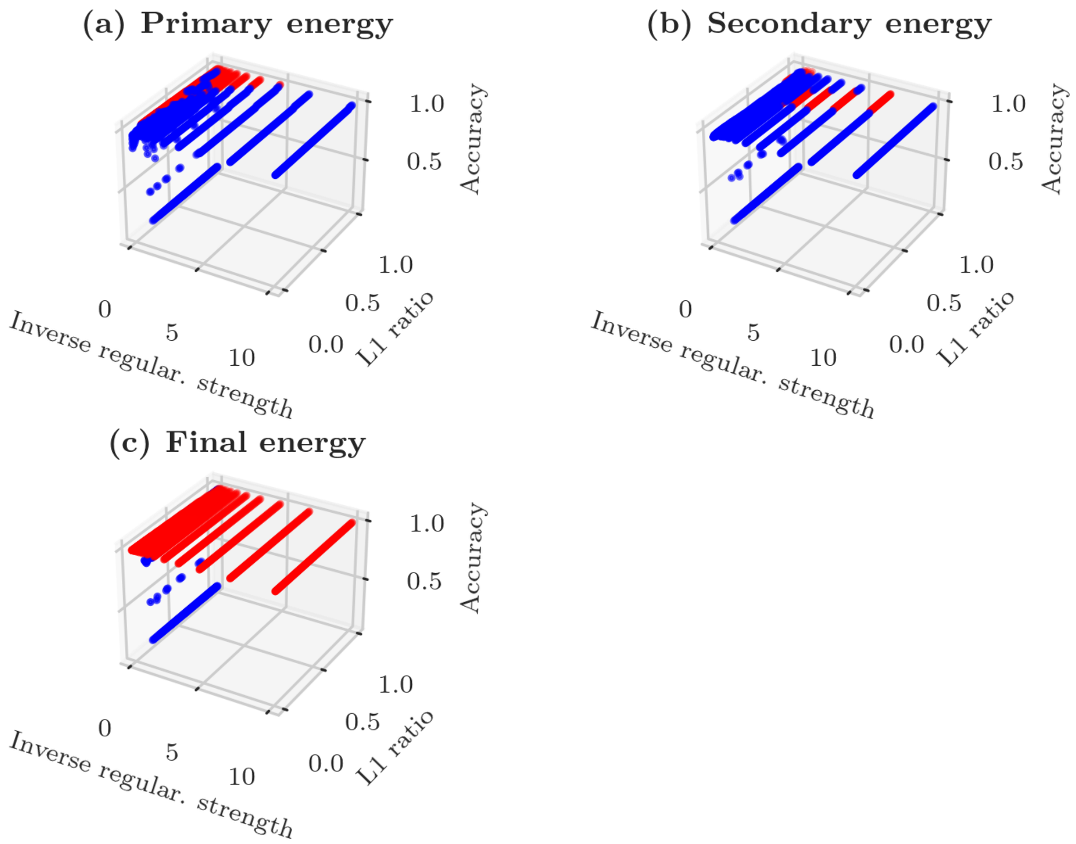

Fig. A2 Results for the hyperparameter tuning for the Elastic Net regressions used to predict whether P use transgresses or not the PB, for **(a)** primary, **(b)** secondary and **(c)** final energy. Dots in red show models with maximum accuracy for **(a)** and **(c)**, where it is 1, and between 1 and 0.98 for **(b)**. We chose a range of values for secondary energy because only two parametrizations lead to an accuracy of 1. The x-axis shows the inverse of the regularization strength; smaller values imply more regularization. The y axis shows the L1 ratio; if it is zero, the L2 penalty is used; if it is 1, the L1 penalty is used; values in between imply that a combination of the L1 and L2 penalties is used. The z-axis shows the accuracy.

A.2.2 Regression results

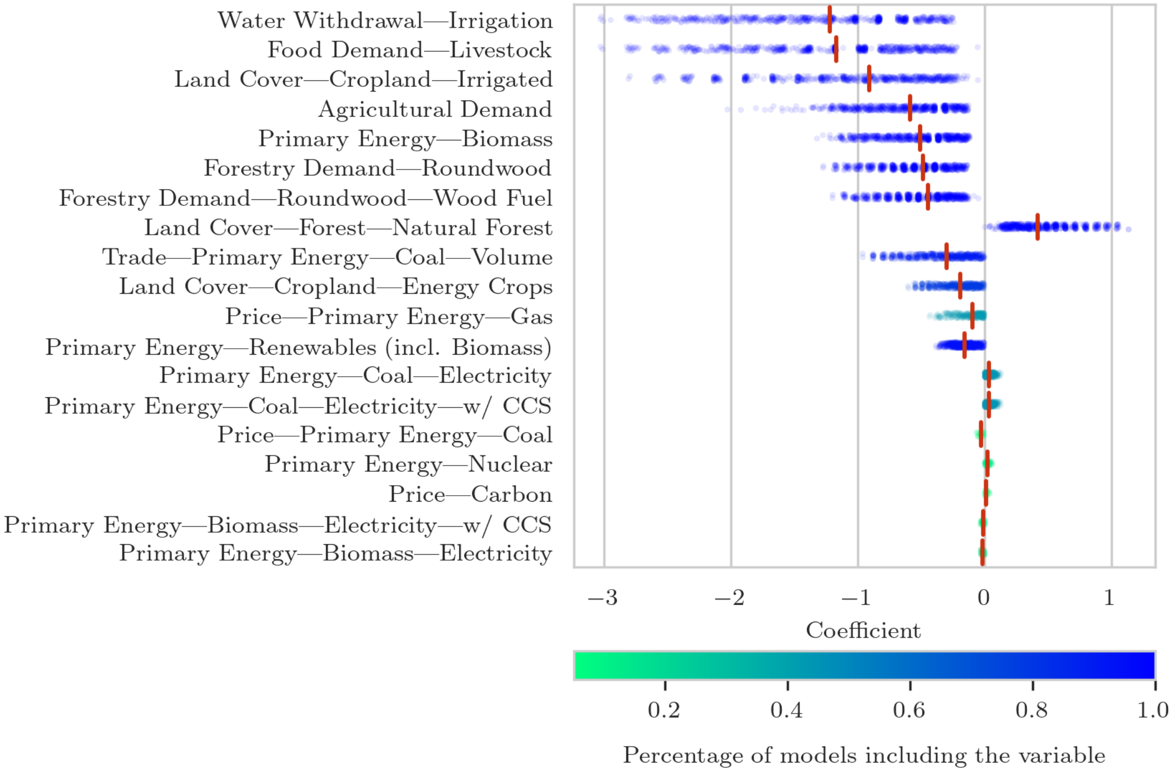

Fig. A3 Coefficient of the P PB regressions for primary energy. The red bar shows the median value. Points are colour-coded according to the percentage of models that include that variable. We fit a total of 621 models, corresponding to the hyperparameter values that achieve maxim accuracy (see Fig. A2).

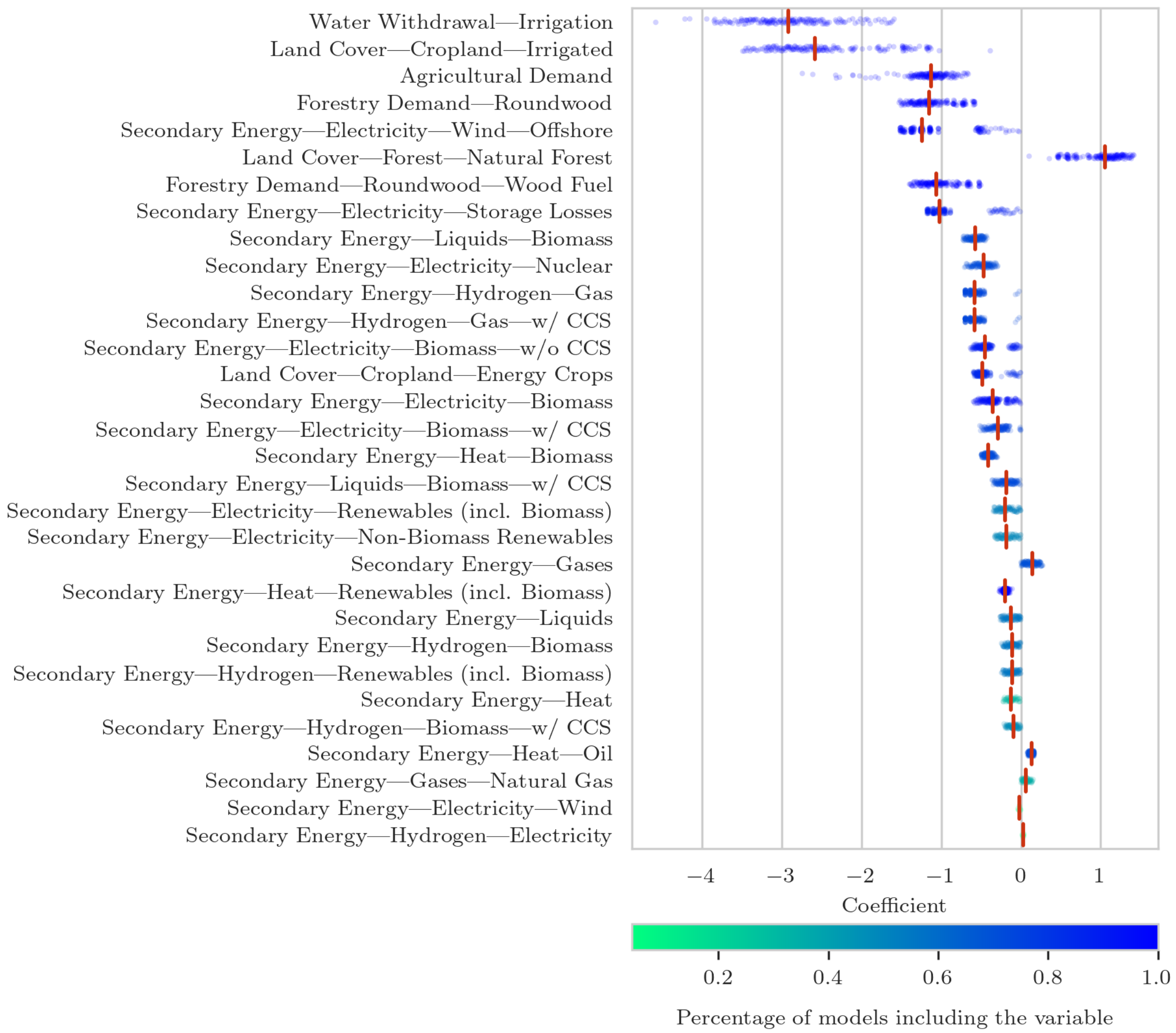

Fig. A4 Coefficient of the P PB regressions for secondary energy. The red bar shows the median value. Points are colour-coded according to the percentage of models that include that variable. We fit a total of 166 models, corresponding to the hyperparameter values that achieve the desired accuracy (see Fig. A2).

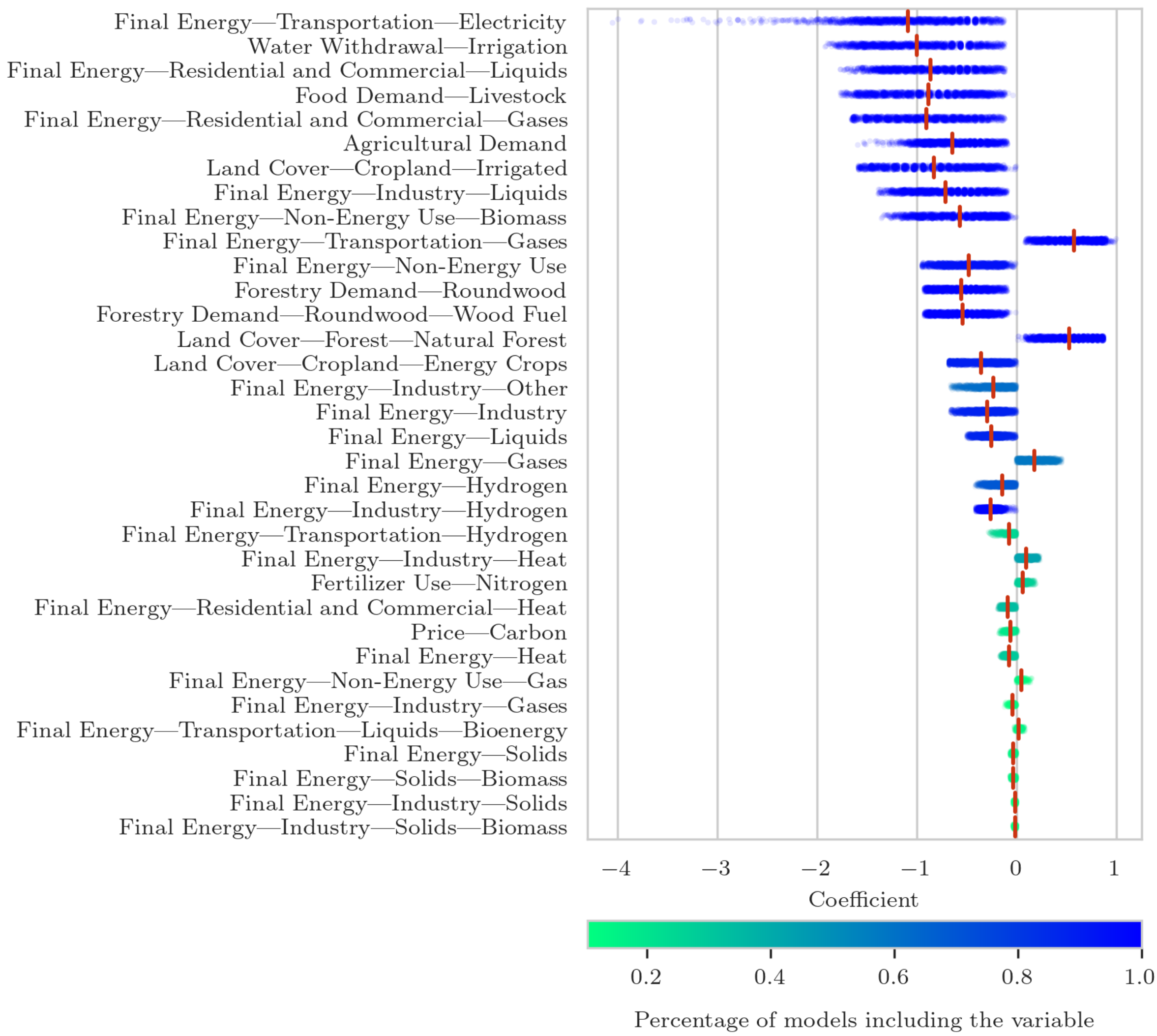

Fig. A5 Coefficient of the P PB regressions for final energy. The red bar shows the median value. Points are colour-coded according to the percentage of models that include that variable. We fit a total of 1,444 models, corresponding to the hyperparameter values that achieve maxim accuracy (see Fig. A2).

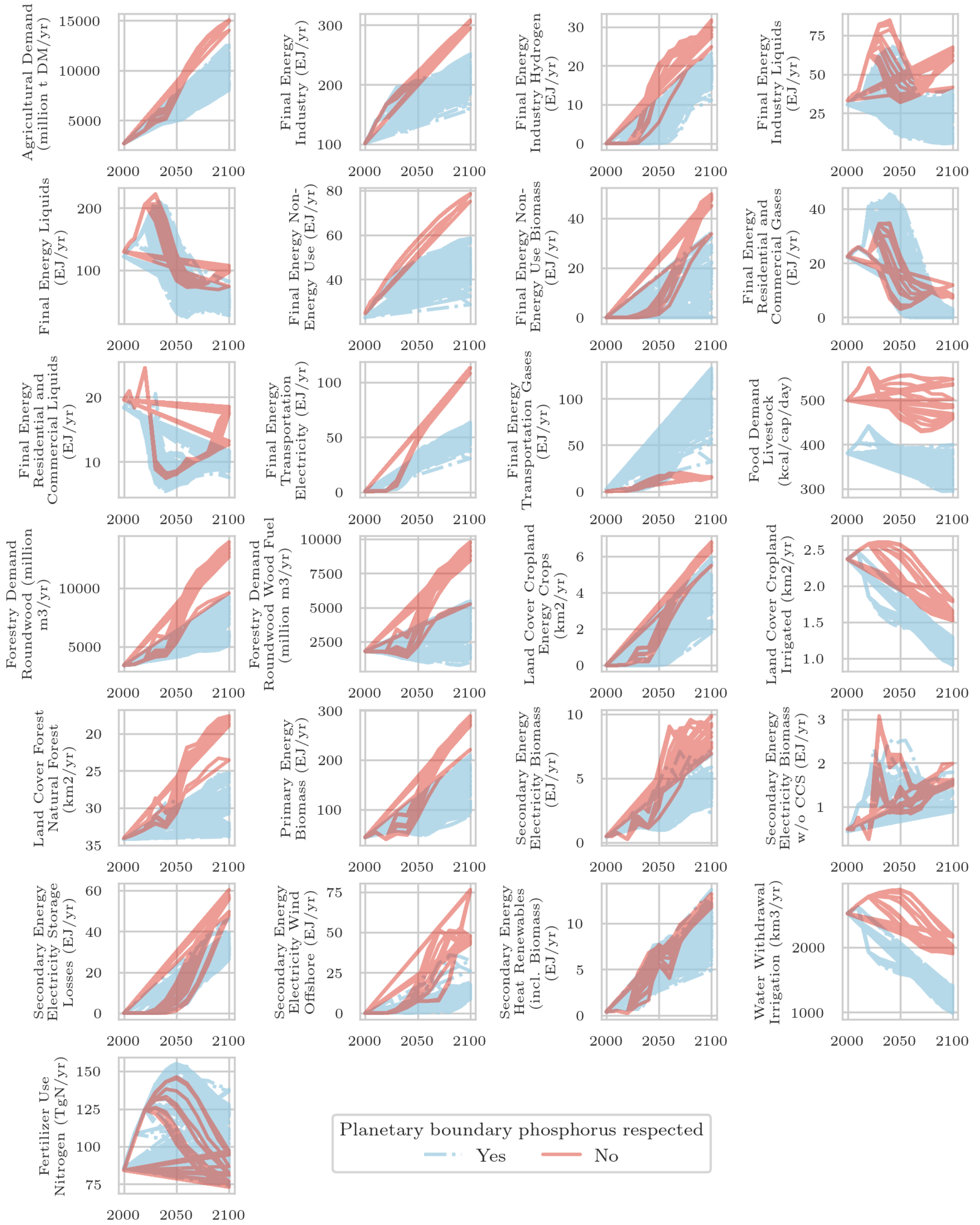

Fig. A6 Variables selected by the 3 Elastic Net regressions for scenarios that stay within the P PB (light blue) and scenarios that do not (red). We show variables that were selected by at least 80% of the models we fit; we also add P and N use.

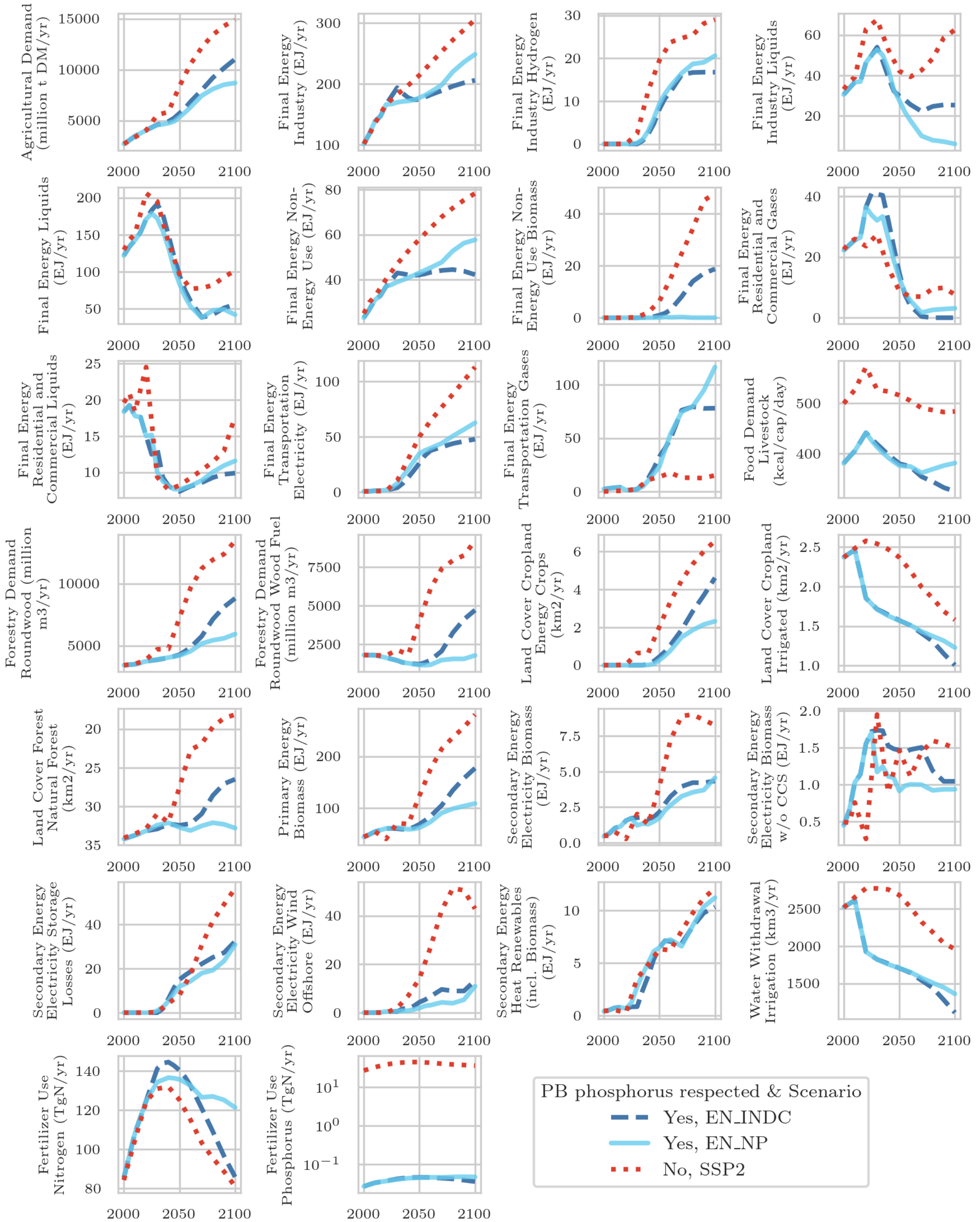

Fig. A7 Variables selected by the 3 Elastic Net regressions for 3 representative scenarios (see Fig. A1) that stay within the P PB(light blue) and scenarios that do not (red). We show variables that were selected by at least 80% of the models we fit; we also show N and P use. In the legend, "EN_INDC" refers to the `EN_INDCi2030_900f_NDCp` scenario and "EN_NP" to the `EN_NPi2020_900` scenario, both use MESSAGEix-GLOBIOM 1.1 [32–34]; "SSP2" refers to the `SSP2_openres_lc_120` scenario, which uses MESSAGEix-GLOBIOM GEI 1.0 [35].

Appendix B Comparison to empirical data

Fig. B8 compares the empirical time series of phosphorus, nitrogen and forest cover with those of the scenarios. We could not compare water consumption as models vary greatly in its coverage, with some, for example, covering only the energy sector, while others have a more extended treatment.

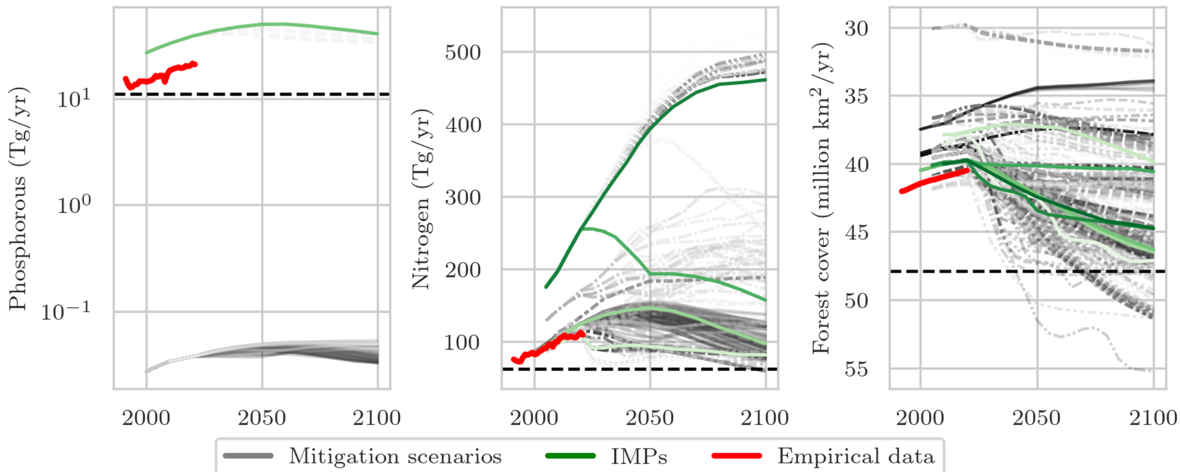

Fig. B8 Comparison of the scenario time series (green for the IMPs and grey for all other scenarios) with empirical data (red). Data for phosphorus and nitrogen consumption are taken from the International Fertilizer Industry Association (IFA, https://www.ifastat.org/databases/graph/1_1). For phosphorus, we follow [46] and use P_2O_5, which we convert to phosphorus by multiplying it by the ratio 62/142. Data for forest cover are taken from the World Bank (https://data.worldbank.org/indicator/AG.LND.FRST.K2) and it is defined as "Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens".

Firm-level production networks: what do we (really) know? *

Andrea Bacilieri[1,2], András Borsos[3,4],
Pablo Astudillo-Estevez[2,5], and François Lafond[1,6]

[1]Institute for New Economic Thinking, University of Oxford, Oxford, UK
[2]Smith School of Enterprise and Environment, University of Oxford, Oxford, UK
[3]Complexity Science Hub Vienna, Vienna, Austria
[4]Department of Financial Systems Analysis, Central Bank of Hungary, Budapest, Hungary
[5]Universidad San Francisco de Quito & School of Economics, Quito, Ecuador
[6]Mathematical Institute, University of Oxford, Oxford, UK

Abstract

Are standard production network properties similar across all available datasets, and if not, why? We provide benchmark results from two administrative datasets (Ecuador and Hungary), which are exceptional in that there is no reporting threshold. We compare these networks to a leading commercial dataset (FactSet) and published results on national firm-level production networks. Administrative datasets with no reporting thresholds have remarkably similar quantitative properties, while a number of important properties are biased in datasets with missing data.

Keywords: Production networks, input-output analysis, firm-level data.

JEL codes: C80, D57, L14

*We would like to thank the Servicio de Rentas Internas (SRI) and its Centro de Estudios Fiscales for providing the raw data of Ecuador for research purposes. András Borsos would like to thank INET for its hospitality. We would like to thank audiences at CCS 2022, ENBES, CSH Vienna, Harvard Growth Lab, ETH Zürich, Conference on Network Science and Economics, and Vasco Carvalho, Banu Demir, Christian Diem, Doyne Farmer, Kerstin Hötte, Luca Mungo and Anton Pichler for comments. This work was supported by Bailie Gifford, the Oxford Martin Programme on the Post-Carbon Transition and the Institute for New Economic Thinking at the Oxford Martin School. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract no. 2019-1902010003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorised to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. FactSet had the opportunity to review the FactSet data description. *Contacts*: andrea.bacilieri@sant.ox.ac.uk; pastudillo@usfq.edu.ec; borsosa@mnb.hu; francois.lafond@inet.ox.ac.uk

1 Introduction

Almost a century after Leontief's *The Economy as a Circular Flow* (1928), national Input-Output (I-O) tables are available for the large majority of advanced economies, have been harmonized and extended to international tables, and serve as the basis for environmentally-extended national accounts. These datasets continue to power the development of major macro-econometric and general equilibrium models used by policymakers across the world.

While these achievements are remarkable, these datasets remain highly aggregated, covering as few as 56 sectors in the World Input-Output Database (WIOD) and a maximum of 405 industries for the most disaggregated tables published by the US Bureau of Economic Analysis (BEA). In comparison, there are about 200 million firms in the world and 6 million in the US.[1] A host of recent papers have started to explore firm-level data on production networks, demonstrating its importance for stock co-movement (Cohen and Frazzini, 2008) and the propagation of shocks (Barrot and Sauvagnat, 2016; Demir et al., 2022; Carvalho et al., 2021; Diem et al., 2022).

In principle, firm-level data is likely to be much more useful than aggregate data since aggregation can create substantial biases (Morimoto, 1970). Even within fine-grained industry classification, firms differ in the extent to which they buy and sell from other industries so that industry-level shock propagation models will generally lead to biased results, especially when shocks do not affect all the firms within an industry in the same way (Diem et al., 2023).

Firm-level production network data is thus very useful in principle. But what data is available and how good is it? Are there generic properties of firm-level production networks that hold across all datasets? We address these questions through a detailed analysis of three important datasets: administrative VAT data from Ecuador and Hungary, and a leading commercial dataset covering large firms in the global supply chain network (FactSet). We complement these results with an extensive synthesis of the literature.

A crucial distinguishing feature of our administrative data is that, for all years in Ecuador and for the last year in Hungary, there is no reporting threshold so that we observe in principle the population of firm-to-firm transactions. We call these our *complete* datasets. As Figure 1 makes clear, the change in the reporting threshold in Hungary had a dramatic effect on the number of transactions in the network. Throughout the paper, we exploit the comparison between the early and recent years of the Hungarian dataset to understand the effect of the reporting threshold.

While several papers have acknowledged the bias due to specific forms of data truncation, and have even attempted to account for it in their analysis (Atalay et al., 2011; Herskovic et al., 2020), the "ground truth" remains unknown. It turns out that there is a remarkable similarity between our complete datasets for key network statistics, providing us with a credible benchmark of what we "really know" – properties of production networks that are very likely to be similar despite country heterogeneity. We compare these results with what we observe on non-complete datasets and interpret the difference as the bias due to incomplete reporting.

To give an example, the average number of suppliers (average degree) in both our complete datasets is around 40, but in our incomplete datasets, it is less than 10, so reporting thresholds strongly bias the observed mean degree downward. By contrast, the tail exponent of the distribution of the value of transactions (weights) does not dramatically change after the change in reporting requirements in Hungary, as can be seen in Figure 1 (for a summary of our results see Table 10 in Section 5).

A key message from our empirical investigation is that very large firms can have a very high number of customers, but not as high a number of suppliers. This is intuitive if we think that firms grow by extending their customer base, but to match the requirement for additional inputs, they buy more from their existing set of suppliers. We provide additional details on joint distributions of (in- or out-) strength (domestic B2B expenses and sales) and (in- or out-) degree (number of

[1] The World Bank reports that 4,304,326 new businesses were created in 2018 (`https://data.worldbank.org/indicator/IC.BUS.NREG`). OpenCorporates reports 201,708,765 as of 17 November 2021 (`https://opencorporates.com/`), while Statista reports around 210 million for 2018, 2019 and 2020 (`https://www.statista.com/statistics/1260686/global-companies/`). For the number of US firms, we used the Statistics of US Businesses dataset provided by the US Census Bureau (`https://www.census.gov/programs-surveys/susb.html`).

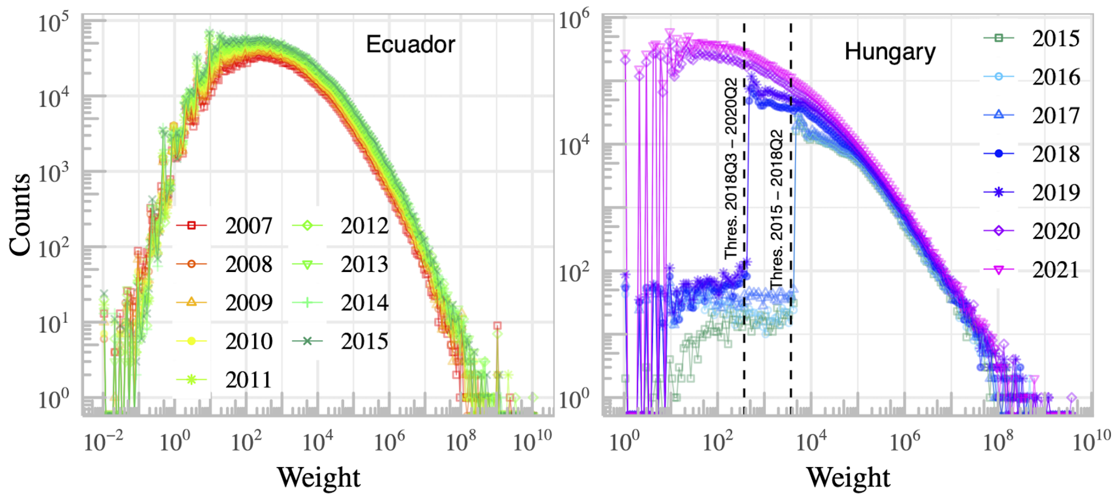

Figure 1: Distribution of the value of the transactions (weights) for Ecuador (left) and for Hungary (right) over time. The values are in USD for Ecuador and in 1,000 HUF (≈ $2.8) for Hungary. For Hungary, the two vertical lines mark the changes in the reporting threshold (i.e., "Thres."). The first threshold was effective from 2015 to the second quarter of 2018. The second threshold was effective until the second quarter of 2020, after which the threshold was removed. The values indicated by the vertical lines apply to the majority of the firms in the economy, but deviations arise depending on the tax rate firms are subject to; for more information see Appendix A.5. We used 200 log-spaced bins for both datasets.

suppliers and customers), documenting new facts. For instance, while the number of partners (customers or suppliers) increases with size (sales or expenses), this relationship features a strong heteroskedasticity: larger firms tend to have more partners, but they may well have very many or very few partners.

Taken together, our results provide the first comprehensive picture of the most fundamental statistics on production networks at the firm level and provide a crucial benchmark for all researchers and statisticians putting together these data. In contrast to other studies which interpret facts qualitatively (e.g., "firms with higher sales have more customers"), here we go beyond "stylized" facts and systematically provide clear *quantitative* estimates. Our results are thus helpful to all researchers who do not have access to administrative data but need key moments to calibrate their macro models or create synthetic datasets.

The paper is organized as follows. In Section 2, we provide a taxonomy of datasets that can be used to study firm-level production networks. Section 3 provides results on the binary structure of firm-level production networks, while Section 4 reports findings on the weighted networks. Section 5 discusses our results and Section 6 concludes.

2 Datasets

In an ideal case, data on firm-level production networks would be time-stamped transaction-level data with a distinction between price and quantities. While prices and quantities are available in rare cases (such as Belgium, Duprez and Magerman, 2018), most datasets contain either a money flow (how much firm j spends on inputs provided by firm i) or simply a binary indicator that i is a supplier of j. We conducted a comprehensive literature review and document the various types of datasets, along with the corresponding reference papers, in Table 1. We distinguish between national and global datasets.[2] Table 1 also provides information about the countries for which these datasets are available.

[2]There have also been a few studies using imports and exports data, which provides a picture of the trade network between a pair, or more, of countries. Table 1 does not include studies that have no information on intra-national production networks. We also omit product-level datasets built for life-cycle analysis and industry-level datasets, such as those available for the automotive industry.

Table 1: Taxonomy of production network datasets, with examples

Type and examples	**Weighted**	**Source(s)**
National		
Data collected for VAT purposes		
Ecuador	Yes	This paper; Mungo et al. (2023)
Hungary	Yes	This paper; Diem et al. (2022, 2023)
Belgium	Yes	Dhyne et al. (2015); Magerman et al. (2016); Dhyne et al. (2016); Dewachter et al. (2017); Dhyne et al. (2021, 2022); Duprez and Magerman (2018); Bernard et al. (2019)
Chile	Yes	Grigoli et al. (2023); Huneeus (2020)
Costa Rica	Yes	Alfaro-Urena et al. (2018, 2022)
Dominican Republic	Yes	Cardoza et al. (2020)
Kenya	Yes	Chacha et al. (2022a)
Turkey	Yes	Demir et al. (2022, 2021)
Rwanda	Yes	Spray and Wolf (2018)
Spain	Yes	Peydró et al. (2020)
Uganda	Yes	Spray and Wolf (2018); Spray (2017)
West Bengal	Yes	Kumar et al. (2021)
Data from payment systems		
Brazil	Yes	Silva et al. (2020)
Japan	Yes	Fujiwara et al. (2021)
Netherlands	Yes	Ialongo et al. (2022)
Data collected for providing business services		
Japan (Tokyo Shoko Research)	No	E.g., Saito et al. (2007); Konno (2009); Ohnishi et al. (2009, 2010); Fujiwara and Aoyama (2010); Carvalho et al. (2021); Inoue (2016); Furusawa et al. (2017); Lu et al. (2017); Zhigang et al. (2018); Yuichi et al. (2019); Bernard et al. (2019)
Japan (Teikoku Databank)	No	Mizuno et al. (2015)
US (Billtrust)	Yes	Costello (2020)
Global		
Data collected from financial reporting requirements		
FactSet	No	This paper and König et al. (2022); Taschereau-Dumouchel (2022) for the US
Bloomberg	Yes/No	E.g., Wu and Birge (2014); Wu (2016)
Compustat (S&P)	Yes/No	E.g., Cohen and Frazzini (2008); Atalay et al. (2011); Herskovic et al. (2020); Atalay et al. (2014); Carvalho and Voigtländer (2014); Barrot and Sauvagnat (2016); Wu and Birge (2014)
Capital IQ (S&P).	Yes/No	E.g., Chakraborty and Ikeda (2020)
Shipment data		
FactSet and S&P.	Yes	This paper and Wu (2016)
Import-export data		
All countries	Yes/No	Examples where matched to national networks: Dhyne et al. (2021); Duprez and Magerman (2018); Spray (2021); Demir et al. (2022); Huneeus (2020)

Table 2: Reporting thresholds by country.

Dataset	Year	Transaction size threshold Raw	Transaction size threshold % of GDPpp	Firm size threshold Raw	Firm size threshold /GDPpp	Source
Ecuador	2008–2015	0 USD	0.00	0 USD	0	This paper
Hungary	2015–2018	3703703 HUF	83.45	0	0	This paper
Hungary	2018–2020	370370 HUF	7.46	0	0	This paper
Hungary	2021	1000 HUF	0.02	0	0	This paper
Belgium	2002-2014	250 EUR	0.70			Bernard et al. (2022)
Costa Rica	2008–2015	2500000 CRC	40.56			Alfaro-Urena et al. (2022)
Domin. Rep.	2012–2017	0 DOP	0.00	0 DOP	0	Cardoza et al. (2020)
Chile	2003–2011	0	0.00	250m CLP	28	Huneeus (2020)
Kenya	2019	0	0.00	5m KES	25	Chacha et al. (2022a)
Spain	2008–2009	3005 EUR	13.05			Peydró et al. (2020)
Turkey	2010–2014	5000 TRY	19.01			Demir et al. (2022)

Notes: The table shows the official reporting thresholds as gathered from the literature, omitting details of each country's idiosyncratic rules. The thresholds on the value of transactions are also shown as a share of GDP per person, taking World Bank data for the last year of the column 'Year'. The thresholds on firm size are also shown as renormalized by GDP per person. The table does not consider thresholds imposed by researchers (e.g., removing small firms)

National datasets. The main source of National datasets is *data collected for VAT purposes*, which mainly gathers supplier-customer relations among firms registered within the country. These datasets usually record money flows, making them well-suited to study the economy from the bottom up. Usually, there is a threshold below which transactions are not reported. For Belgium, this is €250, even though smaller transactions might be reported. The firms and operations exempted from VAT declarations include microenterprises, medical and socio-cultural activities and any financial transactions (Dhyne et al., 2015). Nevertheless, Dhyne et al. (2015) report that for 2012, the revenues of firms in the network represent 95% of national gross output.

A second major source of data is *payment systems*. Brazil collects firm-to-firm transactions through two real-time gross settlement systems provided and operated by the central bank of Brazil (Sistema de Transferência de Reservas and Sistema de Transferência de Fundos). These two datasets collect customer-supplier relations via wire transfers made by firms through their banks (Silva et al., 2020), with no threshold. Silva et al. (2020) reported that the 2014 value of all recorded transactions was 20 times the value of national GDP. Fujiwara et al. (2021) constructs a network from payments among Japanese firms that hold an account with the regional bank Shiga and Ialongo et al. (2022) construct a network from payments made between clients of a bank (they study two banks, ABM AMRO Bank NV and ING Bank NV). These datasets record payments, which may or may not be associated with an economic transaction.

A third source of data is *credit rating companies*. A prominent example is the production network data for Japan collected by two private companies for credit rating purposes and company credit reports (Tokyo Shoko Research, Ltd, and Teikoku Databank, Ltd). When rating and advising firms, these companies collect information on suppliers and customers but do not keep track of the money flows. Depending on the credit rating company, firms are asked to list up to 24 or 60 of their suppliers and customers. Therefore, in- and out-degrees have an artificial cut-off. Providing also company credit report services, Credit2B (acquired by Billtrust) collects supplier-customer transactions, likely for US firms only (Costello, 2020).

Our two national datasets: Ecuador and Hungary. In this paper, we analyse two national datasets: Ecuador (2007–2015) and Hungary (2015–2021). Ecuador requires VAT filings from both firms and natural persons, but we use data on firms only; see Appendix A.4 for more information. For these firms, we know some characteristics, like the sector and location, but we do not have access to firms' revenues, total expenditure or any other financial variable. The monetary values of the transactions are in US dollars.

Hungary's network is collected by the National Tax and Customs Administration of Hungary. Up until the first half of 2020, Hungary required firms to report transactions that exceeded a threshold (which changed over time); for more information on the threshold and Hungary's dataset see Appendix A.5. Although it covers the period 2014–2021, we dismissed the first year because the data quality is poor; this might be due to the inexperience of both the authorities and firms in the new reporting requirements. For most Hungarian firms, we have access to financial statements. The transactions are expressed in 1,000 HUF ($\approx$ \$2.8).

Global datasets. The datasets with global coverage cover mainly listed firms, which account for a large portion of gross output. A key source for this data stems from US Financial Accounting Standards, which require publicly traded firms to report customers that account for 10% or more of their annual revenues – formally called *major customers*. Due to the data collection process, coverage is biased towards companies listed on US stock exchanges. Although companies might report customers that account for less than 10%, this threshold skews the type of relations observed.

Standard and Poor (S&P) provides this information in two datasets: Compustat and Capital IQ. Capital IQ provides information on over 60,000 publicly traded companies worldwide, while Compustat tracks the order of a thousand firms and links per year (Cohen and Frazzini, 2008). Compustat is solely based on relations derived from the disclosure of major customers. Other data providers such as Bloomberg and FactSet collect additional information on supply chains by looking at annual filling/reports, investor presentations, company websites and press releases. As a result, these datasets are still biased in terms of the kind of transactions and companies they keep track of but are much more comprehensive. Comparing these datasets, Wang (2018) found that Bloomberg was the most comprehensive, but it is not possible to observe the network over time and it is also very difficult to access the bulk data. Importantly, the vast majority of these network data is unweighted.

Our global dataset: FactSet. We use FactSet Supply Chain Relationships, which we merge with supply chain relations derived from shipment data (Supply Chain Shipping Transactions). Shipment data are collected daily from the US Bill of Lading required for all seaborne trade; it covers private and listed firms. Considering only companies in our final network and for which we have sales, these represent, on average over time, 43% of US gross output and 29% of world gross output (Figure A.1 and more details in Appendix A.2). By comparison, Atalay et al. (2011) reports that Compustat accounts for 30% of US gross output in the year 1997.

Data cleaning and representativity. For FactSet, Hungary and Ecuador, we keep only firms in the giant weakly connected component (GWCC). Two firms are in the same weakly connected component if they are connected by at least one path, where the direction of edges is ignored. Table A.1 shows that this procedures leaves the number of firms or links virtually unaffected in our complete datasets, but removes 8% of firms and 2% of edges in FactSet.

Aggregating firm-level data does not necessarily lead to quantities comparable to National accounts (see Appendix A.1). Nevertheless, Figure A.1 provides a comparison of firms' sales to gross output (Factset and Hungary), and firm-to-firm sales to the sum of intermediate and investment sales (Ecuador and Hungary). The aggregate value of firms' sales is usually higher than in the national accounts, except for FactSet which covers roughly half of US and a third of world's gross output.

3 Results on binary networks

In this section, we discuss binary network metrics, such as the number of nodes, the average degree, the degree distributions, the correlations between in- and out-degree and degree assortativity. We then describe local patterns, specifically reciprocity, clustering coefficients and average path length.

Here and throughout the paper, our figures usually show all the years for which we have data, from which it is clear that most properties are highly stable over time, except for Hungary due to the change of the reporting threshold. Consequently, when we report results in tables, we report only the last year of Ecuador and FactSet, and 3 different years of Hungary, corresponding to years where different reporting thresholds were in place.

3.1 Density and growth

How many suppliers and customers do firms have? As we will see in the next section, this varies a lot across firms and scales with their size. But before discussing dispersion, we provide detailed statistics about the average because it highlights very well the heterogeneity of the datasets.

We define the *in-degree* k_j^{in} as the number of suppliers of firm j and the *out-degree* k_i^{out} as the number of i's customers. The *average degree* is given by

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i^{\text{in}} = \frac{1}{N}\sum_{i=1}^{N} k_i^{\text{out}},$$

where N is the number of firms.

Mean degree is highly heterogeneous across datasets. We would not expect that the average number of suppliers or customers of firms would differ dramatically across various economies. As a result, heterogeneity in the mean degree helps us to characterise heterogeneity across datasets due to data collection and data cleaning methods. Figure 2 shows the average degree for 12 datasets, often with several data points per dataset corresponding to different years or papers. The mean degree varies from less than 3 to around 50, more than an order of magnitude difference.

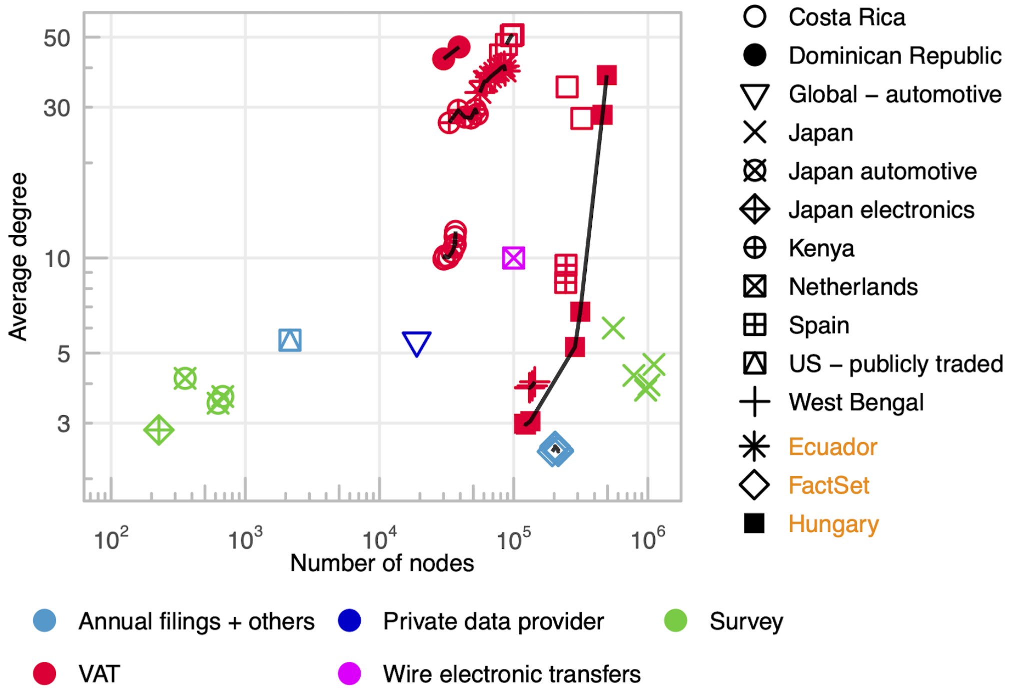

Figure 2: Number of nodes and average degree over time on a log-log scale for Ecuador, Hungary, FactSet and the networks in the literature we reviewed. Colours refer to the data collection method. Names in orange correspond to networks analysed in this paper and in black are the data taken from the literature. See Table C.1 for a list of the networks in the literature we reviewed. We did not include networks that were pooled over years. Connected dots belong to a dataset that has a consistent cleaning procedure over time.

To some extent, this appears to be due to the data collection method. VAT-based datasets (Kenya, Belgium, Ecuador, Hungary, Spain, the Dominican Republic and Costa Rica) have a fairly high degree. The years where Hungary has a low degree correspond to years where a high reporting threshold was in place. The noticeably lower average degree for Spain is likely due to the somewhat high reporting threshold at €3,005 (compared to €250 in Belgium, and 0 in Ecuador and Hungary in 2021), as is the case for Costa Rica (around US$4,200). Datasets collected by private companies (Japan, FactSet and the four smallest networks) tend to have a much lower average degree. The datasets based on transactions in two Dutch banks appear to be in-between – these datasets include only the transactions between accounts within the same bank, so while these banks are large, the data is substantially truncated.

Mean degree tends to increase with network size in the time-series dimension. We might expect that the mean degree would increase with the total number of nodes in the network, both in the real data (i.e., for economic reasons, as firms might choose more partners if there are more partnering opportunities) and in the observed sample (i.e., for statistical sampling reasons, as each time a new node becomes observed, there is a chance that a previously unobserved edge pointing to an existing node becomes observed as well).

Figure 2 (see also Table C.1) presents a mixed picture. The overall cross-sectional relationship is very noisy as small datasets tend to have a smaller mean degree, but the mean degree varies a lot in larger datasets because non-VAT datasets are able to sample many firms but relatively few edges. The time series dimension of each dataset, while very short, provides good evidence that the mean degree increases with size as

$$\bar{k}_t \sim N_t^\eta, \tag{1}$$

as commonly observed in growing networks (Dorogovtsev and Mendes, 2003). From a standard diversification argument, firms with more partners are less volatile, so η is critical to understand aggregate fluctuation since it determines the network sparsity (see Herskovic et al. (2020), who compute $\eta = 0.13$ using Compustat). Table 3 shows estimates of η, taking only years where the datasets are comparable (similar reporting threshold, similar cleaning). While it is difficult to base conclusions on time series of as low as 3 observations, Equation 1 with $\eta > 0$ appears a good hypothesis for administrative datasets, but not for FactSet. In the last column, we pool all the datasets except FactSet and run a standard panel regression with individual fixed effects, leading to a 95% confidence interval for η equal to [0.2, 0.42].[3]

Table 3: Mean degree and network size

	Dependent variable: $\log \bar{k}$							
	Belgium	Costa Rica	Ecuador	FactSet	Hungary	Kenya	West Bengal	Fixed eff.
$\log N$	0.66	0.75***	0.37***	−0.11	0.20	0.10	0.46**	0.31***
	(0.11)	(0.19)	(0.05)	(0.17)	(0.12)	(0.08)	(0.08)	(0.05)
Constant	−3.69	−5.50**	−0.48	2.20	−1.22	2.23*	−4.08**	
	(1.21)	(1.99)	(0.57)	(2.12)	(1.42)	(0.90)	(0.89)	
Obs.	3	8	9	7	3	6	4	35
R^2	0.98	0.72	0.88	0.07	0.73	0.27	0.95	0.54

Notes: See Figure 2 and Appendix C.1 for data sources. For Hungary, we run the regression only for the years 2015–2017, where the reporting threshold did not change. The Fixed Effects model excludes FactSet but includes the two points we have for Dominican Republic (which implies a slope of 0.33). Standard error in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

[3]The unweighted average of the 7 country-level estimates is 0.41 (CI [0.19, 0.63]). We do not include the two observations we have for Spain as they lead to an implausible $\hat{\eta} = 19$.

3.2 Degree distributions, correlations and assortativity

3.2.1 Degree distributions

Like many other networks, production networks tend to have very broad degree distributions: most nodes have a very low degree while some nodes have a very high degree. Nodes with a very high degree may act as "hubs" in the diffusion of shocks. These distributions tend to look linear on a log-log plot, both for the number of suppliers and customers.

Statistical framework. We will use the word *heavy-tailed* to describe distributions that have a complementary cumulative distribution function (CCDF) with a tail that decays slower than an exponential distribution (e.g., lognormal, power-law and Lévy, Voitalov et al., 2019).

An important question is whether these distributions belong to the class of regularly varying distributions, a sub-class of heavy-tailed distribution that can feature power-law tails, so that some of their moments may not exist. In practice (see Appendix B.1 for details), we want to test whether the share of nodes with degree greater than k, $P(k)$, is a regularly varying distribution, that is

$$\mathrm{Prob}(X > k) \equiv P(k) = \ell(k)k^{-\gamma}, \tag{2}$$

where $\ell(k)$ is a slowly varying function. Regularly varying distributions, which are all power-laws asymptotically, have infinite variance if $\gamma \leq 2$. From a statistical point of view, the tail exponent is the key quantity used to characterize the behaviour of regularly varying distributions. It is also the key statistic of interest in many applications. For instance, the theory of Acemoglu et al. (2012) predicts that idiosyncratic shocks average out at a slower rate than predicted by the central limit theorem ($N^{-1/2}$) if the tail exponent of the distribution of network centralities is less than 2, as we confirm in Section 4.4. The lower the exponent is, the higher the probability of finding extremely central firms, and the higher aggregate fluctuations are.

Generally speaking, we expect that in the future many models of production networks will either derive the values of the exponents (of various distributions) based on primitives or make predictions about economic outcomes that depend on the estimated exponents. In this paper, we will seek to characterise these exponents systematically, for unweighted and weighted quantities.

Throughout the main body of the paper, we report the Hill estimator from Clauset et al. (2009), which we call `plfit`, because it is standard in the literature. However, we check all our results using the state-of-the-art implementations of estimators of the tail index of Generalized Extreme Value Distributions (GEVD) provided by Voitalov et al. (2019). This approach allows us to (mostly) avoid a debate on the relative quality of the fit between the power-law and other distributions that may have heavy tails. See Appendix B.1 for details.

Number of customer-only or supplier-only firms. Table 4 shows the share of firms which are supplier only ($k^{\text{in}} = 0$) or customer only ($k^{\text{out}} = 0$). These proportions differ across our datasets. Even considering only our complete datasets, they are almost twice as large in Ecuador compared to Hungary. The VAT datasets reported in the literature suggest that the share of customer-only firms is substantially higher than the share of supplier-only firms.[4] Taken at face value, this means that the share of firms with no domestic non-labour inputs is much smaller than the share of firms with no domestic business customers. We find a similar result for Ecuador, but not for Hungary. For non-VAT dataset, the shares of customer-only and supplier-only appear comparable.

Full distribution. Figure 3 shows the in- and out-degree empirical CCDF of Ecuador, Hungary and FactSet. If the distribution has a perfect power-law tail, the tail of the CCDF $P(k)$ will appear linear on a log-log scale with slope $-\gamma$. Visually, it appears that all distributions display heavy tails. There is a striking difference between the distribution of the number of suppliers (left), where the maximum for our complete datasets is around $10^3 - 10^4$, and the distribution of the number of

[4] An exception is Uganda where Spray (2021) reports 87,000 suppliers but only 13,000 customers. We do not report Uganda in Table 4 because we do not know the total number of firms.

Table 4: Share of customer-only or supplier-only firms

Dataset	Year	Supplier-only	Customer-only	
Ecuador	2015	15.3	20	This paper
Hungary	2021	19.4	13.7	This paper
Hungary	2019	28.1	18.2	This paper
Hungary	2015	36.6	21.1	This paper
FactSet	2020	41.5	44.2	This paper
Belgium	2012	0.1	15.4	Magerman et al. (2016)
Costa Rica	2008–2015	9.7	30.4	Alfaro-Urena et al. (2018)
Dominican Rep.	2012–2017	3	18	Cardoza et al. (2020)
Spain	2009	8	24	Peydró et al. (2020)
US listed	04/2012–06/2013	30	27	Wu and Birge (2014)

Notes: Percentage of firms with no suppliers ('customer-only') and no customers ('supplier-only'). All values are in percent. In our data, these are shares of firms within the largest weakly connected component.

customers (right), where the maximum is an order of magnitude higher ($10^4 - 10^5$), which is just an order of magnitude below the number of nodes in the dataset (Figure 2, $10^5 - 10^6$). In other words, the largest firms are selling to a very big portion of the firms in the economy, while they are buying from just a fraction of all the firms.

These findings confirm an industry-level result discussed in Carvalho and Tahbaz-Salehi (2019): it is uncommon to find industries with many suppliers, but some industries provide almost universal inputs. At the firm level, Figure 1a in Grigoli et al. (2023) for Chile, Figure A1 in Cardoza et al. (2020) for the Dominican Republic, Figure 11 in Chacha et al. (2022b) and Figure 2 in Alfaro-Urena et al. (2018) also show that the range of the distribution of the number of customers tends to be much higher than the range of the distribution of the number of suppliers. In these four cases, as in our data, the maximum number of buyers is roughly an order of magnitude higher than the maximum number of suppliers.

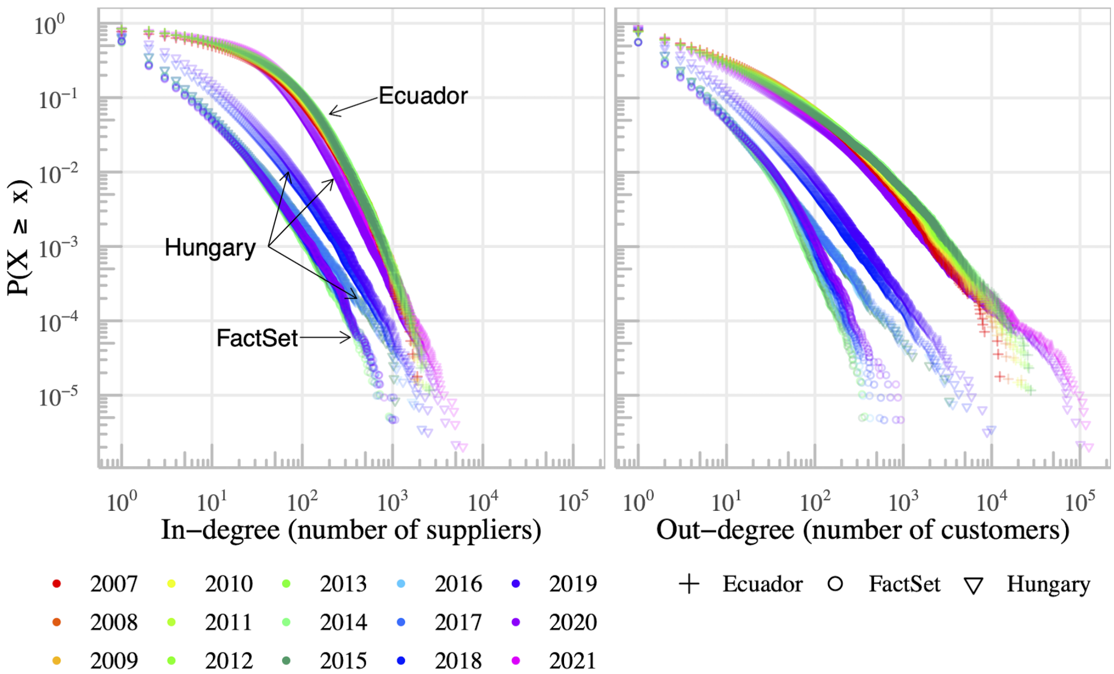

Figure 3: Empirical CCDF of number of suppliers (left) and number of customers (right) over time for the three networks we study. We compute the CCDF as $\bar{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \geq x)$, where $\mathbf{1}$ is the indicator function.

Table 5 shows estimates of the power-law exponents of the degree distributions and confirms the difference we observed between the in- and out-degrees. The distributions of the number of customers exhibit an exponent $\gamma < 2$, with FactSet, the commercially collected data with fewer firms and links, being the only exception (the moment estimator for Ecuador in 2007–2008 is also slightly above 2). By contrast, the distributions of the number of suppliers do not clearly feature power-law tails with a divergent second moment and some might not even feature power-law tails. The results on the distribution of the number of suppliers are not consistent across the different networks. On the one hand, for most datasets, Table 5 suggests $\gamma < 2$. Three studies even find $\gamma < 1$, which would imply an infinite mean, but these studies do not document their estimation methodology; we find these values implausibly low. On the other hand, for our "complete" datasets, we find $\gamma > 2$, so that the distributions are likely to be power-laws with *finite* variance. For the Ecuador dataset, the estimators based on the GEVD (Tables C.2) are often in the region $3 > \gamma > 4$. Thus, while the distribution of the number of suppliers may still be regularly varying, we conclude that it has finite second moments.

To sum up, the degree distributions suggest an interesting difference between the number of customers and suppliers of large firms, whereby while it is possible to have a very large number of customers, the number of suppliers remains somewhat limited in comparison.

Table 5: Tail exponents for the degree distributions.

Dataset	Year	In-degree (n. suppliers)	Out-degree (n. customers)	Source
Ecuador	2015	2.38	1.59	This paper
Hungary	2021	2.69	1.42	This paper
Hungary	2019	1.83	1.62	This paper
Hungary	2015	1.62	1.46	This paper
FactSet	2020	1.72	2.36	This paper
Japan	2005	1.37	1.46	Bernard et al. (2019)
Japan	2005	1.37	1.25	Ohnishi et al. (2010)
Japan	2006	1.35	1.26	Fujiwara and Aoyama (2010)
Netherlands Bank 1	2019	1.44	1.28	Ialongo et al. (2022)
Netherlands Bank 2	2019	1.77	1.31	Ialongo et al. (2022)
Chile	2019	0.28	0.40	Grigoli et al. (2023)
Dominican Republic	2017	0.30	0.43	Cardoza et al. (2020)
Costa Rica	2008–2015	0.58	0.73	Alfaro-Urena et al. (2018)
US listed	04/2012–06/2013	2.76	1.88	Wu and Birge (2014)
US listed	1979–2007	1.00		Atalay et al. (2011)
US listed	1978–2013	1.25	1.44	Barrot and Sauvagnat (2016)
FactSet US	2016	0.97	0.83	Taschereau-Dumouchel (2022)

Notes: Most studies use `plfit` (Clauset et al., 2009). Bernard et al. (2019) regress the log CCDF on the log degree, and a few studies appear to adopt their method (Barrot and Sauvagnat, 2016; Alfaro-Urena et al., 2018; Cardoza et al., 2020; Grigoli et al., 2023). Taschereau-Dumouchel (2022) uses the rank 1/2 estimator of Gabaix and Ibragimov (2011). The first two lines are our "complete" networks.

Correlations between in- and out-degrees Do firms with more customers also tend to have more suppliers? Figure 4 shows that yes, in- and out-degrees are positively correlated. However, we have seen previously that while large firms can have a lot of customers, they hardly have as many suppliers. This suggests that for each doubling of the number of customers, we should see less than a doubling of the number of suppliers; that is, the slope of the in-degree $\sim$ out-degree relationship should be less than 1 (equivalently, the slope of the out-degree $\sim$ in-degree relationship should be more than 1). To quantify this slope, we use Total Least Squares, which finds the line that minimizes the squared residuals measured as the perpendicular distance to the line. In contrast to regressions, it is symmetric (see Appendix B.2.1; in Appendix B.2.2 we provide covariance matrices so regression coefficients and R^2 can be retrieved from there).

We find that, indeed, the slope is substantially less than 1. Taking the value of 0.63 for Hungary, firms with 10 times more customers have only 6.3 times more suppliers. Since firms with more sales have, roughly proportionately, more expenses (Appendix C.4), it suggests that firms grow at the extensive margin on the customer side and at the intensive margin on the supplier side. Broadly speaking, and if we are prepared to make a time series interpretation of our cross-sectional results, firms grow mostly by acquiring more customers but get their extra inputs mostly from existing suppliers.

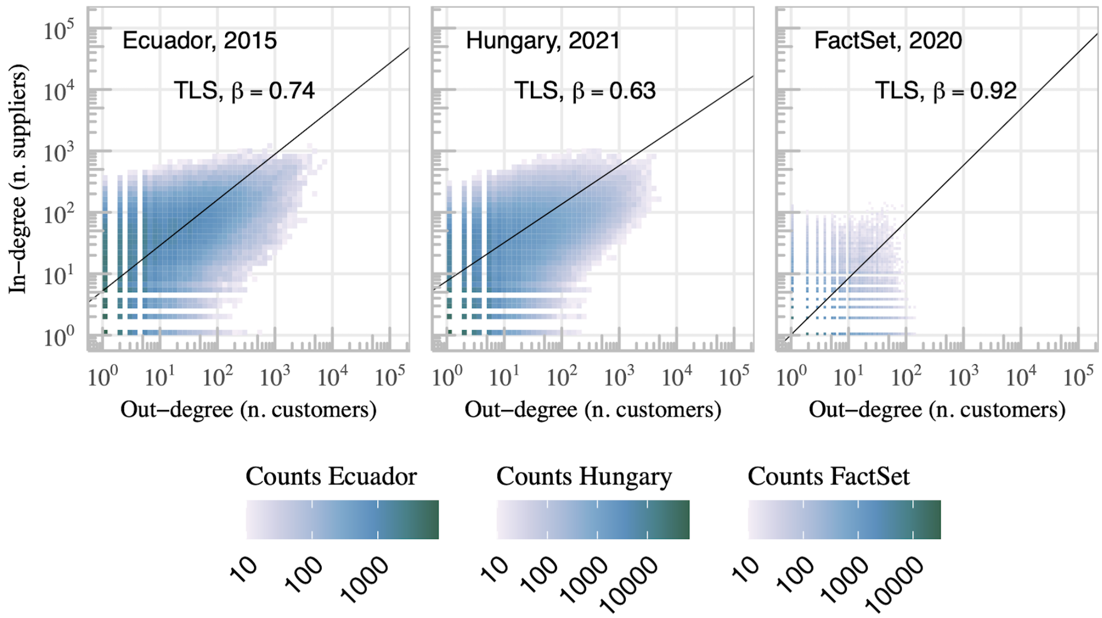

Figure 4: 2-D histogram for the number of customers and suppliers. We divide each axis into 60 log-spaced bins and then count the number of data points falling in each square. We do not show squares that have less than 10 observations. TLS stands for total least squares and β is the estimated coefficient. The TLS fit is shown by the black line.

3.2.2 Assortativity

An interesting hypothesis in the literature is that supply chains are characterized by negative degree assortativity; that is, highly connected firms tend to be connected with less connected firms (Bernard et al., 2019; Fujiwara and Aoyama, 2010; Bernard et al., 2022; Lim, 2018; Alfaro-Urena et al., 2018). Given the highly heterogeneous degree distribution, negative assortativity can be the symptom of nestedness,[5] whereby large firms connect to all types of firms, but small firms connect only to large firms. We compute degree assortativity defined in Newman (2003) as the Pearson correlation coefficient between the degree of firms at the opposite sides of the same edge. Newman's metric is easy to interpret: it varies from -1 for perfectly disassortative networks, to 1 for perfectly assortative networks, and equals zero when there is no correlation between the degree of connected nodes. For directed networks, there are four degree assortativity measures, each of which combines the in- and out-degrees of the suppliers and customers.

To give some context, social networks are frequently characterized by assortative mixing ($r > 0$), while technological and biological networks often have disassortative mixing ($r < 0$) (Newman, 2003). Canonical random graphs such as Erdős-Rényi (ER) or Barabási-Albert models have zero assortativity in the limit of a large number of nodes. For Ecuador and Hungary, we find a weakly negative assortativity between -1.5% and -13% (Table 6) depending on the kind of assortativity

[5]Nestedness means that rows and columns of the adjacency matrix can be rearranged such that the upper left part is mostly full of positive values while the rest is mostly full of zeros. A network is perfectly nested "when the degree of i is smaller than the degree of j, then the neighbourhood of i is contained in the neighbourhood of j". See Mariani et al. (2019) for a thorough review; they note that nested networks are usually disassortative.

measure. Similar values are reported in the literature. In contrast, for FactSet, we find an assortativity mildly positive or close to zero. Of all the possible ways to compute assortativity, the largest in magnitude is the correlation between the out-degree of suppliers and the in-degree of customers; that is, suppliers with many customers tend to sell to customers with few suppliers.

Table 6: Assortativity coefficients

Dataset	Year	$r_{k,k}$	$r_{k^{\text{in}},k^{\text{out}}}$	$r_{k^{\text{out}},k^{\text{in}}}$	$r_{k^{\text{in}},k^{\text{in}}}$	$r_{k^{\text{out}},k^{\text{out}}}$	
Ecuador	2015	-12.2	-3.8	-13.0	-10.5	-5.3	This paper
Hungary	2021	-7.6	-1.5	-8.9	-5.6	-2.4	This paper
Hungary	2019	-4.4	-1.5	-5.6	-3.1	-2.7	This paper
Hungary	2015	-5.5	-2.9	-7.0	-5.3	-3.5	This paper
FactSet	2020	1.9	2.2	0.8	-0.2	4.7	This paper
Japan	2006	-7.5	negative	negative			Fujiwara and Aoyama (2010)
Japan listed	2016	-21					Krichene et al. (2019)
West Bengal	2016Q4	-6.2					Kumar et al. (2021)

Notes: Assortativity coefficients as defined in Newman (2003). $r_{k^{\text{in}},k^{\text{out}}}$ denotes the correlation between the suppliers' in-degrees and the customers' out-degrees, where each edge is a data point. Other columns are interpreted similarly. All values are multiplied by 100.

Bernard et al. (2022), Bernard et al. (2019), Alfaro-Urena et al. (2018) and Cardoza et al. (2020) use a different measure of assortativity. Their downstream assortativity measures how the average number of suppliers of i's customers changes as i's number of customers changes. Likewise, upstream assortativity measures the change in the average number of customers of i's suppliers as i's number of suppliers changes. Regardless of the assortativity measure used, they find a negative degree of assortativity for Belgium (Bernard et al., 2022), Japan (Bernard et al., 2019), Dominican Republic (Cardoza et al., 2020) and Costa Rica (Alfaro-Urena et al., 2018).

3.3 Reciprocity, clustering and path lengths

We start this section by documenting the substantial extent to which links are reciprocal. Then, using the undirected version of the network, we show that the prevalence of closed triangles among all the possible triples (global clustering) is low and can be mostly explained by degree heterogeneity. In contrast, the proportion of node's neighbours that are themselves connected (local clustering) is much higher and higher than a random benchmark that preserves the degree distribution. Finally, we show that the shortest paths between pairs of nodes are very small, typically around 3 steps. For most of these properties, however, we show that non-administrative datasets or datasets with a high reporting threshold provide biased results.

3.3.1 Reciprocity

Reciprocity is the probability that an existing edge is reciprocated. In social networks, it can be very high. For instance, in friendship networks in US schools, the reciprocity is between 0.3 and 0.5 (Ball and Newman, 2013). For firm-level production networks, we found that reciprocity is much lower but still much higher than expected in an equivalent ER random graph, where it is very close to zero. Table 7 shows the empirical values of the reciprocity in Ecuador (around 5%), Hungary (9–4% depending on the threshold) and FactSet (around 3%).

3.3.2 Clustering

In social networks, it is very common that two of a person's friends are also themselves friends. Do we observe a similar pattern among firms? That is, if a firm transacts with two other firms, are these two firms likely to have a supply relationship? We convert each of the networks to an undirected network by assuming that any directed edge is an undirected edge and remove duplicated edges arising because of reciprocal edges. We then look at two standard metrics: the local and the global clustering coefficient.

Table 7: Reciprocity, path lengths and clustering. All values are in percentages, except for path lengths.

Dataset	Year	Recip.	C_g		$\bar{C}_i$		Path lengths			Source
		Empi	Empi	CM	Empi	CM	Empi	ER	CM	
Ecuador	2015	4.6	2.5	4.1	28.0	10.9	2.8	2.9	2.9	This paper
Hungary	2021	3.9	0.5	1.1	19.6	6.8	2.9	3.4	3.0	This paper
Hungary	2019	6.7	1.2	0.7	11.4	1.1	4.1	5.2	3.9	This paper
Hungary	2015	8.7	1.1	0.7	12.9	1.3	4.8	6.9	4.4	This paper
FactSet	2020	2.8	1.8	0.3	3.1	0.3	6.1	8.0	4.9	This paper
FactSet US	2016		2.4				4.8			Taschereau-Dumouchel (2022)
Japan	2005						4.6	10.4		Ohnishi et al. (2010)
Japan	2006		0.2	1.8	4.6		5.6	10.1		Fujiwara and Aoyama (2010)

Notes: "Recip" stands for reciprocity, C_g and C_i are the global and local clustering coefficients. "Empi" stands for Empirical, ER and CM for Erdös-Renyi random graph and Configuration Model. For calculating the global and local clustering coefficient for the 2 null model, we simulate 100 CM models and show the mean over the 100 simulations. For calculating the average shortest path for the 2 null models, we sampled 10^4 node pairs uniformly at random and computed the shortest path between each node pair. We simulate 10 ER or configuration models and show the mean over the 10 simulations. For Hungary 2021, due to the much longer computation time, we sampled 10^3 node pairs. All values are in percent, except path lengths.

The *global clustering coefficient* gives information about the density of triangles in the network. It expresses the probability that two of a node's neighbours are themselves connected; it is defined as

$$C_{\mathrm{g}} = \frac{\text{number of closed triangles}}{\text{number of possible triangles}} . \tag{3}$$

By contrast, the *local clustering coefficient* is the property of a single node:

$$C_i = \frac{\text{number of pairs of neighbours of } i \text{ that are connected}}{\text{number of pairs of neighbours of } i} . \tag{4}$$

Note that C_i is undefined for firms with degree 1 since they do not have a single pair of friends ($C_i = 0/0$); we exclude these firms from the analysis. A low local clustering coefficient is an indicator of centrality, in the sense that firms with low clustering coefficients are by definition connecting pairs of firms that are themselves not connected.

Table 7 shows the average local clustering coefficients $\bar{C}$ and the global clustering coefficient C_{g} in our three networks and in the literature.[6] Both the global and the average local clustering coefficients are substantially larger in Ecuador than in other networks, although the "complete" Hungary (2021) network also features high local clustering. We compare these results to a configuration model (CM), a random benchmark that preserves the nodes' degrees but is otherwise random.[7] The complete administrative datasets provide a relatively clear picture: global clustering is smaller than the random benchmark, something that is not clear from the non-administrative or incomplete data, while local clustering is much higher than the random benchmark, which would have been correctly identified on the incomplete datasets.

As is well-known, the difference between local and global clustering coefficients is partly due to the fact that the degree distribution is highly heterogeneous and that there is a negative correlation

[6]The average local clustering coefficient for globally listed firms reported by Kashiwagi et al. (2018) is slightly higher than what we find for FactSet. While Kashiwagi et al. (2018) also use FactSet's supply chain dataset, first, they also use data prior to 2013, where coverage is less good and, second, they merge FactSet supply chain with other datasets so that their final sample includes only 2,748 firms. Their final dataset is thus very different from ours.

[7]Consider a graph G that we wish to compare to the "random" benchmark. In theory, the CM should provide the set of all possible random graphs that have the exact same degree sequence as G. To compare G to the random benchmark, one should draw graphs uniformly from this set and compute the metric of interest. In practice, it is very hard to sample uniformly from the set of *simple* (no loops and no duplicated edges) *connected* graphs. We use a fast algorithm `sample_degseq(..., method ='simple')` from `iGraph` in `R`, but with the disadvantage that it allows duplicated edges, loops and can return disconnected graphs. The prevalence of loops, duplicated edges and disconnected nodes is small. We omit the comparison to ER graphs, where both clustering coefficients are very close to zero.

between node-level local clustering and degree. To see this, consider a very large degree firm. It usually has a very low local clustering, otherwise the network would be dense. Thus, a large degree firm has a huge number of *pairs* of partners that are not connected and each of these pairs contributes to the overall number of triangles that are not closed, leading to a low global clustering. However, the average local clustering coefficient is an average where each firm weighs equally, so large firms contribute little and the average is driven by the many small degree firms, which can easily have fairly high local clustering coefficients.

Overall, the excess local clustering compared to a configuration model show that matching is not only determined by degree. An intuitive explanation could be geography, as Bernard et al. (2019) find, firms tend to connect with firms that are closer in space, which would make reciprocal links and triads more likely. More generally, to explain the presence of excess clustering, a successful class of models is the one based on hidden geometries, where nodes are more likely to be connected if they are close in some underlying metric space (Serrano et al., 2008).

3.3.3 Paths

In undirected networks, a *walk* between two nodes in the network is a route from one node to another node by travelling along the edges of the network. If the walk does not intersect itself, it is called a *path*. The length of a walk, or a path, is the number of edges that need to be crossed (or the number of hops) to get from node i to node j. The *shortest path* between two nodes is the walk that has to make the least number of hops among all the possible walks (it might not be unique). The *diameter* is the length of the longest shortest path. As in Section 3.3.2, we convert the directed networks to undirected networks.

In production networks, we would expect that short path lengths imply that shocks at the firm level can reach most firms in the network more quickly and, potentially, more strongly. For example, Carvalho et al. (2021) study the impact of the 2011 great Japan earthquake and find that indirect suppliers and customers of directly affected firms were also affected, but the effect decays substantially with network distance. While studies of firm-level production networks typically do not report statistics of path lengths, we provide these because we think that models of endogenous formation of production networks should try to match them. As a point comparison, Carvalho (2014) reports an average path length of 4 in the US BEA I-O tables (417 industries).

Figure 5 shows the distribution of the length of the shortest paths for our three networks. First, the distribution of path lengths is stable over time except for Hungary, where we see the strong effect of the reporting threshold. In the early years, the distribution of path lengths in Hungary is closer to that of FactSet, which is missing many firms and relationships. After the threshold is removed, the distribution of path lengths in Hungary is astonishingly similar to that of Ecuador. Second, the mode for our two "complete" networks is 3 and the average (Table 7) is between 3 and 4, below the estimate of 4 for US I-O tables. This is a surprising result as we might have expected a higher distance in a very large firm-level network than in a small, aggregated industry network.[8] This can be due to differences between the US economy and Hungary/Ecuador, but we think that it is more likely to be due to national accounting conventions, for example, because national accounts do not show links from wholesalers and retailers to their suppliers of goods destined to be resold.

Are these results surprising given the existing density of the network and the degree distributions reported in the previous sections? Given the large number of nodes, one could have expected that if we pick two firms at random, it would typically take many steps to connect them. It turns out that in most networks, the average shortest path length is very small, a phenomenon known as the *small-world* effect (Milgram, 1967; Watts and Strogatz, 1998). This effect is relatively well understood since even simple models of random network formation produce fairly short path lengths. In Table 7, we compare the average shortest path length in our networks with those expected from an ER model and CM.[9] We also append the results reported by Fujiwara and Aoyama (2010) and Ohnishi et al.

[8]In the ER model, the average path length increases with size as $\sim \ln N$, while in growing networks that lead to a power-law degree distribution with $1 < \gamma < 2$, the average path length increases much more slowly (Cohen and Havlin, 2003), scaling as $\sim \ln \ln N$, because the presence of hubs shortens the distance between most pairs of nodes.

[9]In the ER model, each possible edge exists with probability p. We calibrate p to the empirical density of the

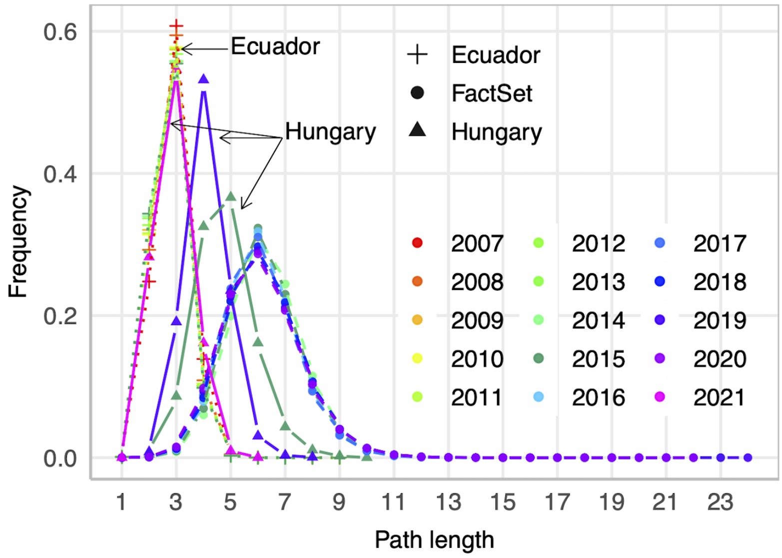

Figure 5: Distribution of the length of the shortest paths in Ecuador, Hungary and FactSet over time. We converted each network to an undirected network by assuming that any directed edge is undirected. We removed duplicated edges. Because of the long computation time for Hungary, we computed the frequency shortest path lengths in a random sample of 10^4 pairs.

(2010) for Japan and compute the ER benchmark for them by making use of their published data on the number of nodes and edges. For our complete datasets, the empirical average path length is similar to that of the ER model or CM. But for incomplete datasets, the empirical average path length is smaller than in the ER model and slightly higher than in the CM.

4 Results on weighted networks

Due to a lack of data, much less is known about weighted networks, compared to binary networks. The Belgian network is probably the most studied firm-level network with information on the monetary values of firm-to-firm transactions, but a number of others have appeared recently (Table 1). In this section, we provide a detailed analysis of the distribution of key quantities of weighted networks. We do not have data for the weights in FactSet, but we do have data for our other two networks, Ecuador and Hungary. Our data is on the in- and out-strengths; that is, on *intermediate* costs, meaning costs excluding factor costs, imports, taxes and subsidies, and on *intermediate* sales, i.e., sales excluding sales to final demand, exports and taxes; although there might be purchases of capital goods. Throughout the section, we use the terms “network sales” and “network expenses” to denote these quantities.

We first discuss the distribution of weights, finding a power-law exponent slightly above 1. We then show that the strength distributions have an exponent very close to 1, confirming studies of firm sizes. Next, we look at the relationships between strengths and degrees, that is, the relationship between network sales and the number of customers, and between network expenses and the number of suppliers. While classic input-output analysis and more modern models focus on technical coefficients and the Leontief inverse, we do not have the necessary data to compute these quantities for Ecuador. To sidestep this issue, as a final result, we consider the distribution of the influence vector, a centrality measure that is motivated by the benchmark Cobb-Douglas model with uniform final demand shares. The influence vector gives the elasticity of aggregate output to a total factor productivity (TFP) shock to a firm (Acemoglu et al., 2012; Carvalho and Tahbaz-Salehi, 2019).

network compared. The ER model has the feature that when p is such that the mean degree is less than 1, the network likely consists of disconnected clusters, and when $p > 1$, a “giant” component emerges. In our case, the mean degree is always substantially above 1, but a draw from the ER ensemble still almost always contains a small number of nodes outside of the GWCC. We remove these before computing path lengths.

4.1 Distributions of the value of transactions

A small number of studies report summary statistics for the value of the transactions (or "network weights") and find that it is heavy-tailed (Dhyne et al., 2015; Magerman et al., 2016; Bernard et al., 2022; Huneeus, 2020); Figure 1 in the Section 1 confirms these findings. More quantitatively, we find that the estimated power-law exponents are remarkably similar and slightly above 1 for both Ecuador and Hungary over time, regardless of the estimation method used (see Table 8 and Table C.7 for a more detailed account). We conclude that the weight distributions very likely have a divergent second moment.

Table 8: Tail exponents for weighted network quantities

	Year	Weight	In-Strength	Out-Strength	Influence	Source
Ecuador	2015	1.14	0.88	0.92	1.28	This paper
Hungary	2021	1.18	1.01	1.02	1.40	This paper
Hungary	2019	1.14	0.99	1.00	1.37	This paper
Hungary	2015	1.15	1.05	0.92	1.44	This paper
Dutch bank 1	2019		1.03	1.05		Ialongo et al. (2022)
Dutch bank 2	2019		0.69	0.72		Ialongo et al. (2022)
Belgium	2012				1.12	Magerman et al. (2016)

Notes: Parameters estimated using `plfit`.

4.2 Distributions of network sales and expenses (strengths)

Our data is only on *network* sales and expenses (i.e., out- and in-strengths), but we would expect that they are highly correlated to other indicators of a firm's "size". It is well established that the distribution of firms' revenues has an exponent close to 1 (Axtell, 2001). Similar exponents are also found when size is measured in terms of employees or capital (Axtell and Guerrero, 2021).

Figure C.2 shows the distribution of network sales and expenses over time for Ecuador and Hungary. Both are markedly stable over time, except of course for Hungary, where we see that the distributions have a break at the reporting threshold (while the threshold applies to the weights, many firms have an in- or out-degree equal to 1). Table 8 shows the estimated power-law exponents, which are close to 1, suggesting a possibly infinite mean. However, Tables C.5 and C.6 show that the GEVD estimators tend to be above 1, so we tentatively conclude that the tail exponents of both strength distributions are slightly above 1.[10]

As expected, firms with higher network sales also tend to have higher network expenses (Figure C.3). However, Figure C.3 suggests substantial heteroskedasticity, with the relationship between network sales and expenses being more predictive for large firms than for small firms. We confirm this in Figure B.1 using estimators of the quantile conditional function.

4.3 Strength-degree relationships

Do firms with more customers have higher network sales? Do firms with more suppliers have higher network expenses? Intuition suggests that yes, but the exact value of the elasticities is important. For instance, we may think that the marginal customer would be less important than the average customer so that growing the customer base by 1% would result in less than a 1% increase in sales.

Figure 6 shows the 2D histogram for our two datasets. Again, the 2015 Ecuadorian network and the 2021 Hungarian network appear remarkably similar. In all four cases, there is a clear positive relationship between strength and degree, but with a fairly high number of large firms (in terms of network sales or expenses) with only a few partners. In Figure B.1, we document heteroskedasticity

[10] By examining qq-plots, we also find that, in contrast to other distributions reported in the paper, the lognormal provides a good fit (a point also noted by Ialongo et al., 2022).

more precisely and find that regressions of strength on degree are fairly homoskedastic while regressions of degree on strength are highly heteroskedastic due to the presence of small degree firms that have high strengths.[11]

We are not aware of any papers documenting this pattern, but we think it is important in the context of systemic risk. For instance, if the importance of a firm depends positively on its size and its vulnerability to shocks depends negatively on its ability to diversify its supplier or customer base (e.g., Herskovic et al., 2020), then large firms with few partners are both important and vulnerable. The top-left quadrant of each panel in Figure 6 shows that there are many such firms.

Table 9 investigates the data from Figure 6 quantitatively. While some papers report regressions of strength on degree, others report regressions of degree on strength, so we show both. We find good consistency among our datasets and with the literature despite some methodological differences in the choice of variable. The elasticities do not appear to change dramatically with the change in reporting threshold, although the customer elasticity of network sales seems more affected than the supplier elasticity of network expenses.

Table 9: Strength-Degree elasticities

	Year	$\ln s \vert \ln k$	$\ln k \vert \ln s$	R^2	Source
Network sales & number of customers					
Ecuador	2015	0.88	0.31	0.27	This paper
Hungary	2021	1.05	0.36	0.37	This paper
Hungary	2015	1.38	0.38	0.53	This paper
Belgium	2014	0.77		0.35	Bernard et al. (2022)
Chile	2014–2020		0.33	0.25	Miranda-Pinto et al. (2022)
Chile	2018–2019		0.42	0.46	Arkolakis et al. (2023)
Costa Rica	2008–2015	1.2			Alfaro-Urena et al. (2018)
Turkey	2015		0.44	0.33	Demir et al. (2021)
Network expenses & number of suppliers					
Ecuador	2015	1.54	0.41	0.63	This paper
Hungary	2021	1.35	0.44	0.60	This paper
Hungary	2015	1.39	0.45	0.63	This paper
Chile	2018–2019		0.45	0.20	Arkolakis et al. (2023)
Costa Rica	2008–2015	0.89			Alfaro-Urena et al. (2018)
Japan	2005–2010		0.33		Bernard et al. (2019)
Turkey	2015		0.58	0.61	Demir et al. (2021)

Notes: OLS regressions of either the log of strength on the log of degree (column $\ln s|\ln k$) or the log of degree on the log of strength (column $\ln k|\ln s$)). All the observations equal to zero are dropped. Bernard et al. (2022) use network sales and add 4-digit industry fixed effects. Miranda-Pinto et al. (2022) use total sales, firms > 5 employees, and degree as the number of suppliers and customers. Alfaro-Urena et al. (2018) use total sales, demeaned by industry, high volume industries only. Arkolakis et al. (2023) use total sales and include year and state fixed effects. Demir et al. (2021) consider manufacturing firms.

[11] Figure 6 gives the visual impression that the strength-degree relationship might be heteroskedastic, but it is important to bear in mind that the sample maximum and minimum increase with sample size even for light-tailed distributions. In other words, the reason for which we observe a narrower range of values for strength conditional on high degree, compared to conditional on low degree, turns out to be simply due to the fact that there are fewer high-degree values to sample from, not due to a lower variance.

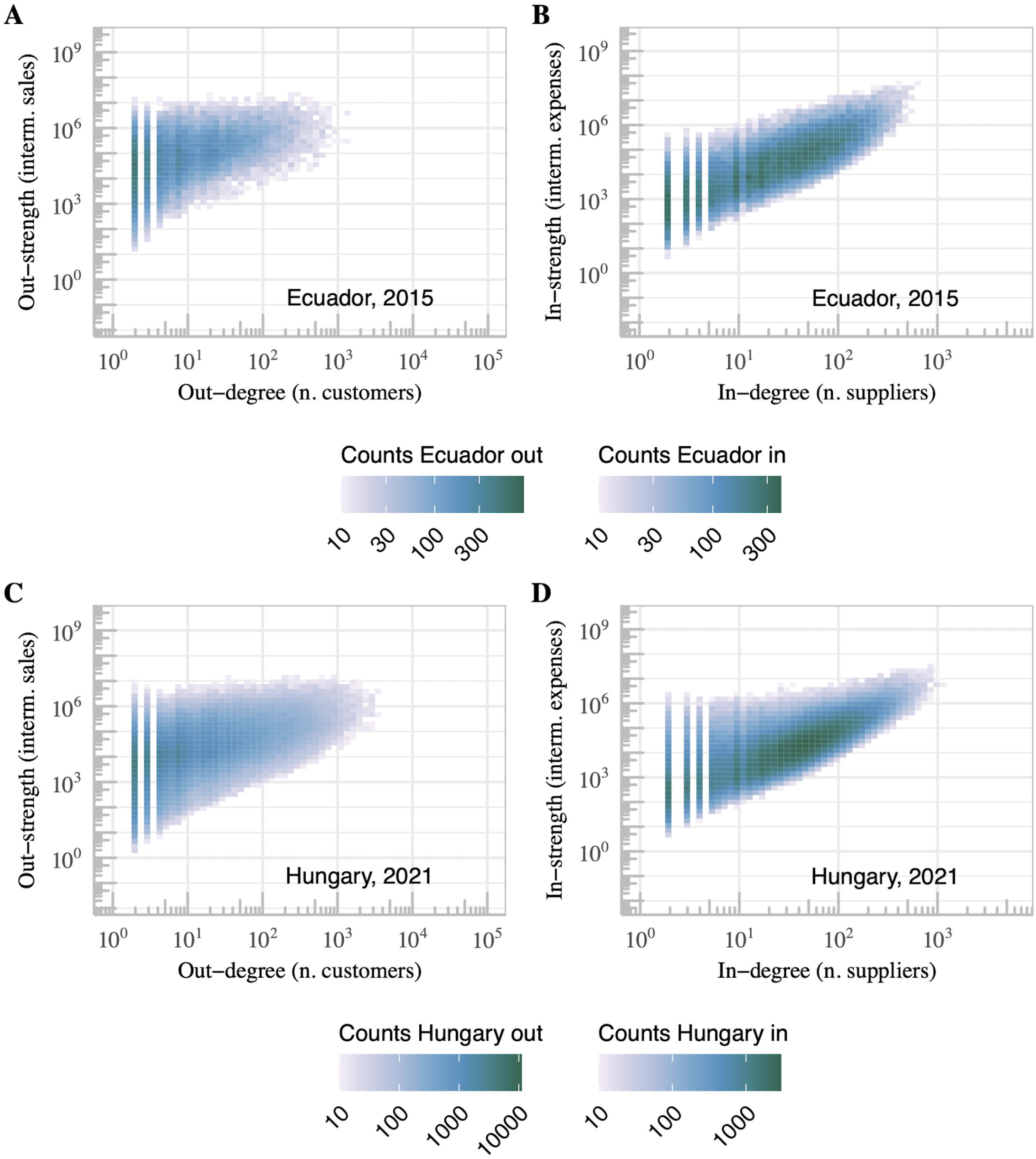

Figure 6: **A**, **C**, 2D histogram for the number of customers on the x-axis and network sales on the y-axis for Ecuador in 2015 and Hungary in 2021, respectively. **B**, **D**, 2D histogram for the number of suppliers on the x-axis and network expenses on the y-axis for Ecuador in 2015 and Hungary in 2021, respectively. Ecuador is in 2015 USD and Hungary in 2021 HUF. We bin both axes into 60 equally-spaced bins, we then count the number of observations in each square. We do not show squares that have less than 10 observations.

4.4 The influence vector

In a standard I-O equilibrium model with Cobb-Douglas production functions, no capital and uniform final demand shares (Carvalho and Tahbaz-Salehi, 2019; Acemoglu et al., 2012), the impact of firm-level TFP shocks on aggregate output is given by the *influence vector*

$$\boldsymbol{v} \equiv \frac{\alpha}{N}\Big[\boldsymbol{I} - (1-\alpha)\boldsymbol{P}^{\text{in}}\Big]^{-1}\mathbb{1} ,$$

where $\alpha \in (0,1]$ is the labour share of gross output,[12] $\boldsymbol{P}^{\text{in}}$ is the (column-stochastic) matrix of input shares computed as $P^{\text{in}}_{ij} = Z_{ij}/\sum_i Z_{ij}$, N is the number of firms, $\boldsymbol{I}$ is an identity matrix and $\mathbb{1}$ is a vector of ones.[13] The distribution of the influence vector is critical to understand aggregate fluctuations (see, e.g., Acemoglu et al., 2012; Carvalho, 2014; Carvalho and Tahbaz-Salehi, 2019). If the distribution has infinite variance, shocks at the micro level average out at a slower rate than would have been implied by the (non-generalized) central limit theorem. In the Cobb-Douglas model (Acemoglu et al., 2012; Carvalho and Tahbaz-Salehi, 2019; Magerman et al., 2016), the dependence of aggregate fluctuations on the influence vector is given by

$$\text{std}(\log \text{GDP}) \sim N^{-(1-1/\gamma)} ,$$

where $1 < \gamma \leq 2$ is the power-law exponent of the distribution of the influence vector.

Does the distribution of the influence vector feature power-law tails? If so, with what exponent? Figure 7 shows the distribution of the influence vector for Ecuador and Hungary.[14] It clearly displays heavy tails with a constant slope over three orders of magnitude and an overall shape very similar to that reported for US industries by Carvalho (2014). Table 8 reports the estimated power-law exponents for our networks and for Belgium (Magerman et al., 2016), showing very good consistency and suggesting a potentially slightly smaller exponent in firm-level datasets compared to the industry-level estimate of 1.48 by Carvalho (2014). Importantly, the estimates for Hungary are fairly constant, indicating that they are not sensitive to the reporting threshold. Table C.8 reports the estimates for each year and for different estimation methods, showing that these estimates are fairly robust.

[12]The labour share is assumed constant across firms. In a model without capital, $1-\alpha$ is the share of intermediate inputs in gross output.

[13]The equilibrium result when final demand shares are heterogenous is $\log \text{GDP} = \boldsymbol{\lambda}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ are the TFP shocks and $\boldsymbol{\lambda} = \boldsymbol{v}'\boldsymbol{b}$, where $\boldsymbol{b}$ if the vector of final demand shares; see Magerman et al. (2016). The fact that the Domar weights $\lambda_i \equiv \text{sales}_i/\text{GDP}$ are equal to $\boldsymbol{v}'\boldsymbol{b}$ is in principle a matter of accounting (see Equation 1 in Baqaee and Farhi, 2020, and Equation S51 in McNerney et al., 2022), but with firm-level data, not all intermediate sales are observed and, even when they are, the distinction between intermediate and final sales (which include investment) is difficult.

[14]First, to compute the influence vector, one needs to choose the parameter α, we consider that $1-\alpha$ represents the share of intermediates in gross output and choose $\alpha = 0.5$ as in Carvalho (2014); see Magerman et al. (2016) for a discussion. Second, the influence vector is equivalent to PageRank applied to the weighted matrix appropriately transposed and with damping $1-\alpha$. We use an existing implementation of PageRank in `iGraph` in `R` because it is very fast and with well-understood error tolerance.

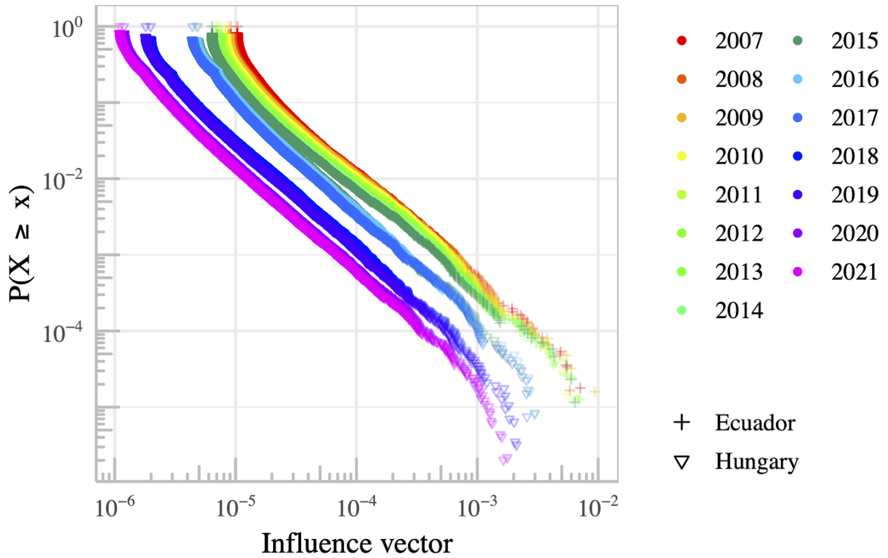

Figure 7: Distribution of the influence vector for Ecuador (cross) and Hungary (triangle) over time.

5 Discussion

Table 10 summarizes our results and provides a qualitative assessment of the agreement between complete datasets. Overall, we find that for most properties there is a strong agreement between VAT datasets with no or low reporting thresholds – there are properties of production networks that we think are solid enough to be considered "really known". If incomplete datasets feature different patterns and depart from good VAT datasets in a clear direction, our results allow us to put a sign on the bias that results from incomplete reporting. While the direction of the bias is not always clear, there are many cases where it is both clear and intuitive.

An overall pattern that emerges from Table 10 is that weighted quantities are usually both in very good agreement between complete VAT datasets and not dramatically biased in incomplete datasets. In contrast, binary statistics are more affected by the reporting thresholds. This is intuitive since a lower threshold leads to the presence of more links, but with low weights.

Of course, this assessment is qualitative. The behaviour of quantitative models can depend very sensitively on estimated moments, such as elasticities or power-law exponents, which can easily vary by 10-20% in our complete datasets.

Table 10: Summary of results.

Property	Results	Consistency	Bias	Section
Binary network				
Share supplier-only	$[0-20]\%$	Low	Upward	2
Share customer-only	$[14-20]\%$	Low	Upward	2
Mean degree	$30-50$	High	Downward	3.1
Mean degree $\sim$ #firms	Elasticity around 1/3	High	Unknown	3.1
In-Degree distribution	Tail exponent $\approx [1.4-1.6] > 2$	Very High	Unclear	3.2.1
Out-Degree distribution	Tail exponent $\approx [2-3] < 2$	Very High	Downward	3.2.1
In-Degree $\sim$ Out-Degree	Total Least Squares slope $[0.6-0.8] < 1$	High	Unknown	3.2.2
Degree assortativity	$\approx -[0.015-0.13] < 0$	High	Closer to zero?	3.2.2
Reciprocity	Much higher than random, $\approx [4-5.5]\%$	Very High	Unclear	3.3.1
Global Clustering	Low, lower than in CM	Moderate	Higher than CM	3.3.2
Average Local Clustering	$\approx 20-28\%$, much higher than CM	High	Downward	3.3.2
Average path length	$\approx 2.7-3$	Very High	Upwards	3.3.3
Weighted network				
Weights	Tail exponent $\approx [1.1, 1.2]$	Very High	None	4.1
Strengths	Tail exponent close to 1	Very High	None	4.2
Out-Strength $\sim$ Out-Degree	OLS slope $\approx$ [0.9,1.05]	Very High	Upward	4.2
Out-Degree $\sim$ Out-Strength	OLS slope $\approx$ [0.31,0.36]	Very High	Upward	4.2
In-Strength $\sim$ In-Degree	OLS slope $\approx$ [1.35-1.54]	High	Unclear	4.2
In-Degree $\sim$ In-Strength	OLS slope $\approx$ [0.41-0.45]	High	Unclear	4.2
In-Strength $\sim$ Out-strength	TLS slope ≈ 1	High	Unknown	4.2
Input shares	mode $\approx 0.1\%$, mean $\approx 2\%$	Very High	Upward	C.3
Output shares	mode $\approx 0.02\%$, mean $\approx 2\%$	High	Upward	C.3
Influence vector	Tail exponent $\approx [1.05, 1.4]$	High	Slightly upwards?	4.4

Notes: See the relevant section for the definition of the properties and evidence for the reported results. The edge direction is from a supplier to a customer, so the in-degree is the number of suppliers and the out-degree is the number of customers. The column "Consistency" provides our qualitative evaluation of the extent to which the reported results are similar across complete administrative datasets (Ecuador 2015, Hungary 2021, and Belgium when available). The column "Bias" provides our qualitative evaluation of whether non-complete datasets have a clear and systematic deviation from complete datasets, and is "unknown" if we could not evaluate it, or "unclear" if non-complete datasets departed from the complete datasets in different directions. "Upward" means that non-complete datasets tend to feature higher values. Properties marked $y \sim x$ refer to OLS or TLS estimates for the log-transformed values. All these results are very persistent over time. CM stands for Configuration Model, TLS stands for Total Least Squares.

6 Conclusions

There is a large consensus that modern macroeconomics should be bottom-up, data-rich and take into account interactions. This agenda is hampered by the fact that we know very little about firm-level production networks, raising concerns that observed differences across datasets may come from differences in data collection methods more than from genuine cross-country differences.

In this paper, we have made the first systematic attempt at summarizing what is known: what data is available? Are there generic properties of firm-level production networks that hold across different countries and over time? Do differences between datasets come from data collection and data cleaning methods?

As expected, some properties of production networks hold across all datasets, at least qualitatively; for instance sparsity, heavy-tailed degree and strengths distributions, or high local clustering and small average path length. However, our paper shows that we can be much more precise, thanks to the fact that many quantities are very similar across "complete" datasets, using incomplete datasets to calibrate models can lead to targeting clearly biased moments.

Aside from our systematic attempt at comparing datasets, we have also established a few facts of economic significance, for instance, the fact that the distribution of the number of customers exhibits much heavier tails than the distribution of the number of suppliers. This is intuitive, as firms may tend to grow by acquiring customers but tend to rely on their existing suppliers when scaling up. Second, we have found that many large firms actually have very few customers or suppliers; this suggests the existence of very large firms with limited downstream and upstream diversification, with potential consequences for systemic risk. Finally, we have shown that the distribution of firms' centrality (the "influence" vector) has a diverging second moment, a key property to establish the role of production networks in aggregate fluctuations.

There are several limitations to this work, which we regard as a key step of an important research agenda. A first line of research will need to dig deeper into the data collection methods and the comparability of firm-level datasets to classical national account objects. A second line of research should look at more sophisticated properties, perhaps driven by theoretical research. For instance, we have refrained from computing any quantity that makes use of industry classification systems or geographical locations, even though this is a clear avenue for applications.

To conclude, our paper provides a reference point for those interested in datasets of firm-level production networks with two objectives. First, it makes key moments and statistics publicly available, which should be useful for disciplining theoretical research and for researchers who do not have access to administrative data but need key moments to calibrate their macro models or create synthetic datasets. Second, it contributes to an emerging agenda to develop standards of data collection, cleaning and matching for micro-level production networks data around the world.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Alfaro-Urena, A., Fuentes, M. F., Manelici, I., and Vásquez, J. P. (2018). Costa Rican Production Network: Stylized Facts. *Central Bank of Costa Rica Research Documents*, No. 002.

Alfaro-Urena, A., Manelici, I., and Vasquez, J. P. (2022). The effects of joining multinational supply chains: New evidence from firm-to-firm linkages. *The Quarterly Journal of Economics*, 137(3):1495–1552.

Arkolakis, C., Huneeus, F., and Miyauchi, Y. (2023). Spatial production networks. *NBER Working paper*, 30954.

Astudillo-Estevez, P. A. (2021). *Towards a Post-Oil Economy: A Complexity Approach to Understanding Natural Resource Dependency and Economic Diversification in Ecuador [Doctoral dissertation]*. PhD thesis, University of Oxford.

Atalay, E., Hortacsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202.

Atalay, E., Hortaçsu, A., and Syverson, C. (2014). Vertical integration and input flows. *American Economic Review*, 104(4):1120–48.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *Science*, 293(5536):1818–1820.

Axtell, R. L. and Guerrero, O. (in press, 2021). *Dynamics of Firms from the Bottom Up: Data, Theories and Models.* MIT Press Cambridge, MA.

Ball, B. and Newman, M. E. (2013). Friendship networks and social status. *Network Science*, 1(1):16–30.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Barrot, J.-N. and Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., and Moxnes, A. (2022). The origins of firm heterogeneity: A production network approach. *Journal of Political Economy*, 130(7):000–000.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, 127(2):639–688.

Borsos, A. and Stancsics, M. (2020). Unfolding the hidden structure of the Hungarian multi-layer firm network. MNB Working Papers 2020/139, Magyar Nemzeti Bank (Central Bank of Hungary).

Brintrup, A., Barros, J., and Tiwari, A. (2015). The nested structure of emergent supply networks. *IEEE Systems Journal*, 12(2):1803–1812.

Buda, G., Hansen, S., Rodrigo, T., Carvalho, V., Ortiz, Á., and Mora, J. V. R. (2022). National accounts in a world of naturally occurring data: A proof of concept for consumption. *Cambridge Faculty of Economics Working paper.*

Cardoza, M., Grigoli, F., Pierri, N., and Ruane, C. (2020). Worker mobility and domestic production networks. *IMF Working paper*, No. 20/205.

Carvalho, V. M. (2014). From micro to macro via production networks. *Journal of Economic Perspectives*, 28(4):23–48.

Carvalho, V. M., Nirei, M., Saito, Y. U., and Tahbaz-Salehi, A. (2021). Supply chain disruptions: Evidence from the great east japan earthquake. *The Quarterly Journal of Economics*, 136(2):1255–1321.

Carvalho, V. M. and Tahbaz-Salehi, A. (2019). Production networks: A primer. *Annual Review of Economics*, 11:635–663.

Carvalho, V. M. and Voigtländer, N. (2014). Input diffusion and the evolution of production networks. Working Paper 20025, National Bureau of Economic Research.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2019). On binscatter. *arXiv preprint arXiv:1902.09608.*

Chacha, P. W., Kirui, B., and Wiedemann, V. (2022a). Mapping Kenya's Production Network Transaction by Transaction. *Oxford WP.*

Chacha, P. W., Kirui, B. K., and Wiedemann, V. (2022b). Supply chains in times of crisis: Evidence from Kenya's production network. CSAE Working Paper Series 2022-09.2, Centre for the Study of African Economies, University of Oxford.

Chakraborty, A. and Ikeda, Y. (2020). Testing "efficient supply chain propositions" using topological characterization of the global supply chain network. *PloS one*, 15(10):e0239669.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011.

Cohen, R. and Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):058701.

Costello, A. M. (2020). Credit market disruptions and liquidity spillover effects in the supply chain. *Journal of Political Economy*, 128(9):3434–3468.

Demir, B., Fieler, A. C., Xu, D., and Yang, K. K. (2021). O-ring production networks. *Journal of Political Economy*, forthcoming.

Demir, B., Javorcik, B., Michalski, T. K., and Ors, E. (2022). Financial constraints and propagation of shocks in production networks. *The Review of Economics and Statistics*, pages 1–46.

Dewachter, H., Tielens, J., and Hove, J. (2017). Credit supply shock propagation and amplification in the real economy: Firm–level evidence. *SSRN Electronic Journal.*

Dhyne, E. and Duprez, C. (2015). Has the crisis altered the Belgian economy's DNA? *Economic Review*, ii:31–43.

Dhyne, E., Duprez, C., et al. (2016). Three regions, three economies? *Economic Review*, iii:59–73.

Dhyne, E., Kikkawa, A. K., and Magerman, G. (2022). Imperfect competition in firm-to-firm trade. *Journal of the European Economic Association*, 20(5):1933–1970.

Dhyne, E., Kikkawa, A. K., Mogstad, M., and Tintelnot, F. (2021). Trade and domestic production networks. *The Review of Economic Studies*, 88(2):643–668.

Dhyne, E., Magerman, G., and Rubínová, S. (2015). The belgian production network 2002-2012. NBB Working Paper No. 288, National Bank of Belgium, Brussels.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2022). Quantifying firm-level economic systemic risk from nation-wide supply networks. *Scientific Reports*, 12(1):1–13.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2023). Estimating the loss of economic predictability from aggregating firm-level production networks. *arXiv preprint arXiv:2302.11451.*

Dorogovtsev, S. N. and Mendes, J. F. (2003). *Evolution of networks: From biological nets to the Internet and WWW.* Oxford university press.

Duprez, C. and Magerman, G. (2018). Price Updating in Production Networks. Working Paper Research No 352, National Bank of Belgium.

Fujiwara, Y. and Aoyama, H. (2010). Large-scale structure of a nation-wide production network. *The European Physical Journal B*, 77(4):565–580.

Fujiwara, Y., Inoue, H., Yamaguchi, T., Aoyama, H., Tanaka, T., and Kikuchi, K. (2021). Money flow network among firms' accounts in a regional bank of japan. *EPJ Data Science*, 10(1):1–26.

Furusawa, T., Inui, T., Ito, K., and Tang, H. (2017). Global sourcing and domestic production networks. *CESifo Working Paper Series*, No. 6658.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772.

Gabaix, X. and Ibragimov, R. (2011). Rank- 1/2: a simple way to improve the ols estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Grigoli, F., Luttini, E., and Sandri, D. (2023). Idiosyncratic shocks and aggregate fluctuations in an emerging market. *Journal of Development Economics*, 160:102949.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162.

Huneeus, F. (2020). Production network dynamics and the propagation of shocks. *Draft*. https://www.fedehuneeus.com/s/JMP_FHL.pdf.

Ialongo, L. N., de Valk, C., Marchese, E., Jansen, F., Zmarrou, H., Squartini, T., and Garlaschelli, D. (2022). Reconstructing firm-level interactions in the dutch input–output network from production constraints. *Scientific Reports*, 12(1):1–12.

Inoue, H. (2016). Controllability analyses on firm networks based on comprehensive data. arXiv preprint arXiv:1604.01322.

Kashiwagi, Y., Todo, Y., and Matous, P. (2018). Propagation of shocks by natural disasters through global supply chains. Discussion papers 18041, Research Institute of Economy, Trade and Industry (RIETI).

Kikkawa, A. K., Magerman, G., and Dhyne, E. (2019). Imperfect competition in firm-to-firm trade. *Available at SSRN 3389836*.

König, M. D., Levchenko, A., Rogers, T., and Zilibotti, F. (2022). Aggregate fluctuations in adaptive production networks. *Proceedings of the National Academy of Sciences*, 119(38):e2203730119.

Konno, T. (2009). Network structure of japanese firms. scale-free, hierarchy, and degree correlation: Analysis from 800,000 firms. *Economics: The Open-Access, Open-Assessment E-Journal*, 3:1–13.

Krichene, H., Fujiwara, Y., Chakraborty, A., Arata, Y., Inoue, H., and Terai, M. (2019). The emergence of properties of the japanese production network: How do listed firms choose their partners? *Social Networks*, 59:1–9.

Kumar, A., Chakrabarti, A. S., Chakraborti, A., and Nandi, T. (2021). Distress propagation on production networks: Coarse-graining and modularity of linkages. *Physica A: Statistical Mechanics and its Applications*, 568:125714.

Leontief, W. (1928). The Economy as a Circular Flow. (in German), Archiv für Sozialwissenschaft und Sozialpolitik, 60, 577-623. Reprinted and translated into English in *Structural Change and Economic Dyamics*, 2 (1991), 181-212.

Lim, K. (2018). Endogenous production networks and the business cycle. Working paper, University of Toronto, Toronto, Canada. Available at: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxsaW1rdm58Z3g6NTcxNzdlNjFkNmExYzQ5ZQ.

Lu, Y., Ogura, Y., Todo, Y., and Zhu, L. (2017). Supply Chain Disruptions and Trade Credit. RIETI Discussion Paper Series 17-E-054, Research Institute of Economy, Trade and Industry (RIETI).

Luo, J., Baldwin, C. Y., Whitney, D. E., and Magee, C. L. (2012). The architecture of transaction networks: a comparative analysis of hierarchy in two sectors. *Industrial and Corporate Change*, 21(6):1307–1335.

Luo, J. and Magee, C. L. (2011). Detecting evolving patterns of self-organizing networks by flow hierarchy measurement. *Complexity*, 16(6):53–61.

Magerman, G., De Bruyne, K., Dhyne, E., and Van Hove, J. (2016). Heterogeneous firms and the micro origins of aggregate fluctuations. *National Bank of Belgium Working Paper*.

Mariani, M. S., Ren, Z.-M., Bascompte, J., and Tessone, C. J. (2019). Nestedness in complex networks: observation, emergence, and implications. *Physics Reports*, 813:1–90.

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V. M., and Farmer, J. D. (2022). How production networks amplify economic growth. *Proceedings of the National Academy of Sciences*, 119(1).

Milgram, S. (1967). The small world problem. *Psychology Today*, 2(1):60–67.

Miller, R. E. and Blair, P. D. (2009). *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press.

Miranda-Pinto, J., Silva, A., and Young, E. R. (2022). Business Cycle Asymmetry and Input-Output Structure: The Role of Firm-to-Firm Networks. *Working Paper, Available at https://asilvub.github.io/assets/papers/CRN_MPSY_2022.pdf*.

Mizuno, T., Souma, W., and Watanabe, T. (2015). Buyer-supplier networks and aggregate volatility. In *The Economics of Interfirm Networks*, pages 15–37. Springer.

Morimoto, Y. (1970). On aggregation problems in input-output analysis. *The Review of Economic Studies*, 37(1):119–126.

Mungo, L., Lafond, F., Astudillo-Estévez, P., and Farmer, J. D. (2023). Reconstructing production networks using machine learning. *Journal of Economic Dynamics and Control*, 148:104607.

Nair, J., Wierman, A., and Zwart, B. (2022). *The fundamentals of heavy tails: Properties, emergence, and estimation*, volume 53. Cambridge University Press.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

Ohnishi, T., Takayasu, H., and Takayasu, M. (2009). Hubs and authorities on japanese inter-firm network: Characterization of nodes in very large directed networks. *Progress of Theoretical Physics Supplement*, 179:157–166.

Ohnishi, T., Takayasu, H., and Takayasu, M. (2010). Network motifs in an inter-firm network. *Journal of Economic Interaction and Coordination*, 5(2):171–180.

Panigrahi, P. (2023). Triad trade and small worlds of large spatial production networks. *Working Paper*. Available at: https://piyushpanigrahi.github.io/TriadTrade_PiyushPanigrahi.pdf.

Peydró, J. L., Jiménez, G., Kenan, H., Moral-Benito, E., and Vega-Redondo, F. (2020). Production and financial networks in interplay: Crisis evidence from supplier-customer and credit registers. *CEPR Discussion Paper*.

Saito, Y. U., Watanabe, T., and Iwamura, M. (2007). Do larger firms have more interfirm relationships? *Physica A: Statistical Mechanics and its Applications*, 383(1):158–163.

Serafino, M., Cimini, G., Maritan, A., Rinaldo, A., Suweis, S., Banavar, J. R., and Caldarelli, G. (2021). True scale-free networks hidden by finite size effects. *Proceedings of the National Academy of Sciences*, 118(2).

Serrano, M. Á., Krioukov, D., and Boguná, M. (2008). Self-similarity of complex networks and hidden metric spaces. *Physical review letters*, 100(7):078701.

Silva, T. C., Amancio, D. R., and Tabak, B. M. (2020). Modeling economic networks with firm-to-firm wire transfers. arXiv preprint arXiv:2001.06889.

Sornette, D. (2006). *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media.

Spray, J. (2017). Reorganise, Replace or Expand? The role of the supply-chain in first-time exporting. Cambridge-INET Working Paper Series No: 2017/18, Institute for New Economic Thinking, Department of Economics, Cambridge University, Cambridge.

Spray, J. (2021). Search externalities in firm-to-firm trade. *IMF Working paper*, 2108.

Spray, J. and Wolf, S. (2018). *Industries without Smokestacks: Industrialization in Africa Reconsidered*. Oxford University Press, Oxford, UK.

Taschereau-Dumouchel, M. (2022). Cascades and fluctuations in an economy with an endogenous production network. *Available at https://www.mathtd.info/files/papers/Network/paper.pdf*, ''.

United Nations (2010). *System of National Accounts 2008*. European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, World Bank.

United Nations (2018). *Handbook on supply, use and input-output tables with extensions and applications*. United Nations, New York, US.

Voitalov, I., van der Hoorn, P., van der Hofstad, R., and Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research*, 1(3):033034.

Vom Lehn, C. and Winberry, T. (2022). The investment network, sectoral comovement, and the changing us business cycle. *The Quarterly Journal of Economics*, 137(1):387–433.

Wang, Y. (2018). *Understanding Supply Chain Disruptions-Empirical Analysis of Supply Chain Structures*. PhD thesis, University of Michigan, Ann Arbor, Michigan, US.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Wu, D. (2016). Shock spillover and financial response in supply chain networks: Evidence from firm-level data. Unpublished working paper. Available at: https://www.semanticscholar.org/paper/Shock-Spillover-and-Financial-Response-in-Supply-%3A-Wu/9bf39f16bf5aa2f663ff196dfeae6a5a8848af3c.

Wu, J. and Birge, J. R. (2014). Supply chain network structure and firm returns. Available at SSRN: https://ssrn.com/abstract=2385217 or http://dx.doi.org/10.2139/ssrn.2385217.

Yuichi, K., Takashi, I., Hiroshi, I., and Hiroyasu, I. (2019). Hierarchical and Circular Flow Structure of the Interfirm Transaction Network in Japan. Discussion papers 19063, Research Institute of Economy, Trade and Industry (RIETI).

Zhigang, L., Shang-Jin, W., and Hongyong, Z. (2018). Production Chains, Exchange Rate Shocks, and Firm Performance. Discussion papers 18058, Research Institute of Economy, Trade and Industry (RIETI).

Appendices

A Data

In this appendix, we describe our three datasets in turn. We focus on data sources and on comparing aggregates with those available from national accounts. Our goal here is not to try to recompile national accounts quantities from firm-level data – this would require an entirely separate paper, as we explain below.

A.1 Differences between firm-level networks and national I-O tables

We provide a non-technical discussion of the differences between firm-level data and the Supply-Use (SU) or Input-Output (I-O) tables framework from national accounts. For a detailed handbook on the compilation of SU tables, see United Nations (2018). We omit a discussion of missing firms and missing transactions, as this is discussed throughout the main text and is highly dataset-specific.

Investment vs intermediate consumption. I-O tables are central to national accounts as they make it possible to compute GDP (i.e., total value-added) by subtracting intermediate consumption from gross output (i.e., total sales). Because net investment spending (i.e., gross fixed capital formation) is part of final demand, national accounts record transactions for intermediate consumption in the inter-industry transaction matrix, while transactions for investment goods and services are in a separate column. By contrast, firm-to-firm transaction networks include both intermediate and investment transactions indiscriminately. In practice, this can lead to a substantial bias: the total transactions observed within the firm network should in principle be higher than those in the inter-industry transaction matrix. If investment is 25% of GDP,[15] and intermediate transactions are about as large as GDP, the network transactions should be 25% higher than in the I-O table. This bias should be highly heterogenous across industries: Vom Lehn and Winberry (2022) reconstruct the investment network for 37 industries in the US, showing that a few industries represent a very high share of investment goods: construction, machinery, professional and technical services, and motor vehicles.

Wholesale trade, retail trade and transport. In national accounts, the convention is that wholesale, retail and transport should be better thought of as "pass-through sectors" rather than producing and consuming in a similar way as the other sectors do. More precisely, national accounts consider that the output of these industries is not their total sales but the *margins* they apply over the goods they buy and sell or transport. When industry j buys goods produced by industry i through a wholesaler k, I-O tables are thus able to record the flow of goods directly between industries i and j. The total cost paid by industry j is then split into the sales proceeds for industry i and a "trade margin" received by k. Another way to think about this is to consider that the wholesaler is a service provider – its true inputs are, say, labour, electricity and real estate, not the goods that it buys only for reselling. In sharp contrast, in firm-to-firm transaction data, we would observe the wholesaler buying the goods and reselling them and we would not observe a transaction between industry i and j. Therefore, we expect that the total sales of wholesale trade, retail trade and transport would be much higher in firm data than in the I-O tables (roughly 5 times higher if margins are 20%). Furthermore, we also expect the structure of the inter-industry matrix to be substantially different.

Financial institutions and financial services. The measurement of financial services in national accounts is complex. Additionally, financial institutions usually obey specific accounting rules and regulations. As a result, it is customary to remove financial firms when analyzing firm-level datasets. Although we have opted to keep all the firms in our analysis, we expect that firm-level network datasets based on VAT, surveys or payment systems would show quite different flows in and out of financial firms. Reconciling this with national accounts should proceed on a case-by-case basis.

Unit of analysis and industrial classification. To construct I-O tables, national statistical offices conduct surveys at the establishment level rather than the firm level. Establishments are preferred to aggregate production units into sectors because they tend to have a more homogeneous

[15] This is roughly the ratio at the world level according to World Bank data (accessible here https://data.worldbank.org/indicator/NE.GDI.TOTL.ZS).

production. However, the unit in micro-level datasets is typically a firm; this can cause significant issues as multi-establishment firms are common.[16] Having multi-establishment firms implies that to aggregate firm-level data into proper I-O tables, one needs to split a firm's output into various industries and then decide how to split the firm's inputs into the various output, a well-known problem for constructing symmetric input-output tables from SU tables (Miller and Blair, 2009).

In datasets based on transaction-level records, such as those that may be obtained from banks or payment providers, another issue arises as firms hold multiple accounts. For instance, consider a large multi-product firm that operates in multiple regions and that buys legal services from a large legal services firm with offices across the country. It is possible that the customer firm would centralize its payments so we would see a transaction from one headquarter to another, rather than multiple firm-to-firm links.

Prices and volume measures. National I-O tables use different concepts of prices. Roughly speaking and omitting imports and exports, on the resources side (output + net taxes + transport and trade margins), a unit of output is evaluated at the basic price; that is, the price that the producer can obtain, omitting taxes that go back directly to the government and including subsidies received. On the uses side (intermediate consumption + final demand), prices are market prices; that is, including VAT, and transport and trade margins. In firm-level data, it is more likely that the prices paid are market prices, but each dataset would have idiosyncrasies, e.g., the possibility or not to actually observe VAT. In addition, national accounts elaborate price indices that can be used to deflate the nominal values of transactions, while firm-level datasets often cannot distinguish between prices and quantities.

Timing of transactions. In principle, both national and financial accounts are compiled on an accrual basis; that is, they record "flows at the time economic value is created, transformed, exchanged, transferred or extinguished. This means that flows that imply a change of ownership are entered when the change occurs, services are recorded when provided, output at the time products are created and intermediate consumption when materials and supplies are being used." (United Nations, 2010, 3.166 p. 57). This can create substantial inconsistencies with firm-level datasets, particularly those created from direct money flows rather than from financial accounts, due to the prevalence of trade credit.

International Trade. Multinational firms typically file their accounts (and taxes) in various countries so that national accounts can ultimately try to separate the contributions of foreign firms domestically and domestic firms abroad. When using firm-level data, the ability to reconstruct tables close to national accounts would depend on the ability to access detailed financial accounts. Here again, the specifics of the data collection method would matter.

Taxes and government sector. In most countries, the public sector represents a large share of GDP. National accounts can represent this rather accurately, but it is more difficult for firm-level datasets, as the presence of transactions would depend on the legal nature of the entity and confidentiality issues (e.g., for defence spending).

Informal sector. In most countries, national accounts make an estimate of the value of the informal economy, which is unlikely to have a counterpart in tax-based administrative data.

All considered, reconstructing or reconciling national accounting quantities from firm-level datasets is a serious challenge, which we do not attempt here. Buda et al. (2022) provide a proof of concept

[16] Another issue is that different classification systems may be used. In the Ecuadorian dataset, the sectoral codes used in the national I-O tables differ from the ISIC classification codes used in the firm-level dataset. A one-to-one mapping from one classification system to the other is available only for the highest level of aggregation. For FactSet and Hungary, the firm-level dataset and the I-O table use the same sectoral codes – i.e., ISIC Codes Rev. 4.

that this can be done for consumption using payments data, but we are not aware of any study having done this for network data, which is more difficult. Having recognised these issues, we proceed to describe the datasets and provide some comparison of our data to relevant national accounting quantities.

A.2 Aggregate comparison

Keeping the largest connected component. Throughout the paper, we keep firms in the largest connected component. A network is *connected* if there is at least one path between all pairs of firms. The network is directed, so we keep firms in the giant *weakly* connected component (GWCC); that is, we require to be at least one path in either direction between all pairs of firms. Table A.1 shows that this data truncation leads to removing a larger fraction of nodes than edges and that the overall effect is very small on our VAT networks, but not insignificant for FactSet.

Table A.1: Share of nodes and edges not in the Giant Weakly Connected Component

Dataset	Year	% nodes removed	% edges removed
Ecuador	2015	0.046	0.00077
Hungary	2015	4.2	0.83
Hungary	2019	0.61	0.048
Hungary	2021	0.031	0.00046
FactSet	2020	8.7	2.3

Comparing firm-level data to national accounts. When we describe each dataset in subsequent sections, we show a comparison to national accounts at the industry level. Here, we start with a comparison at the aggregate level. Table A.2 summarises what data is available (we provide more details in the following subsections) and what quantities we chose as national account benchmarks. In particular, we compare the sum of network sales to the sum of national accounts intermediates plus investment, as we think network sales include transactions related to capital goods. Although our network sales probably include sales to other final demand categories such as government consumption, we do not add other national accounts' final demand categories.

Table A.2: Data underlying the comparison of firm-level aggregates to National Accounts

Firm dataset	Variable	NA concept	Source
Ecuador	B2B sales	Interm. sales + GFCF	NA[1]
Hungary	B2B sales	Interm. sales + GFCF	NA[2]
Hungary	Total sales	Gross output	NA[2]
FactSet US	Total sales (when available)	Gross output	BEA[3]
FactSet World	Total sales (when available)	Gross output	WIOD[4] & World Bank[5]

Notes: Data sources underlying Figure A.1. B2B stands for Business-to-business, NA stands for National Accounts and GFCF stands for Gross Fixed Capital Formation.

[1] Available at https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Catalogo/CuentasNacionales/Anuales/Dolares/MenuMatrizInsumoProducto.htm, downloaded on 9 August 2019.

[2] Available at https://statinfo.ksh.hu/Statinfo/themeSelector.jsp?&lang=en, downloaded on 16 March 2023.

[3] Available at https://apps.bea.gov/iTable/?reqid=150&step=2&isuri=1&categories=gdpxind, downloaded on 16 March 2023.

[4] Available at https://www.rug.nl/ggdc/valuechain/wiod/wiod-2016-release, downloaded on 19 December 2019, 2016 version.

[5] Available at https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, downloaded on 27 July 2020.

Figure A.1 compares the sum of the values of transactions or revenues to the most relevant quantity in national accounts for each of our three production networks.

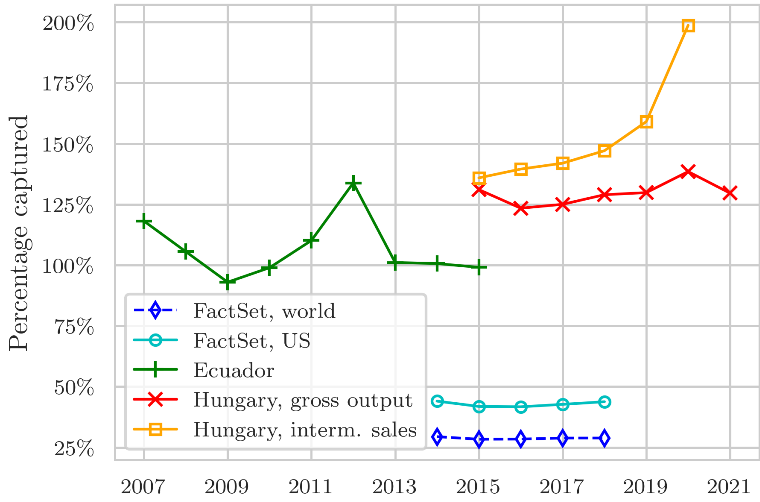

Figure A.1: Percentage of national gross output or total intermediate transactions captured by our network data. For Ecuador, we compare national and firms' total intermediate transactions (green cross). There is no I-O table for 2008, so we interpolate 2007 and 2009. For Hungary, we compare national gross output and firms' revenues (red x). When firms did not disclose revenues or revenues were negative or zero, we use firms' total intermediate sales (out-strength). For Hungary, we also compare intermediate sales (orange square). For FactSet, we compare firms' revenues to (1) world gross output as reported by the WIOD (blue diamond) and (2) to US gross output (light blue dot), for firms for which we have a financial statement.

Our VAT datasets usually have higher aggregate values than national accounts. As discussed previously, there are many sources of upward and downward bias when comparing to national accounts, but as we will see below a likely source for the upward bias of the VAT data compared to national accounts is the value of wholesale and retail trade, which is much higher in VAT data because it likely includes the value of goods bought for resale.

As a point of comparison, Dhyne et al. (2015) find that for Belgium, total turnover represents 95% of the total production from the national accounts.

Ecuador. We do not have firms' revenues but only firm-to-firm transactions. These transactions include intermediate and investment goods and services, so we compare the sum of all transactions with the sum of two components: the sum of all transactions in the national inter-industry transaction matrix and the sum of Gross Fixed Capital Formation (GFCF).

Hungary. For Hungary, in addition to the network sales, we have access to the revenues of many firms (there is variation across the years, see Table A.3) so we can compare both network sales and revenues to their closest equivalent in national accounts. For network sales, we use the sum of intermediate sales and GFCF in national accounts. For revenues, we use gross output. Some firms do not report their revenues and a few percent report negative or zero revenues so for these firms we use their network sales as a proxy for total sales (see Table A.3).

FactSet. We assess how much of world gross output we capture using firms for which we have financial statements (we describe data cleaning and network construction in more detail below). In our dataset, there are 410,584 unique firms over time; we have financial statements with positive revenues for only 26,141 of them. Since the number of firms for which we have financial statements is considerably lower than the overall number of firms in the network, the figures given below have to be interpreted as a *lower bound*.

We calculate world gross output by summing up firms' sales as declared in their yearly financial statements. We then compare firms' cumulative sales to world gross output taken from the World Input-Output Database (WIOD), extrapolated to 2018.[17] We also do a similar exercise for companies in the US, using gross output data from the BEA, which is available until 2018.

A.3 FactSet

Data sources. We use two data sources provided by FactSet: Supply Chain Relationships and Supply Chain Shipping Transactions. The Supply Chain Relationships data come in part from companies' filings required by US Federal Accounting Standards[18] and in part from information on supply chain relationships released in investor presentations, company websites and press releases. The second source ("Supply Chain Shipping Transactions") records shipment declarations at ports from the US Bill of Lading. FactSet collects this information from the US Customs and Border Protection.

The supply chain dataset goes from 2003 to the present date, while the shipment dataset starts in 2013 and goes until the present date (we downloaded these two datasets on 11 February 2021). Due to the nature of the data collection process, coverage is biased towards companies listed on US stock exchanges, large firms and large transactions. We keep the dataset from 2014 to 2020. We do not use years prior to 2014 because in 2013 FactSet changed the data collection methodology, enhancing the quality of the dataset.

The monetary values of customer-supplier transactions are rarely available and when recorded, they are reported as a revenue percentage earned by the seller from a specific customer. However, it is unclear to what disclosed revenue figure that percentage refers to (e.g., quarterly or annual income statement). Similarly, in the shipment dataset, the cumulative value of the goods shipped is only partially disclosed and with valuation methods that do not necessarily match balance sheet information. Therefore, we use only the binary topology.

In the raw data, links report the year, month, day and hour. The start and end dates correspond to when the record was first published and when the ending was announced. We consider that a relationship exists in a given year if it exists at any point during that year. For each company, we also have information on the sector (NACE Rev.2 codes at the 4-digit level) and the country where the company's headquarters are located.

For Figure A.1 and to assess the sectoral composition, we merged the network data with Fundamentals (downloaded in April 2020), which contains firms' financial statements. To avoid double counting, we aggregate all three FactSet datasets at the parent company level, meaning that we use consolidated income statements. We rely on the latest available information on a company's ownership structure as it is impossible to know the evolution of companies' ownership structures (mergers, acquisitions, buy-backs, etc.).

All the descriptive statistics below are for firms for which the information is available, which typically is a subset of all the firms in the network.

Coverage. We have information about the country for 99.96% of firms in our network. Figure A.2a shows that the US and China are the most represented countries. While only listed companies have to disclose information on their major customers, customers can be of any type. We have information for 99.99% of our 410,584 firms, showing more than 30 different types of firms (for instance, private companies, subsidiaries, listed companies, non-profit organisations or government). Figure A.2b shows the number of listed companies globally and in the US, comparing FactSet with data from

[17]The WIOD is available from 2003 to 2014 but our firm-level dataset is available from 2014 to 2020. We forecast world gross output from 2015 until 2018 using GDP from the World Bank as follows. We take the ratio of gross output to GDP, which is known to be fairly stable over time, and assume that after 2014 this ratio stays constant. This gives us gross output q_t as a function of GDP y_t and the gross output/GDP ratio ζ_t, thus $q_t = \zeta_{2014} \cdot y_t$ for all $t = 2015, \ldots, 2020$.

[18]The Statement of Financial Accounting Standards No. 131 requires publicly traded firms on US stock exchanges to report customers that account for 10% or more of their annual revenues, formally called *major customers*.

the World Bank.[19] Factset covers roughly half of the listed firms, with increasing coverage over time.

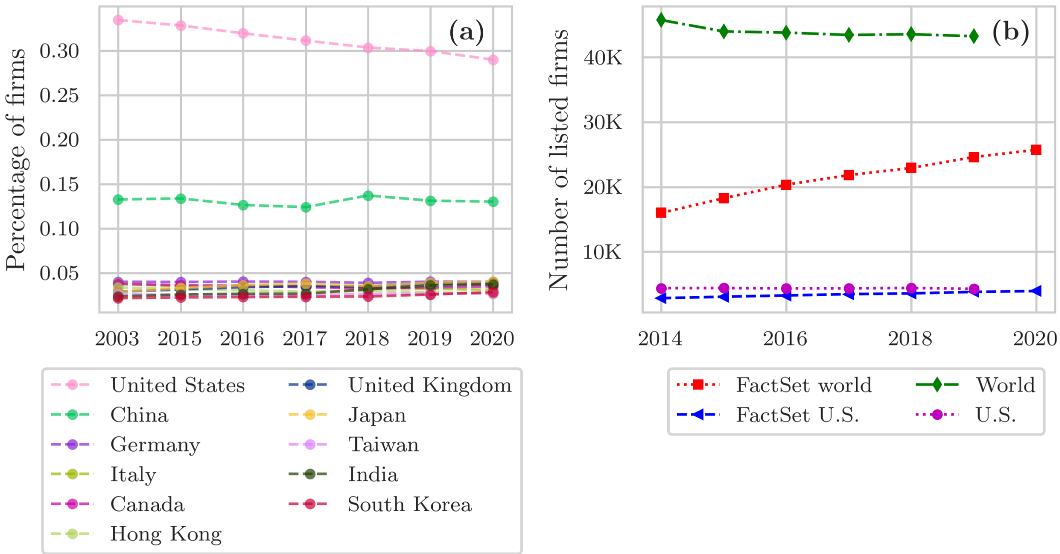

Figure A.2: **(a)**, Share of each country in all FactSet firms. **(b)**, Percentage of listed firms in FactSet compared to World Bank data, for the world for the US.

Sectoral composition. Figure A.3 shows the sectoral composition of the WIOD (black bars) and FactSet (for those firms for which we have financial statements) aggregated at the sector level (green bars) for the year 2014. We assess sectoral composition using the sectors' shares of gross output.

[19]Data available at `https://data.worldbank.org/indicator/CM.MKT.LDOM.NO`, downloaded on 16 March 2023.

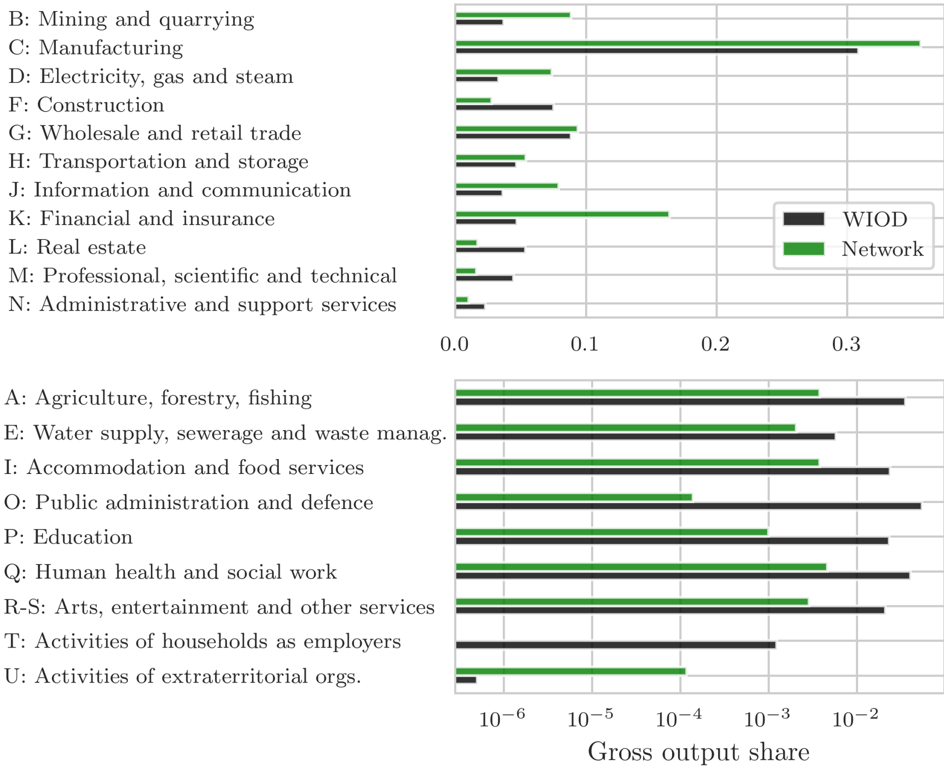

Figure A.3: Gross output shares in the WIOD (black) and in FactSet (green) in 2014. Sectoral codes are ISIC codes at the 1-digit level (Rev. 4). The top panel shows gross output shares that are bigger than 10^{-2} and the bottom panel those that are smaller than or equal to 10^{-2}.

A.4 Ecuador

Data source. The production network of Ecuador is collected through VAT filings and it is provided by the Internal Revenues Service (IRS) of Ecuador (Astudillo-Estevez, 2021). It comprises fine-grained information about all the legal (firms) and natural persons registered in the country between 2007 and 2015. We consider a dataset with 328,640 unique firms over the whole time period. In principle, it includes information on every transaction between all the entities, aggregated per year. For individual firms, it provides information on the industrial classification: ISIC Code at the 4-digit level, Rev. 4,[20] the taxpayer classification (public sector, private sector, IGO or NGO) and the fiscal address at the municipal level.

Firms need to report both their suppliers and customers. Sometimes the value of the transaction reported by the customer and by the supplier can differ. The IRS takes care of cleaning possible mismatches and we do not know which of the reported relationships are kept. The IRS is particularly concerned with detecting possible frauds related to transactions with large firms, which they identify using the weighted degree centrality. In the first couple of years of the data collection, numbers were reported manually, so the latest years are more reliable.

The dataset includes transactions between registered entities and some foreign companies that are not registered in Ecuador. Since the focus of the analysis is on the domestic product chain and because their information is incomplete, all these transactions were excluded. The dataset also contains self-loops, which represent transfers among establishments of the same firm. These transactions are not taxed and are purely for accounting purposes within the firm. We also remove these from all analyses.

Finally, we replaced one implausible value for a transaction (of the order 10^{12}) by its value in the previous year (of the order 10^6).

[20] ISIC stands for International Standard Industrial Classification of all economic activities. It is a standard classification of economic activities where entities are classified according to the main activity they carry out.

Sectoral composition. Figure A.4 shows the sectoral composition of the Ecuadorian economy according to national I-O tables (black bars) and our firm-level dataset aggregated at the sector level (green bars), for the year 2015. Since we do not have access to firm-level final demand or total revenues, we compare network sales to the sum of intermediate sales and GFCF in national accounts. We use a concordance table to translate National I-O tables (CICN codes) into the ISIC system.[21] The most important discrepancies are for Construction, which is vastly under-represented in the network, and Wholesale and retail trade, which, as expected, is vastly over-represented in the network.

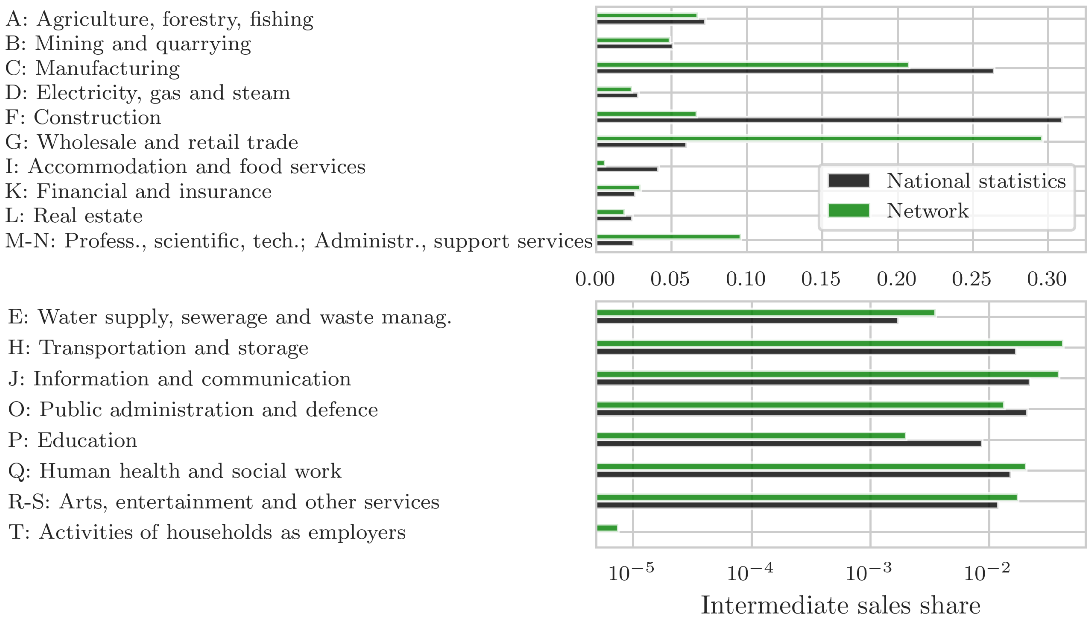

Figure A.4: Sectoral composition (intermediate sales shares) in the national I-O table at the sector level (black) and in our firm-level dataset (green) in 2015. Sectoral codes are ISIC codes at the 1-digit level (Rev. 4); sectors M and N, and R and S are grouped together to align the codes at the firm and sector level. The top panel shows intermediate sales shares in national accounts that are bigger than 0.023 and the bottom panel those that are smaller than or equal to 0.023.

A.5 Hungary

Source and reporting threshold. Hungary's network is collected by the National Tax and Customs Administration of Hungary. Before the first half of 2020, firms had to report a supply chain relation if the tax content of their cumulative trades or invoices exceeded a certain value during the reporting period (see below for more information). There are exceptions, however, depending on the sector the firm is in or for certain goods. There is heterogeneity in the reporting period, which can be annual, quarterly or monthly. The reporting period depends on the size of the firms, so they cannot choose it freely.

From 2015 to the second quarter of 2018, the threshold was 1 million HUF. It is calculated on the tax content of the sum of the transactions between two firms in a given reporting period. Given that the tax rate is 27% (although there are exceptions as noted above), the value of the transaction (without the taxes) above which reporting is required is HUF 3,703,703. During this period firms had to report both directions, i.e., both their purchase and their sales connections that were above the threshold. We use the network constructed from the information reported by the buyers, because they have a clear incentive to report (claiming back VAT), and the network appeared much more complete.

[21]The crosswalk is available at `https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Catalogo/CuentasNacionales/ClasProdSCN_12042013.xlsx`, downloaded on 1 December 2019. We use the crosswalk that goes from 69 to 13 industries.

From the third quarter of 2018 to the second quarter of 2020 there were three important changes. First, the threshold was lowered to 100,000 HUF. Second, it became interpreted at the invoice level regardless of the reporting frequency. So, only invoices above 100,000 tax content had to be reported, which means that the value of the transaction (without the taxes) needs to be above 370,370 (with exceptions as discussed above). On the one hand, more links are observed as a result, but, on the other hand, some edges are lost, especially those where the typical transaction amount is low but firms trade often enough to reach the previous threshold on the sum of the transactions between the firms in a given reporting period. Third, only the purchases had to be reported. Finally, since the second half of 2020, there is *no threshold anymore*, so all the invoices have to be reported.

The dataset covers the period 2014–2021. However, we dismiss the first year because the quality of the data is poor; this might be due to the inexperience of both the authorities and firms in the new reporting requirements.

Sectoral composition. For an in-depth description of the Hungarian dataset, we refer to Borsos and Stancsics (2020). Figure A.5 shows the sectoral composition of the Hungarian economy according to the national I-O tables at the sector level and in our firm-level network aggregated at the sector level, both for the year 2020 (left) and 2021 (right); data for 2021 are preliminary and might still be adjusted by the office of national statics. We assess sectoral composition using the sectors' shares of gross output. As for Ecuador, Wholesale and retail trade is vastly over-represented in the network compared to National Accounts.

Some firms do not report their sector; these account for 9.9% of total firms' revenues in 2020 and 8.8% in 2021. Not all firms report their revenues and some firms report zero or negative revenues, for these firms, we use their network sales. Table A.3 reports the percentage of firms in our network that report revenues and the percentage that report zero or negative revenues. The percentage of firms reporting negative revenues is very small, the highest percentage is in the 2021 network where it equals 0.01%. In 2021, we observe the highest percentage of firms that report zero revenues and the lowest percentage of firms that report financial information. Since in 2021 there was no reporting threshold, we observe almost all firms in Hungary, thus it makes sense to observe those differences.

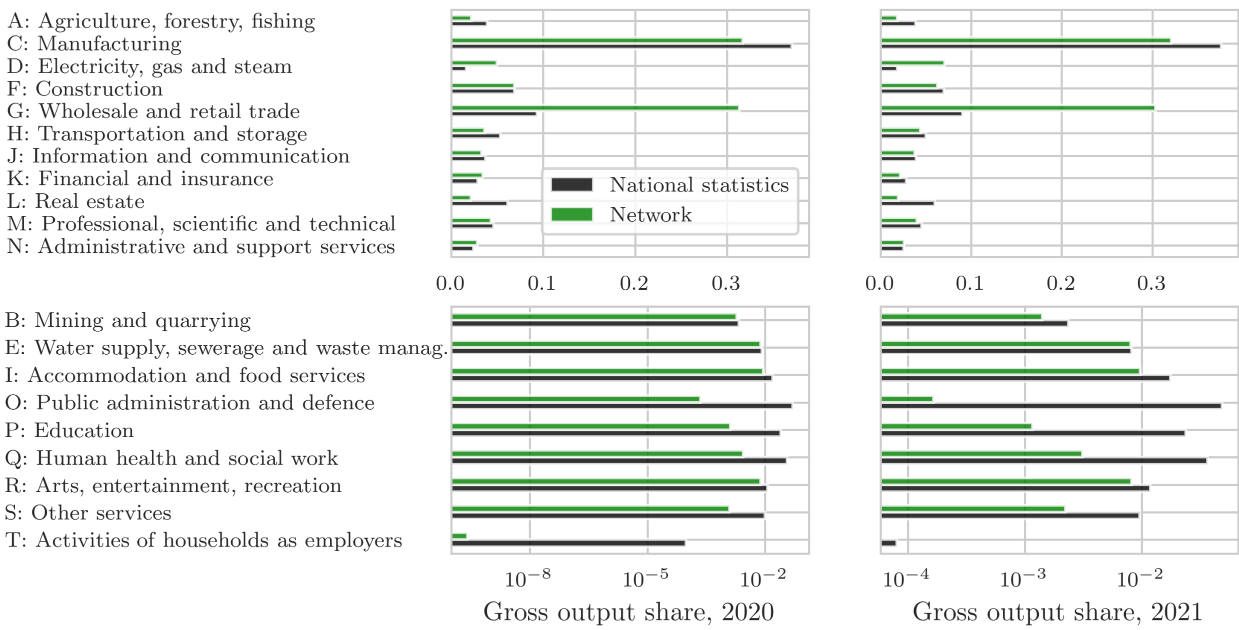

Figure A.5: Gross output shares in the national I-O table at the sector level (black) and in our firm-level dataset (green) in 2020 (left) and 2021 (right). The top panels show gross output shares that are bigger than 10^{-2} and the bottom panels show those that are smaller than or equal to 10^{-2}.

Table A.3: Percentage of firms in the GWCC that are in corporate tax data and those reporting zero or negative revenues.

Year	Pct with financials	Pct zero revenues	Pct negative revenue
2015	75.17%	1.46%	0.004%
2016	73.83%	1.62%	0.004%
2017	71.84%	1.60%	0.008%
2018	61.98%	1.81%	0.005%
2019	57.26%	1.99%	0.006%
2020	53.27%	4.52%	0.007%
2021	48.59%	5.05%	0.011%

B Characterizing distributions

B.1 Estimating power-law exponents

Power-laws as regularly varying distributions. Power-laws are sometimes defined as distributions with a Pareto tail, that is, the tail of the (complementary cumulative) distribution is exactly proportional to $k^{-\gamma}$, perhaps after some threshold $k_{\min}$. In applications where the tail of the distribution is of interest, it is better to consider the larger class of *regularly varying distributions*, which have a complementary cumulative distribution function (CCDF) of the form $\bar{F}(k) = \ell(k)k^{-\gamma}$, where $\ell(k)$ is a slowly varying function; that is, it satisfies

$$\lim_{x\to\infty} \frac{\ell(tx)}{\ell(x)} = 1 \ .$$

The presence of the slowly varying function implies that the shape of the distribution can deviate noticeably from a pure power-law for low and moderately high degrees (the body of the distribution) – as one would expect with any real-world phenomena in the presence of noise, heterogeneity, etc. – but keeps the key feature we are interested in: the extreme tail behaviour. For instance, in models of the granular origins of aggregate fluctuations (Gabaix, 2011; Acemoglu et al., 2012), what matters is that the variance of a given distribution diverges; this will be the case for all regularly varying distributions with $\gamma < 2$.

How can we test whether a distribution is regularly varying and if it is, how can we estimate γ? Extreme value theory (the Fisher–Tippett–Gnedenko theorem) tells us that the asymptotic distribution of the (suitably normalized) maximum of a sample of i.i.d. random variables, if it exists, is the Generalized Extreme Value Distribution (GEVD) with extreme value index ξ,

$$\Pr(V < v) = \exp\Big(- (1 + \xi v)^{-1/\xi} \Big) \ .$$

There are three subfamilies, characterized by the value of ξ. For $\xi < 0$, the GEVD is the Weibull distribution; for $\xi = 0$, it is the Gumbel distribution; and for $\xi > 0$, it is the Fréchet distribution. It turns out that the maximum domain of attraction (MDA) of the Fréchet distribution is exactly the set of all regularly varying distributions. So for any regularly varying distribution (RVD) with tail index γ, the distribution of the suitably normalized sample maximum is a GEVD with tail index $\xi = 1/\gamma > 0$.

Estimating power-law exponents. Estimating tail exponents require two choices: a choice of how many order statistics to keep (a threshold) and a choice of the tail exponent estimator to be applied to this restricted sample. The standard estimator in the literature is the method from Clauset et al. (2009), which uses the Maximum Likelihood Estimator (MLE) for pure Pareto tails (the Hill estimator), with the threshold chosen to minimize the Kolmogorov-Smirnov distance between the estimated and empirical CCDFs.

Voitalov et al. (2019) provide a review of tail index estimators and an implementation of the double bootstrap procedure to select the threshold. The double bootstrap takes two estimators that have been proven to be consistent, applies them to various sample sizes and picks the value of the lower bound that makes the two estimators in closest agreement. These methods have been shown to be consistent; see Voitalov et al. (2019) and Nair et al. (2022) for further details and references.

For discrete distributions, such as degree distributions, Voitalov et al. (2019) approximate the degrees to continuous reals by adding small symmetric noise sampled uniformly at random in the interval $[-0.5, 0.5]$. They show that adding such noise does not substantially affect the estimated value of the tail index as long as the distribution is regularly varying with Fréchet as the MDA and that the noise improves the convergence and stability of the estimators.

In the main text, we report the estimates from the method of Clauset et al. (2009), to make our results comparable with published results and because it has been shown to be fairly robust to finite-size effects, typically observed in network data.[22] We do not report the standard errors from Clauset et al.'s (2009) MLE because we do think that pure Pareto tails are the correct benchmark.[23]

With regularly varying distributions, it is not possible to do hypothesis testing. Regularly varying distributions do not form a parametric class of distributions; they are non-parametric with infinite degrees of freedom due to the unspecified slowly varying function $\ell(k)$. Importantly, there is an infinite number of regularly varying distributions for which a sampled sequence of finite length does not appear to be regularly varying. Likewise, there is an infinite number of distributions that are not regularly varying for which a sampled sequence of finite length may appear to be regularly varying. Therefore, the best strategy one can adopt is to consider all the available consistent estimators and check for agreement on the ranges of the estimated γ's.

As a result, we do not use a formal criteria for classifying distributions as regularly varying or not, but informally we are guided by the classification scheme adopted by Voitalov et al. (2019), where a distribution is *not a power-law* if at least one of the extreme value estimators returns $\xi \leq 0$; *hardly a power-law* if $\xi > 0$ for all the extreme value estimators and at least one $\xi \leq 1/4$ (i.e., $\gamma \geq 4$); *a power-law* if for all the extreme value estimators $\xi > 1/4$ (i.e., $\gamma < 4$); and *a power-law with divergent second moment* if for all the extreme value estimators $\gamma \leq 2$ (i.e., $\xi \geq 1/2$).

The $1/4$ threshold is set because if ξ is positive but very small, it is not possible to test whether $\xi = 0$. If $\xi = 0$, then the distribution is in the Gumbel MDA, which includes both light-tailed distributions and heavy-tailed distributions that are not regularly varying. The value $1/4$ is somewhat subjective and we may have wanted to make it depend on the number of observations. For instance, Dorogovtsev and Mendes (2003, Figure 3.32) provide a heuristic argument: to estimate a power-law with a reasonable degree of precision, we need data that span at least a 2-3 orders of magnitude and, given a sample size, the range of the data is heavily affected by γ; power-laws with $\gamma > 4$ would require a tremendous amount of data to span enough orders of magnitude to be measured properly.

When we report our detailed estimates (Tables C.2, C.3, C.5, C.6, C.7, and C.8), we report the estimated value of γ and the number of data points used to estimate the tail exponent, that is, the lowest order statistic used in the estimation, as determined by the double-bootstrap in the GEVD-based estimators, and by minimizing the K-S distance in `plfit`. These values are interesting because they show that various estimators use much more data than others, so this provides an additional robustness check.

Finally, a note on the lognormal distribution. Crucially, it has a finite second moment and is among the heavy-tailed distributions that are in the Gumbel MDA. However, when the lognormal distribution has a high variance, it can be easily mistaken for a power-law. This can be seen from

[22] In some data-generating processes, including some canonical models of growing networks, the asymptotic distribution is a power-law but for finite sizes, the distribution is a power-law with an exponential cutoff (which has finite moments). Serafino et al. (2021) have shown that the Clauset et al. (2009) estimator performs well to retrieve the true power-law exponent even with finite-size networks. See also Figure 8, panels c, h and m in Voitalov et al. (2019), showing the same result but for finite-size i.i.d sequences drawn from a power-law with "natural" exponential cutoff, $k^{-\gamma}e^{-k/n^{\omega}}$, where n is the sample size. This has finite moments for fixed n but pure power-law behaviour asymptotically. The Clauset et al. (2009) estimator performs relatively well at estimating γ for reasonable sample sizes $(10^3 - 10^5)$.

[23] Further, the MLE standard errors are based on the assumption that the data are i.i.d., which is unlikely to be a good assumption with network data. In practice, the MLE standard errors for the distributions we study are very small, and we think deceptively so.

the probability density function $\log p(x) = -\log x - \log(\sigma\sqrt{2\pi}) - (\log x - \mu)^2/2\sigma^2$, where as $\sigma \to \infty$ the quadratic term tends to zero; in these cases, the lognormal can look very similar to a power-law. Sornette (2006) shows that the lognormal can be rewritten as $p(x) = (x_0\sqrt{2\pi\sigma^2})^{-1}(x/x_0)^{-1-m(x)}$, where $x_0 = \exp(\mu)$ and $m(x) = \log(x/x_0)/(2\sigma^2)$ is a slowly varying function of x; when σ^2 is large enough, there is a large range of values x such that $m(x)$ is very small, and the lognormal looks like a power law in this region. In our case, we have found by examining qq-plots that lognormal fits are good for the distribution of strengths, but not for other quantities.

B.2 Characterizing joint distributions

In this appendix, we collect a number of technical details and empirical results related to the characterization of the joint distributions.

B.2.1 Total Least Squares

In many instances, we are interested in characterizing the direction of the relationship between two variables. For instance, we expect that firms with large sales also have large expenses, so we can hypothesize the deterministic relationship $s^{\text{in}} \propto s^{\text{out}\theta}$, perhaps with $\theta \approx 1$. Regressing (log) sales on (log) expenses only characterizes the slope of the *conditional* relationship. Therefore, the estimate of θ will differ if we regress sales on expenses or the other way around unless they are perfectly correlated.[24]

To characterize the "slope", we use the Total Least Square (TLS) estimator, which is well-known as a specific "errors-in-variables" estimator (see also "Deming" regressions). Essentially, it minimizes the squared *perpendicular* distances (rather than horizontal or vertical) between each point and the line, exactly as in principal component analysis. In fact, in our bivariate case, the TLS slope is equal to the ratio of the first two entries of the leading eigenvector of the covariance matrix. In practice, we demean the data, estimate the TLS slope $\hat{b}$ and find the intercept as $\hat{a} = \bar{y} - \hat{b}\bar{x}$, where $\bar{y}$ and $\bar{x}$ are sample averages.

B.2.2 Covariances

Here we report the covariance matrices for the strengths and degrees of Ecuador (2015) and Hungary (2021). Since we are interested in log-transformed values, we need to drop the zeros. When a node has an in-strength of zero, it always has an in-degree of zero; similarly for out-strength and out-degree. However, it is possible for a node to have suppliers but not to have any customers, or the other way around. As a result, we need to report two covariance matrices. Table B.1 shows the variances and covariances computed by removing only the nodes that have a value of zero for a specific metric or pair of metrics. Instead, Table B.2 reports the covariance matrix computed after all the nodes that have at least one zero value are removed. These tables allow the reader to compute results that we do not report explicitly in the main text. We provide 3 examples.

[24] If β is the coefficient of the OLS regression of y on x, $\beta = \text{Cov}(y,x)/\sigma_x^2$, and $\tilde{\beta}$ is the coefficient of the regression of x on y, we have $\beta = \rho^2(1/\tilde{\beta})$, where ρ^2 is the squared correlation coefficient, that is, the R^2 from both regressions. Thus $\beta = (1/\tilde{\beta})$ iif $\rho = 1$ or -1.

Table B.1: Covariance matrix keeping only nodes with pairwise positive values

	Ecuador (2015)				Hungary (2021)			
	k^{out}	k^{in}	s^{out}	s^{in}	k^{out}	k^{in}	s^{out}	s^{in}
k^{out}	2.83	1.37	2.49	2.39	2.73	1.14	2.85	1.92
k^{in}	1.37	2.36	1.98	3.66	1.14	2.00	1.20	2.71
s^{out}	2.49	1.98	8.17	4.34	2.85	1.20	7.99	2.98
s^{in}	2.39	3.66	4.34	8.99	1.92	2.71	2.98	6.10
Mean	1.85	2.83	10.62	9.91	1.67	3.05	8.61	9.25

Notes: All variables are log-transformed. The row "Mean" shows the average of the log-transform of the positive values.

Table B.2: Covariance matrix keeping only nodes which have positive values for all four metrics

	Ecuador (2015)				Hungary (2021)			
	k^{out}	k^{in}	s^{out}	s^{in}	k^{out}	k^{in}	s^{out}	s^{in}
k^{out}	2.89	1.37	2.12	2.39	2.75	1.14	2.19	1.92
k^{in}	1.37	2.06	1.98	3.07	1.14	1.65	1.20	2.19
s^{out}	2.12	1.98	6.97	4.34	2.19	1.20	5.90	2.98
s^{in}	2.39	3.07	4.34	7.63	1.92	2.19	2.98	5.22
Mean	2.07	3.19	11.08	10.53	2.04	3.30	9.38	9.64

Notes: All variables are log-transformed.

Example 1: Total Least Squares. In Section 4.2, we report the TLS estimate for the relationship between in- and out-strengths as 0.93. This is the ratio between the two values of the first eigenvector of the covariance matrix. For Ecuador, the matrix (from Table B.2) $\begin{pmatrix} 6.97 & 4.34 \\ 4.34 & 7.63 \end{pmatrix}$ has eigenvector $(0.679, 0.733)$, leading to a TLS estimate of $0.679/0.733 = 0.93$, as reported.

Example 2: Least squares. In Table 9, we report regressions of (log) strengths on (log) degrees. For instance, for the regression of out-strengths on out-degrees for Hungary in 2021, the coefficient is $\hat{\beta} = \frac{Cov(\ln s^{\text{out}}, \ln k^{\text{out}})}{Var(\ln k^{\text{out}})} = 2.85/2.73 = 1.044$, as reported (up to rounding errors).

Example 3: Large variances. In Section 4.2, we report that the strength distributions can also be well-fitted by lognormal distributions. It is well-known that it is very hard to distinguish lognormal distributions from distributions with regularly varying tails when the lognormal scale parameter is large. Sornette (2006, pg. 95), for instance, uses a value up to $\sigma = 3$ to make this point. We can calculate that fitting a lognormal for the in-strength distribution of Ecuador, for instance, leads to $\hat{\sigma} = \sqrt{8.99} \approx 3$.

B.2.3 Heteroskedasticity and non-linearities

To investigate whether the conditional relations feature non-linearities and/or heteroskedasticity, we use binned scatter plots. To gauge nonlinearities, we use GLM, and to gauge heteroskedasticity, we use quantile regressions at the 10^{th} and 90^{th} level, both as implemented by Cattaneo et al. (2019) and using 100 bins. Figure B.1 shows the results for Ecuador and Hungary, considering all two conditional relationships from the two (i.e., in- or out-) strength-degree joint relationships.

Overall, linearity appears to hold fairly over large ranges. However, it is interesting that the deviations from linearity (computed by simple OLS) are almost systematically the same in Ecuador and Hungary. Regarding homoskedasticity, there is a noticeably smaller variance of the strengths when considering intermediate values of the number of partners (top two rows). For the degree-strength relationship, there is a very clear trend of increasing variance of the number of customers as we condition on higher and higher sales. In other words, firms with very high sales may have many customers – on average, they do – but it is not uncommon to find very large firms having just a few customers. This fact does not appear to have been noted in the literature. It could be due to the fact that we observe only intermediate domestic customers.

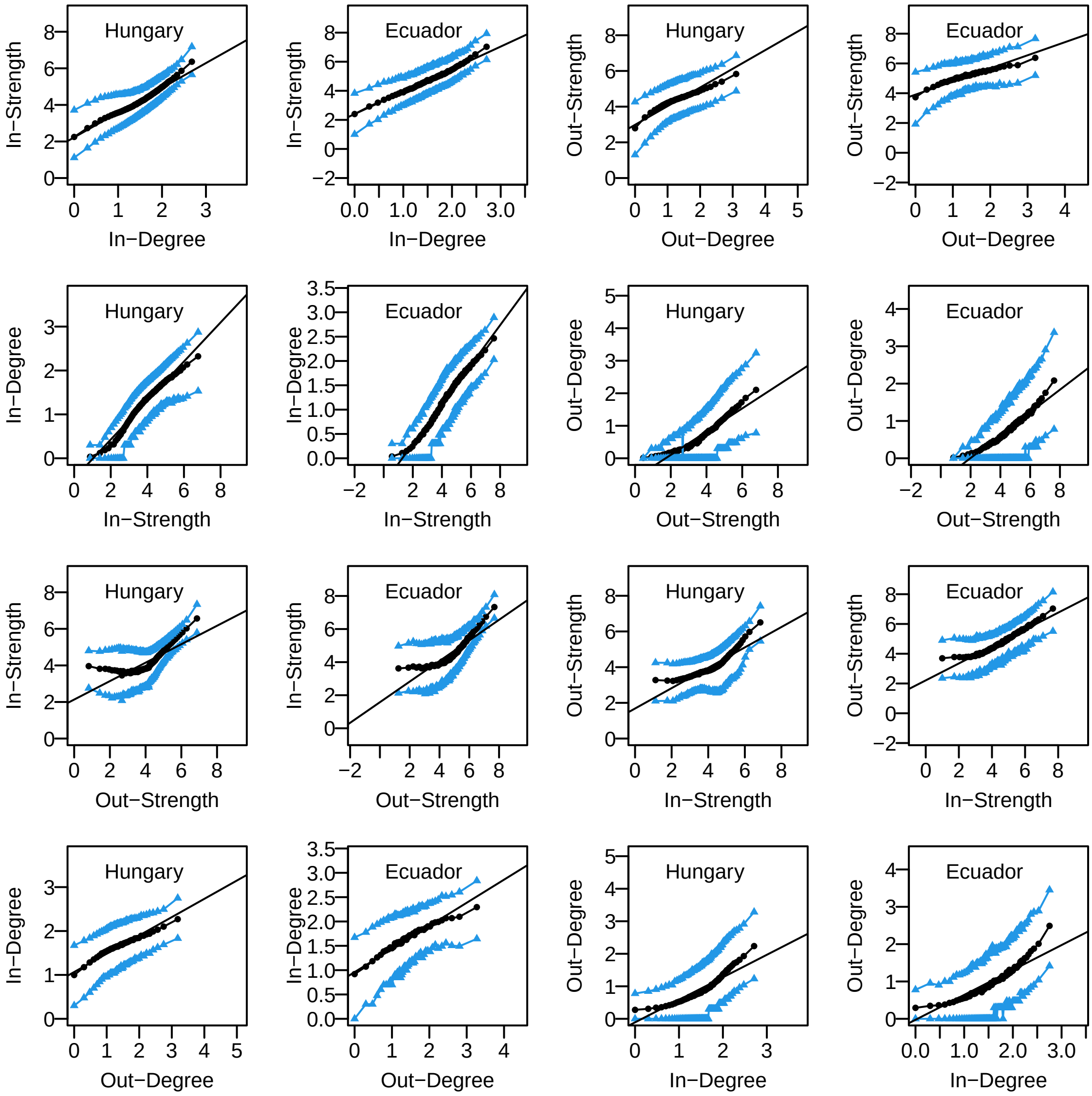

Figure B.1: Binned scatter plots for the conditional relations. The black dots show the non-parametric GLM estimates and the blue dots show the non-parametric 10^{th} and 90^{th} quantile regressions, using the implementation of Cattaneo et al. (2019) and 100 bins. The ranges of the x- and y-axis are determined by the range of the data to highlight that the last bins in the tails reflect a large range of data. All axes are in log base 10.

C Additional results

C.1 Density and growth

Table C.1 shows binary network statistics for the networks analysed in the literature we reviewed and for Ecuador, Hungary and FactSet, the network we analyse. The rest of the section gives additional details on how we constructed the Table.

For Costa Rica and the Dominican Republic, we have to infer the number of edges. For Costa Rica (Alfaro-Urena et al., 2018), we take the number of firms from Table 8 in Alfaro-Urena et al.

(2018). For the total number of edges, we multiply the number of customers by their average degree, (Alfaro-Urena et al., 2018, Table 6).

For the Dominican Republic (Cardoza et al., 2020), we take the number of firms in 2017 from Table 1 in the reference paper and infer the number of firms in 2012 using the 2012–2017 growth rate of 30.3% reported on pg. 9. Cardoza et al. (2020) report that only 3% of firms are supplier-only, so we estimate the number of customer firms by multiplying our estimated number of firms by 0.97. Finally, we get the number of edges by multiplying the number of customers by their average degree (Table A1b in the reference paper).

Table C.1: Binary network statistics.

Country	Year	N	E	$\bar{k}$	Source(s)
Ecuador	2007	56,058	1,873,023	33.41	This paper
	2011	72,200	2,774,900	38.43	This paper
	2015	86,345	3,372,929	39.06	This paper
Hungary	2015	119,469	356,788	2.99	This paper
	2019	313,117	2,116,912	6.76	This paper
	2021	493,616	18,710,235	37.90	This paper
FactSet	2015	201,389	502,429	2.50	This paper
	2017	203,445	515,668	2.53	This paper
	2020	214,605	523,648	2.44	This paper
Belgium	2002–2012	400,000	88,437,335	221.09	Dhyne et al. (2015)
	2002	88,301	4,187,000	47.42	Dhyne et al. (2021)
	2007	95,941	4,848,000	50.53	Dhyne et al. (2021)
	2012	98,745	5,026,000	50.9	Dhyne et al. (2021)
	2012	79,788	3,505,207	43.93	Magerman et al. (2016)
	2012	250,000	8,700,000	34.8	Dhyne and Duprez (2015)
	2014	321,824	8,900,000	27.65	Dhyne et al. (2016)
	2014	94,334			Bernard et al. (2022)[25]
Turkey manufacturing	2010–2014	600,000	6,000,000	10	Demir et al. (2022)[26]
5 Indian states	2011–2012 & 2015–2016	2,500,000	130,000,000	10	Panigrahi (2023)
Uganda	2009–2016	83,000	420,000	5.06	Spray (2017)
	2010–2015	100,428			Spray and Wolf (2018)
	2010	29,274			Spray and Wolf (2018)
	2014	41,578			Spray and Wolf (2018)
Rwanda	2009–2014	65,193			Spray and Wolf (2018)
	2010	18,714			Spray and Wolf (2018)
	2014	32,330			Spray and Wolf (2018)
Brazil	2003–2014	6,200,000	410,000,000	66	Silva et al. (2020)
Chile	2005		5,670,000		Huneeus (2020)
	2008		6,830,000		Huneeus (2020)
	2011		6,580,000		Huneeus (2020)
	2014–2020			20.3	Miranda-Pinto et al. (2022)
Costa Rica	2008–2015	60,478	1,995,970	33.00	Alfaro-Urena et al. (2018)

[25] Firms in the GSCC; these have at least two customers and suppliers.
[26] The number of firms and links are approximate.

Country	Year	N	E	$\bar{k}$	Source(s)
Dominican Republic	2017	39,161			Cardoza et al. (2020)
Kenya	2015	33,090	88,6940	26.80	Chacha et al. (2022a)
Kenya	2016	38,655	1,134,159	29.34	Chacha et al. (2022a)
Kenya	2017	43,145	1,204,754	27.92	Chacha et al. (2022a)
Kenya	2018	48,027	1,332,150	27.74	Chacha et al. (2022a)
Kenya	2019	51,749	1,528,410	29.54	Chacha et al. (2022a)
Kenya	2020	53,584	1,528,109	28.52	Chacha et al. (2022a)
Netherlands	2019	100,000	1,000,000	10	Ialongo et al. (2022)
Spain	2008	245,524	2,328,908	9.49	Peydró et al. (2020)
Spain	2009	243,936	2,040,869	8.37	Peydró et al. (2020)
Japan	2005	785,939	3,338,319	4.25	Bernard et al. (2019)
	2005	961,318	3,667,521	3.82	Ohnishi et al. (2010)
	2006	1,019,854	4,041,442	3.96	Fujiwara and Aoyama (2010).
	2008	552,145		6	Mizuno et al. (2015)
	2009	541,816			Mizuno et al. (2015)
	2010	518,565			Mizuno et al. (2015)
	2010	1,600,000			Lu et al. (2017)
	2011	520,087			Mizuno et al. (2015)
	2012	525,836			Mizuno et al. (2015)
	2012	1,109,549	5,106,081	4.6	Inoue (2016)
	2013	1,610,000			Lu et al. (2017)
Japan electronics	1993	227	648	2.85	Luo and Magee (2011); Luo et al. (2012)
Japan automotive	1983	356	1,480	4.16	Luo et al. (2012)
	1993	679	2,437	3.59	Luo and Magee (2011); Luo et al. (2012)
	2001	627	2,175	3.47	Luo et al. (2012)
Global automotive	10/2013 to 01/2014	18,942	103,602	5.47	Brintrup et al. (2015)
U.S. listed	04/2012 to 06/2013	2,152	11,819	5.49	Wu and Birge (2014)
	1979–2007	39,815	14,204	0.36	Atalay et al. (2011)
	1980–2004	30,622 min = 390 max = 1,470 mean = 918 median = 889 SD = 291	11,484	0.38	Cohen and Frazzini (2008)
	1980–2009		48,839		Herskovic et al. (2020)
	1978–2013		21,528		Barrot and Sauvagnat (2016)
Global listed	1994–2015	23,059	2,257,761	97.91	Wu (2016)
Global listed cleaned	1994–2015	10,930	1,007,998	92.22	Wu (2016)

Notes: "year" is the year of observation, N is the number of firms, E is the number of edges and $\bar{k}$ is the average degree. The supply chain network of global listed firms in Wu (2016) is taken from FactSet, Bloomberg, 8-K filings and the US Customs Bill of Lading; they subsequently merge this with customer-supplier relations provided by Capital IQ. They report summary statistics before and after cleaning the data set so that every firm in the final data set has cleaned financial statements.

C.2 Degree distributions

Table C.2 and C.3 show the power-law exponents of the in- and out-degree distributions (CCDFs) estimated using the method of Clauset et al. (2009) (marked γ^{plfit}) and the three estimators of Voitalov et al. (2019) based on extreme value theory.

Table C.2: Tail exponents for in-degree distributions

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007	2.07	2,018	2.89	272	5.83	401	2.90	25,420
2008	2.06	2,391	4.09	10	5.88	156	2.76	8,725
2009	2.16	2,239	3.61	88	4.85	168	2.99	10,005
2010	2.25	2,403	3.01	274	5.20	302	3.14	13,013
2011	2.16	3,353	3.05	155	4.24	364	3.26	18,001
2012	2.36	2,462	3.15	209	4.05	558	3.33	21,053
2013	2.33	2,734	3.68	70	3.07	4,893	2.62	61,805
2014	2.70	1,127	2.68	890	3.07	4,059	3.47	39,256
2015	2.38	2,900	2.66	553	2.97	5,275	3.61	32,169
Hungary								
2015	1.62	1,162	1.65	628	1.75	920	1.41	33,200
2016	1.66	836	1.62	530	1.74	845	1.39	25,303
2017	1.35	6,663	1.71	365	1.77	1,144	1.38	44,455
2018	1.66	3,916	2.03	172	2.26	971	1.97	8,376
2019	1.83	2,696	2.24	115	2.02	4,800	2.12	7,711
2020	2.51	3,545	2.50	3,615	2.68	13,046	2.70	102,261
2021	2.69	2,246	2.72	1,578	2.86	13,167	2.83	103,178
FactSet								
2014	1.74	2,789	2.20	179	2.00	3,303	2.09	5,530
2015	1.78	2,044	1.83	1,656	2.01	2,903	1.91	12,831
2016	1.77	1,928	1.81	1,280	1.95	3,307	2.07	6,260
2017	1.82	1,351	1.85	979	1.94	3,654	2.07	5,215
2018	1.83	1,058	2.37	19	1.92	3,890	2.03	7,460
2019	1.80	1,217	2.54	15	1.93	3,090	2.02	4,924
2020	1.72	1,853	2.37	15	1.88	4,119	1.99	7,049

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points).

Table C.3: Tail exponents for out-degree distributions

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007	1.26	1,963	1.79	66	2.01	128	1.63	2,385
2008	1.82	190	1.71	76	2.23	78	1.39	5,630
2009	1.38	934	1.63	80	1.98	218	1.41	4,829
2010	1.13	3,126	1.63	61	1.82	118	1.33	6,863
2011	1.40	944	1.60	97	1.86	141	1.74	1,513
2012	1.36	972	1.56	83	1.79	181	1.77	958
2013	1.65	210	1.59	95	1.83	269	1.60	1,753
2014	1.64	228	1.59	106	1.87	158	1.57	1,811
2015	1.59	220	1.58	90	1.76	236	1.55	1,616
Hungary								
2015	1.46	2,771	1.45	2,029	1.44	13,531	1.32	81,169
2016	1.43	3,739	1.45	2,148	1.47	8,219	1.42	43,544
2017	1.44	4,752	1.45	3,752	1.49	5,215	1.49	16,399
2018	1.61	1,687	1.59	1,215	1.65	2,794	1.72	2,616
2019	1.62	1,444	1.63	1,334	1.65	2,607	1.72	7,187
2020	1.43	3,577	1.41	4,542	1.40	12,007	1.49	20,985
2021	1.42	4,081	1.41	3,722	1.40	10,346	1.38	34,717
FactSet								
2014	2.69	871	2.61	276	2.93	1,049	3.33	5,257
2015	2.81	668	2.81	338	3.46	1,068	3.83	3,323
2016	2.55	1,057	2.74	99	3.30	1,553	4.12	2,627
2017	2.51	969	2.73	133	3.13	1,358	3.25	5,789
2018	2.71	414	2.75	221	3.07	1,085	3.53	3,417
2019	2.29	802	2.40	493	2.71	1,187	3.01	2,673
2020	2.36	573	2.37	407	2.65	1,018	2.80	3,888

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points).

C.3 Input and output shares

The input shares are computed as $P_{ij}^{\text{in}} = Z_{ij}/\sum_i Z_{ij}$, and the output shares as $P_{ij}^{\text{out}} = Z_{ij}/\sum_j Z_{ij}$, where Z_{ij} is the payment from j to i. Magerman et al. (2016) calculate the input shares of Belgian firms and find that its distribution is heavy-tailed with a mean of 0.02 and a standard deviation of 0.08: a supplier accounts for 2%, on average, of a firm's intermediate input mix. Table C.4 shows that two of our complete networks have a mean and standard deviation strikingly similar to those of Belgium. Output shares have similar moments.

Table C.4: Summary statistics for the input and output shares.

Type	Country	Year	Mean	Median	Standard dev.	Source
Input share	Ecuador	2015	0.0217	0.0008	0.0907	This paper
Input share	Hungary	2015	0.2122	0.0481	0.3181	This paper
Input share	Hungary	2019	0.1064	0.0137	0.2210	This paper
Input share	Hungary	2021	0.0213	0.0012	0.0834	This paper
Input share	Belgium	2012	0.0200	0.0030	0.0800	Magerman et al. (2016)
Input share	Belgium	2012	0.0180	0.0019		Kikkawa et al. (2019)
Output share	Ecuador	2015	0.0205	0.0002	0.1045	This paper
Output share	Hungary	2015	0.2640	0.0679	0.3566	This paper
Output share	Hungary	2019	0.1211	0.0114	0.2530	This paper
Output share	Hungary	2021	0.0228	0.0003	0.1128	This paper

Similar findings are reported by Kikkawa et al. (2019), who also characterize the distribution of input shares as (roughly) lognormal. Figure C.1 shows the distributions of Ecuador (top) and Hungary's (bottom) input and output shares, displaying roughly a bell-shaped pattern for the log shares. In all the distributions for our complete networks, there is a clear mode around 0.1%. While small input shares are the most common, it is not rare that a supplier or customer represents a large fraction of costs or sales (including 100%).

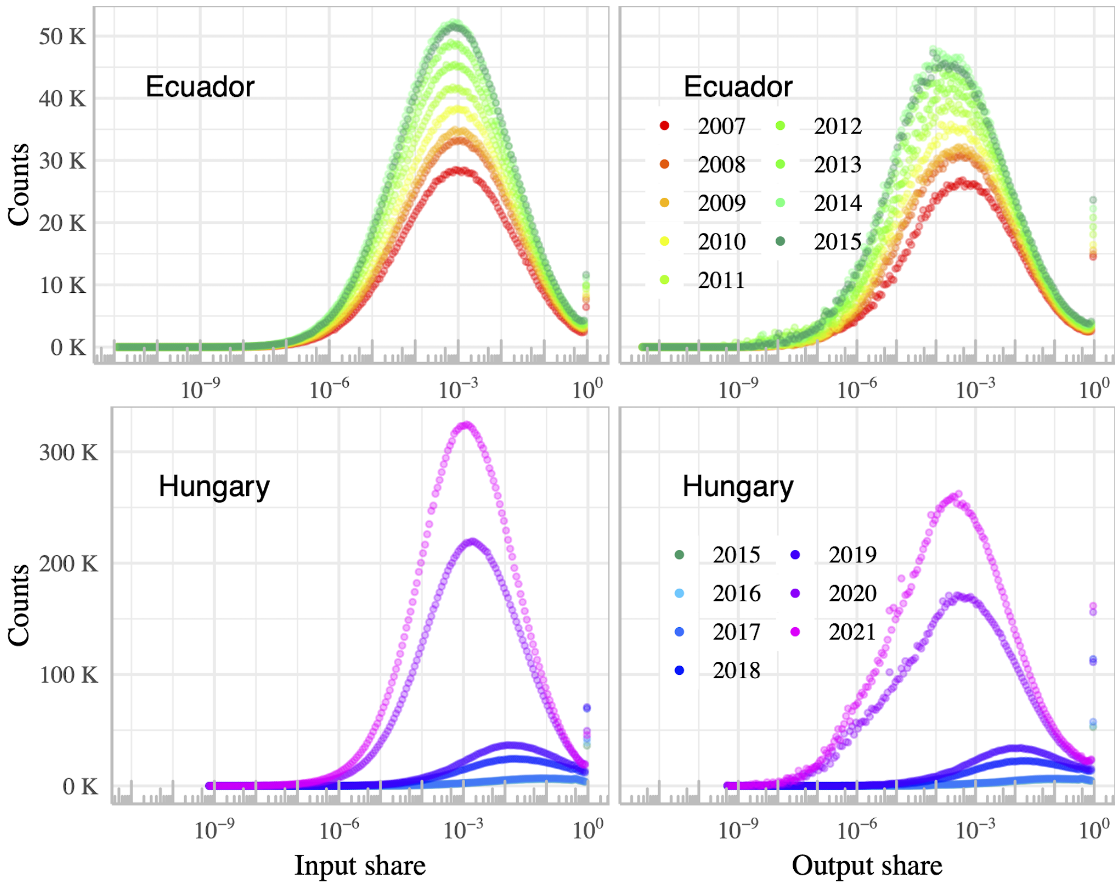

Figure C.1: Empirical pdf of the input shares (left) and the output shares (right) for Ecuador (top) and Hungary (bottom) over time on a semi-log scale. We binned the data into 200 log-spaced bins.

C.4 Strength distributions

Figure C.2 shows the distribution of in- and out-strengths for Ecuador and Hungary, while Figure C.3 shows the 2-D scatter of in- and out-strength with the Total Least Squares estimate.

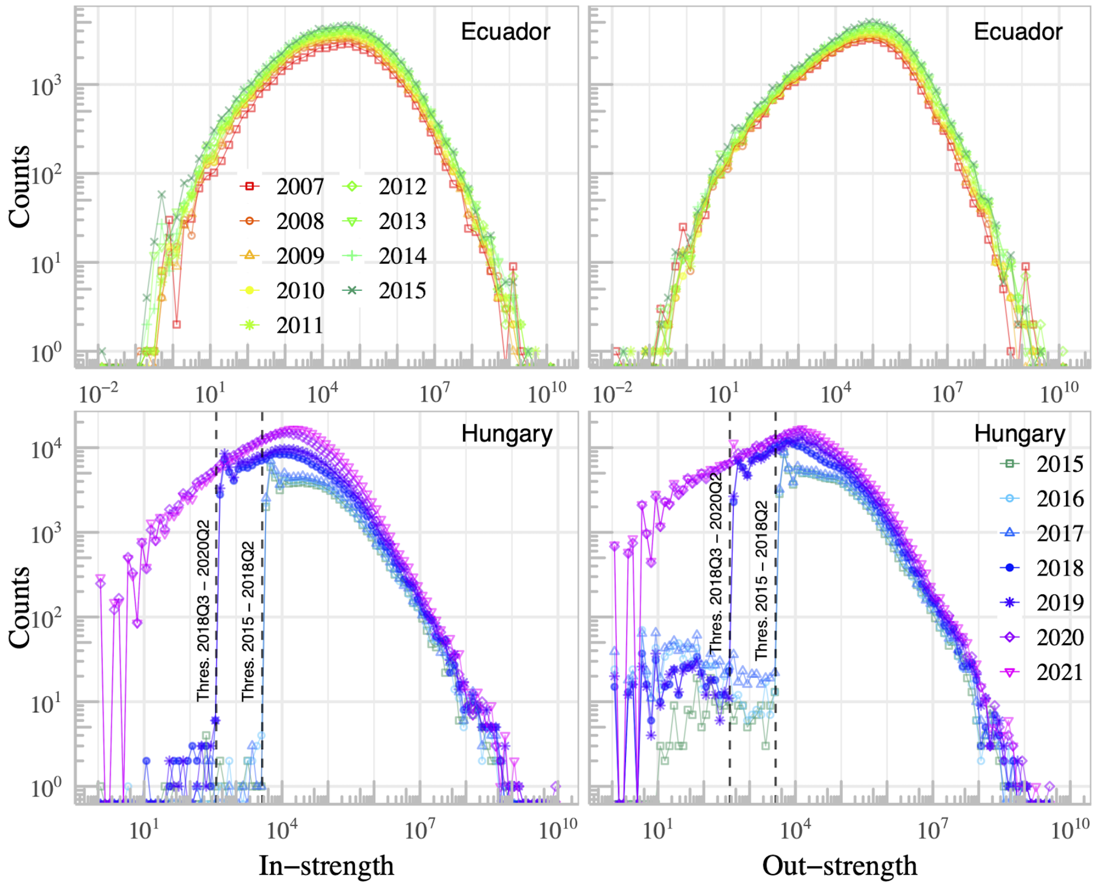

Figure C.2: Empirical pdf of the in- and out-strength (network expenses and sales, respectively) for Ecuador (left) and Hungary (right) over time. We used 80 log-spaced bins for Ecuador and 100 for Hungary. The two vertical lines for Hungary mark the reporting thresholds; see description of Figure 1 and Appendix A.5. The values are in USD for Ecuador and in 1,000 HUF for Hungary.

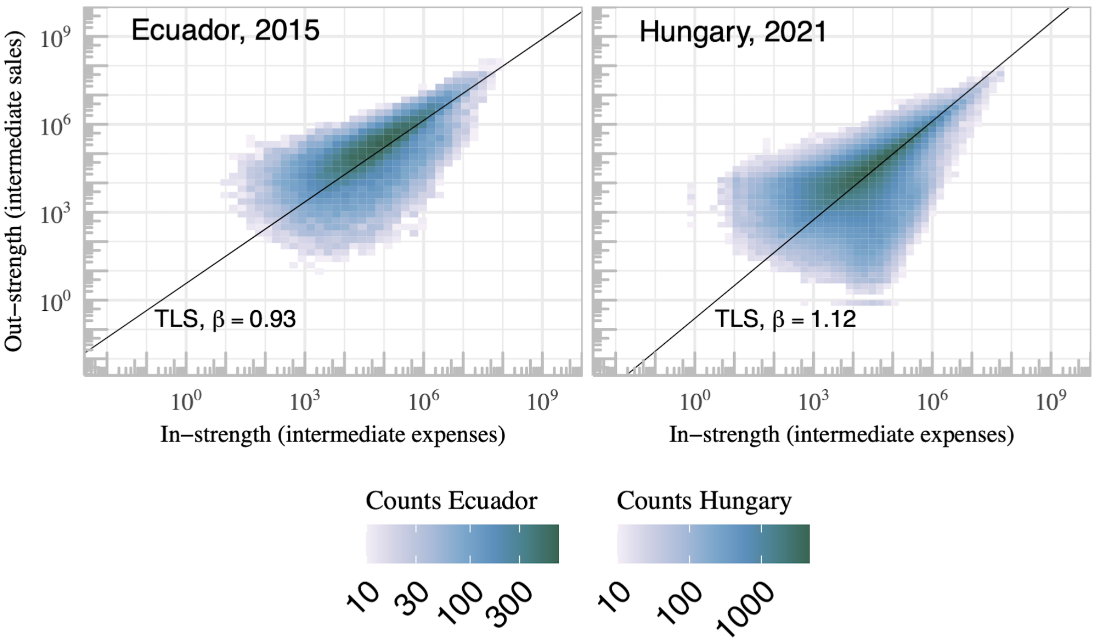

Figure C.3: 2D histogram for network expenses on the x-axis and network sales on the y-axis for Ecuador in 2015 (left) and Hungary in 2021 (right). Ecuador is in 2015 USD and Hungary in 2021 1,000 HUF. We divide both axes into 60 equally-spaced bins and count the number of data points in each square. We do not show squares that have less than 10 observations. The counts are log-transformed.

Tables C.5 and C.6 show the estimated power-law exponents using the three estimators of Voitalov et al. (2019) based on extreme value theory and the estimator of Clauset et al. (2009).

Table C.5: Tail exponents for in-strength distributions

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007	2.07	41,244	0.96	539	0.96	1,579	0.97	3,843
2008	2.06	48,030	0.93	1,391	0.98	1,835	1.00	3,358
2009	2.16	49,920	1.05	448	1.02	1,706	1.29	604
2010	2.25	53,239	1.04	448	1.05	1,365	1.16	904
2011	2.16	57,618	1.07	347	1.14	448	1.11	1,486
2012	2.36	60,903	0.96	950	1.05	813	1.04	1,447
2013	2.33	63,513	1.11	278	1.20	398	0.94	7,489
2014	2.70	66,375	1.12	300	1.21	349	0.92	8,900
2015	2.38	68,123	1.09	280	1.14	596	0.91	10,879
Hungary								
2015	1.62	75,717	1.06	887	1.11	1,086	0.93	20,238
2016	1.66	79,806	1.11	467	1.17	757	0.95	15,809
2017	1.35	87,531	2.78	1	0.92	10,331	0.95	12,453
2018	1.66	191,109	1.14	305	0.96	10,704	0.96	25,189
2019	1.83	225,218	1.00	2,922	1.05	2,963	1.00	20,751
2020	2.51	332,591	0.97	10,532	1.00	14,183	1.00	35,474
2021	2.69	356,679	1.01	10,331	1.02	19,466	1.03	53,149

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points).

Table C.6: Tail exponents for out-strength distributions

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007	1.26	43,227	1.04	454	1.05	278	0.95	5,542
2008	1.82	41,225	1.13	322	1.22	317	0.90	15,116
2009	1.38	45,252	1.09	383	1.20	562	0.92	12,474
2010	1.13	50,272	1.22	28	1.22	480	0.92	13,931
2011	1.40	52,107	1.15	209	1.32	160	0.92	14,850
2012	1.36	55,443	1.09	279	1.08	322	0.91	13,953
2013	1.65	54,394	1.23	147	0.96	4,869	0.98	7,570
2014	1.64	57,045	1.45	35	0.99	4,000	1.01	6,517
2015	1.59	59,178	1.27	225	1.54	131	1.00	7,160
Hungary								
2015	1.46	94,198	1.56	86	1.53	115	0.97	23,157
2016	1.43	96,040	1.60	71	1.54	88	0.98	22,143
2017	1.44	103,225	1.39	125	0.96	12,594	0.98	19,780
2018	1.61	233,494	0.99	4,513	1.01	10,339	1.05	6,494
2019	1.62	255,898	1.00	5,404	1.02	10,293	1.04	17,404
2020	1.43	314,797	0.98	5,763	1.00	10,538	1.01	20,985
2021	1.42	329,869	1.03	5,938	1.47	118	1.07	10,609

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points).

C.5 Weight distributions

Table C.7 shows the estimated power-law exponents using the three estimators of Voitalov et al. (2019) and the estimator of Clauset et al. (2009). Since Clauset et al.'s (2009) method is computationally intensive on the weight distributions, we have used their method for one year only, for Ecuador, and for three years, for Hungary. For Hungary, we choose the years corresponding to the changes in the reporting threshold.

C.6 Influence vector distributions

Table C.8 shows the estimated power-law exponents for the CCDF of the influence vector over time using the three estimators of Voitalov et al. (2019) and the estimator of Clauset et al. (2009).

Table C.7: Tail exponents for weight distributions

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007			1.17	1,175	1.14	3,223	1.16	5,584
2008			1.19	812	1.23	611	1.26	1,903
2009			1.21	304	1.31	741	1.36	1,674
2010			1.22	278	1.19	3,335	1.19	5,561
2011			1.22	723	1.22	677	1.21	4,555
2012			1.09	3,587	1.12	2,566	1.11	5,913
2013			1.13	2,594	1.16	3,537	1.18	8,206
2014			1.15	1,650	1.19	3,178	1.21	6,119
2015	1.14	5,093	1.02	14,847	1.22	2,541	1.22	4,995
Hungary								
2015	1.15	15,095	1.15	10,503	1.16	21,424	1.07	195,787
2016			1.18	6,587	1.18	13,412	1.05	213,150
2017			1.13	13,860	1.14	21,887	1.02	265,206
2018			1.19	3,680	1.21	7,675	1.12	159,117
2019	1.14	10,879	1.15	6,663	1.17	8,705	1.19	16,201
2020			1.06	53,908	1.13	8,914	1.13	49,951
2021	1.18	15,128	1.18	9,953	1.18	17,327	1.20	35,569

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points). We do not compute all the years for `plfit` due to computational constraints.

Table C.8: Tail exponents for the distributions of the influence vector

	plfit		Hill		Moment		Kernel	
	γ	κ	γ	κ	γ	κ	γ	κ
Ecuador								
2007	1.37	3,347	1.33	2,257	1.37	7,809	1.36	27,922
2008	1.31	3,012	1.29	2,551	1.34	9,475	1.32	31,471
2009	1.28	2,274	1.30	2,709	1.34	9,081	1.32	32,570
2010	1.26	2,060	1.27	2,328	1.32	9,276	1.31	36,427
2011	1.29	3,515	1.29	3,199	1.32	9,451	1.32	38,962
2012	1.28	2,838	1.28	2,633	1.32	9,412	1.30	45,742
2013	1.27	2,827	1.27	2,902	1.32	9,742	1.32	43,119
2014	1.25	3,292	1.25	3,423	1.30	9,377	1.30	50,265
2015	1.28	3,472	1.28	2,991	1.30	8,576	1.33	48,382
Hungary								
2015	1.44	5,301	1.44	5,385	1.48	28,369	1.42	24,044
2016	1.39	2,784	1.40	3,886	1.43	15,081	1.37	16,535
2017	1.40	3,375	1.39	3,043	1.44	15,661	1.39	87,531
2018	1.37	11,694	1.38	13,534	1.41	62,979	1.42	91,644
2019	1.37	6,831	1.39	9,844	1.41	37,422	1.36	36,709
2020	1.40	12,383	1.40	8,258	1.40	29,169	1.30	366,587
2021	1.40	11,249	1.40	8,442	1.39	29,218	1.30	398,050

Notes: Parameters estimated using `plfit` and the three estimators of the tail index of the generalized extreme value distribution. κ is the smallest order statistics used for estimation (i.e., the number of data points).

Reconstructing firm-level input-output networks from partial information

Andrea Bacilieri[1,2,3,*] and Pablo Austudillo-Estevez[3,4]

[1]Institute for New Economic Thinking, University of Oxford, Oxford, OX1 3UQ, UK
[2]Smith School of Enterprise and Environment, University of Oxford, Oxford, OX1 3QY, UK
[3]School of Geography and the Environment, University of Oxford, Oxford, OX1 3QY, UK
[4]Universidad San Francisco de Quito & School of Economics, Quito, 170902, Ecuador
[*]Corresponding author. *E-mail address:* andrea.bacilieri@sant.ox.ac.uk (A. Bacilieri).

Abstract

There is large consensus on the fundamental role of firm-level supply chain networks in macroeconomics. However, data on supply chains at the fine-grained, firm level are scarce and frequently incomplete. For listed firms, some commercial datasets exist, but only contain information on the existence of a trade relationship between two firms, not the value of the monetary transaction. We use a recently developed maximum entropy method to reconstruct the values of the transactions based on information about their existence and aggregate information disclosed by firms in their financial statements. We test the method on the administrative dataset of Ecuador and reconstruct a commercial dataset (FactSet). We test the method's performance on the weights, the technical and allocation coefficients (microscale quantities), two measures of firms' systemic importance and GDP volatility. The method reconstructs the distribution of microscale quantities reasonably well but shows divergent results for the measures of firms' systemic importance. Due to the network structure of supply chains and the sampling process of firms and links, quantities that more prominently depend on the number of customers firms have (out-degrees) are harder to reconstruct. We also reconstruct the input-output table of globally listed firms and merge it with a global input-output table at the sector level (the WIOD). Differences in accounting standards between national accounts and firms' financial statements significantly reduce the quality of the reconstruction.

Keywords: Network reconstruction, supply chain, production network, input-output table, maximum entropy, missing information

JEL codes: C80, D57, E32, L14, F12

1 Introduction

Recent events such as the war in Ukraine and the increasing frequency of natural disasters have highlighted the fragility of global supply chains. Shocks to a single firm or a cluster of firms, sometimes located in a specific geographical region, can quickly spread through the network, with severe repercussions on the global economy. Most research to date has been conducted at the sectoral level (Acemoglu et al., 2012; Carvalho, 2014; Pichler and Farmer, 2021). However, analysing such shocks at the sector level can lead to misleading results (Diem et al., 2022, 2023). The value of firm-level data is therefore increasingly recognised, but research efforts are constrained by data availability.

Due to confidentiality and the data collection process, supply chain data are scarce, hard to access, and frequently incomplete. Countries collect the best data sources through VAT filings, but only a handful of countries collect them; for a comprehensive review and discussion of the different datasets, see Bacilieri et al. (2023). Some datasets of broader global coverage derived from US disclosure requirements are available. Given their wider breadth of coverage and relatively easier access, such datasets are used in many studies (e.g., Wu, 2016; Wang et al., 2021; Pankratz and Schiller, 2019; Taschereau-Dumouchel, 2020; Boehm and Sonntag, 2022; Barrot and Sauvagnat, 2016; Atalay et al., 2011; Herskovic et al., 2020). In contrast to national datasets, which report the monetary value of the transactions (i.e., the weighted network; an exception is Japan), global datasets do not provide this valuable information.

Reconstructing firm-level production networks is thus an important topic, showing growing research interest. The reconstruction problem concerns mainly two features of the production network: supplier-customer relations and transaction values. Several studies develop methods to infer both links and transaction values (Reisch et al., 2021; Ialongo et al., 2022; Hooijmaaijers and Buiten, 2019; Hillman et al., 2021) or links only (Brintrup et al., 2018; Mungo et al., 2022; Kosasih and Brintrup, 2022), and two focus on inferring weights given the binary topology (Inoue and Todo, 2019; Welburn et al., 2020). The assessment of the quality of the weights reconstruction is usually missing and sometimes carried out on aggregate quantities or compared to empirical facts about another country's network.

In this paper, we focus on reconstructing the transaction values using partial information on the supply chain relations and aggregate information on firms' revenues and expenditures. We provide the first rigorous assessment of a recently developed maximum entropy reconstruction method (Parisi et al., 2020) using the administrative dataset of Ecuador. We evaluate the method on microscale, higher-order and macroscale quantities that are widely used in economic input-output (I-O) models. We also reconstruct the production network of globally listed firms (FactSet). However, we can only judge the quality of the reconstructed network as FactSet does not provide the monetary value of the transactions.

We assess the method's performance at recovering the link weights, and the technical and allocation coefficients (i.e., normalised weights). We also use two indicators of firms' systemic importance, the output multipliers and the influence vector, which are prominent in macroeconomic I-O models of shock propagation. While the reconstruction method reproduces the weights (normalised or not) rather poorly, it reconstructs their distributions reasonably well. In contrast, the reconstruction shows diverging results for higher-order quantities: the output multipliers are in remarkable agreement with the empirical values, while the influence vector is overestimated. We then use a general equilibrium I-O model (Acemoglu et al., 2012) to assess how shocks to firms' total factor productivity (TFP) propagate through the network, ultimately affecting aggregate GDP fluctuations. We show that aggregate volatility is overestimated by the reconstruction method we employ.

Our results suggest that quantities that rely more on the number of customers firms have are more adversely affected by missing firms and links due to the structure of supply chain networks and the sampling process of firms and links. We also find that including a proxy node to represent the rest of the economy not captured by the network improves the prediction of microscale quantities and measures of systemic importance, but worsens the prediction of aggregate volatility.

An additional contribution we make in this paper is to construct the I-O table of globally listed firms using the dataset collected by FactSet. We merge FactSet with the World Input-Output Database (WIOD). Key challenges related to differences in accounting standards between national accounts and firms' financial statements prevent us from (1) merging the two datasets at the desired country-sector or even sector level and (2) carrying out an accurate quantification (at the firm level) of the key variables making up an I-O table. We then reconstruct and compute weights, coefficients and higher-order quantities for FactSet as well. The inability to accurately quantify the variables of the I-O table at the firm level dramatically reduces the quality of the final dataset and thus also the quality of the reconstruction.

The remainder of the paper is organised as follows. In Section 2, we review the literature on network reconstruction. Section 3 delves into the network reconstruction method. Specifically, in Section 3.1, we explain the notation and define the production network at the firm level. In Section 3.2, we concisely describe the reconstruction method we employ. Section 3.3 details the metrics we use to assess the performance of the reconstruction method. In Section 4, we discuss the two datasets we use. We then present and discuss the results for microscale (Section 5.1), higher-order and macroscale properties (Section 5.2) for Ecuador and FactSet. Section 5.3 shows results for different numbers of unknown links and Section 6 concludes.

2 Literature review

Methods for reconstructing networks with missing information have mostly been developed for financial or trade networks (e.g., Moussa, 2011; Mastrandrea et al., 2014; Cimini et al., 2015b; Gandy and Veraart, 2017; Anand et al., 2015) and I-O tables at the sector level (e.g., Golan et al., 1994; Robinson et al., 2001; Lenzen et al., 2009). Only a few studies develop methods for reconstructing firm-level production networks (Inoue and Todo, 2019; Welburn et al., 2020; Reisch et al., 2021; Hooijmaaijers and Buiten, 2019; Ialongo et al., 2022; Hillman et al., 2021). We start with a brief overview of the different reconstruction methods developed in the literature, which we divide into deterministic and ensemble methods, and subsequently give a more detailed account of the methods developed to reconstruct firm-level networks. We refer to Squartini et al. (2018) and Cimini et al. (2021) for reviews on reconstruction methods developed mostly for financial and trade networks, and to Miller and Blair (2009), McDougall (1999) and Lahr and De Mesnard (2004) for reviews on sector-level reconstruction methods and matrix balancing problems.

The network reconstruction problem, be it for financial, trade or production networks, boils down to inferring a matrix of bilateral flows among entities (e.g., banks, countries or sectors) given constraints on the total in- and out-flows of each entity (i.e., in- and out-strengths) and other information when available, such as the degrees (i.e., the number of connections entities have) or a prior bilateral flows matrix. Most of the reconstruction methods are based on the maximum entropy principle, of which there are two strands: deterministic and ensemble methods. Deterministic methods yield a single reconstruction of the weighted network while meeting the constraints exactly. Instead, ensemble methods sample many networks from a distribution that is constructed to respect the constraints on average. Therefore, while ensemble methods generate a probability distribution over the likely networks, deterministic methods assign a probability of one to the reconstructed network and a zero probability to all the other networks – which likely include the true network (Parisi et al., 2020). The shortcomings that most deterministic and ensemble methods share are that they tend to create a fully connected network with weights distributed as uniformly as possible given the imposed constraints on the in- and out-flows. To create sparser networks, algorithms with tunable (Moussa, 2011; Upper, 2011; Mastromatteo et al., 2012) or exact network density (Mastrandrea et al., 2014; Cimini et al., 2015b) have been developed.

The most well-known deterministic method is known as *MaxEnt*. It maximises an entropy-like functional subject to constraints on the in- and out-strength of each node. The solution to this maximisation yields the well-known gravity model (without distance) in the international trade literature (first proposed by Tinbergen, 1962 and Pöyhönen, 1963; see also Squartini and Garlaschelli, 2014 for a discussion). MaxEnt displays all the shortcomings mentioned above: it generates a fully connected network and the weights are distributed as equally as possible given the constraints. To enhance the MaxEnt reconstruction, if some prior information about the binary topology or the weights is available, it can be integrated using a cross-entropy method (Golan et al., 1994; Di Gangi et al., 2018; Upper, 2011; Wells, 2004). The cross-entropy method reconstructs a network that has minimum distance to the prior while accounting for the imposed constraints. The cross-entropy method is equivalent to the iterative proportional fitting (IPF) algorithm, which iteratively distributes the weights (coming from the MaxEnt solution or any other prior) among the non-zero edges until the row and column sums are satisfied. The IPF algorithm is also known as the RAS technique in the I-O literature (Miller and

Blair, 2009). If the network is fully connected, the IPF algorithm is equivalent to MaxEnt.

There are different ensemble methods depending on the information used for the reconstruction (e.g., in- and out-degrees or strengths sequences). We discuss the method developed by Cimini et al. (2015b) since it is one of the best performing (Anand et al., 2018; Lebacher et al., 2021). To enhance the reconstruction of the weighted network, Cimini et al. (2015b) impose constraints on both the in- and out-strength sequences and on the degrees. Their strategy is motivated by recent results showing that the strengths do not encode information about the binary topology (although they are correlated with degrees) and that the degrees are "fundamental" local structural properties of weighted networks (Mastrandrea et al., 2014). Combining constraints on the degrees and strengths thus greatly enhances the reconstruction of weighted networks since the degrees provide information about the binary topology that strengths do not, enabling to identify better the matrix of link probabilities as well as higher-order properties (Mastrandrea et al., 2014; Gandy and Veraart, 2017).

The method proposed by Mastrandrea et al. (2014) requires knowledge of the degrees and strengths of all the nodes, which are not always available. Cimini et al. (2015b) note that in financial networks, the in- and out-strengths are usually known while the degrees might be known for a subset of nodes only. To account for these two pieces of information, Cimini et al. (2015b) restore to the *fitness* ansatz. The fitnesses are nodes' non-topological futures that relate to the ability of nodes to establish connections: nodes with higher fitness attract more connections and are thus likely to become hubs (Squartini et al., 2018; Mazzarisi and Lillo, 2017). Given the empirical correlation frequently observed among strengths and degrees, strengths are often used as a proxy for nodes' fitnesses. To estimate the binary topology, Cimini et al. (2015b) thus develop a fitness-induced configuration model that estimates link probabilities using the nodes' fitnesses and the degrees of only a few nodes. The weights are estimated in a second step using a degree-corrected gravity model that accounts for the sparsity of the adjacency matrix inferred in the first step.

How does one choose which reconstruction method to use? Anand et al. (2018), Lebacher et al. (2021) and Ramadiah et al. (2020) find that the choice of the reconstruction method ultimately depends on the feature of the network one aims to reconstruct, which usually boils down to connectivity structure versus weights. If one cares about inferring links, methods that focus on reconstructing a sparse connectivity structure are better suited. Anand et al. (2018), Lebacher et al. (2021) and Mazzarisi and Lillo (2017) conclude that the best-performing ones are the degree-corrected gravity model (Cimini et al., 2015b), the minimum density (Anand et al., 2015) and the Bayesian hierarchical fitness (Gandy and Veraart, 2017). If one cares about inferring the link weights, methods based on MaxEnt or the IPF algorithm perform best. Since financial networks have many link weights of relatively equal size (something that is less likely to be the case in firm-level networks; Bacilieri et al., 2023), they score well on weight-based similarity measures (Anand et al., 2015). In their horse races, Anand et al. (2018) and Lebacher et al. (2021) identify the methods developed by Cimini et al. (2015b) and Baral and Fique (2012) to be the best performing. Altogether, the reconstruction method proposed by Cimini et al. (2015b) seems to be the best in reconstructing both binary and weighted topological features.

Firm-level network reconstruction. Welburn et al. (2020) reconstruct the network of listed firms in the US, using a dataset that covers only the major customers of these firms.[1] Therefore, they know the binary topology only partially. To reconstruct the weighted network, they use a two-step procedure. In the first step, they infer missing links using a logistic regression. In the second step, they develop a linear programming method to reconstruct the unknown weights given the links. Since they do not observe the whole economy, they introduce a proxy node that captures the rest of the economy and to which each firm is linked. The cumulative in- and out-flows of the proxy node are then minimised subject to constraints on firms' revenues and the cost of goods sold. While there are almost 6,000 firms in their network, they only reconstruct the network composed of 1,000 firms due to the high computational complexity of the second step of their procedure. Inoue and Todo (2019) reconstruct the weighted network of Japanese firms given the binary topology.[2] Firstly, they assume that the link weight is proportional to the supplier and customer's sales. Subsequently, they re-adjust the estimated weights using the I-O table at the sector level to ensure that if the firm-level network is aggregated

[1] A major customer accounts for 10% or more of a firm's annual revenues. They retrieved the customer-supplier relations from firms' filings available through the EDGAR database.

[2] The network is collected by a private company. Although the coverage is extensive, comprising almost 890,000 firms, it is not exhaustive.

at the sector level, it is consistent with national accounts. In a similar fashion and using the same Japanese dataset, Carvalho et al. (2021) assume that the technical coefficient between two firms is proportional to the technical coefficient between the sectors those two firms are in.

Hillman et al. (2021) reconstruct the global network of private and public firms in the ORBIS database. As done in other reconstruction methods, they build the firm-level network so that it is consistent with sector-level I-O data (OECD). In each step, a firm i is chosen at random and its sales are split into n units that are then sold to n different customers. i's customers are chosen according to the industry they are in based on sector-level I-O tables, meaning that if firm i is in industry s, its customers need to be in one of the industries to which s sells (in the sector-level I-O table). They further calibrate their model on the observation that larger firms tend to have more customers (Bernard et al., 2019). This feature can be controlled by changing how, at the beginning of each step, firm i's sales are split into n units. For computational reasons, they only use 5,000 firms among the more than 200 million firms in the Orbis database. Hooijmaaijers and Buiten (2019) reconstruct the supply chain network of the Netherlands using several microdata sources available to the Office of National Statistics. They describe their method as being akin to maximum entropy methods with exact link density (as classified in Squartini et al., 2018), but they pose additional constraints thanks to the richness of their microdata and to findings in the literature about empirical facts of firm-level supply chain networks. A novel feature of their reconstruction is the disaggregation of firms' output into different goods.

None of the studies just described can assess how well their method recovers the empirical weighted network because none of them has access to it. Two studies assess their reconstruction method, at least to some extent. First, Reisch et al. (2021) reconstruct a firm-level production network using mobile communication data. To guarantee anonymity, the company providing the data and the country are not disclosed. Roughly speaking, links are inferred by assuming that if two firms communicate with each other, they are involved in a supply-chain relationship. To determine the link direction, they use the national I-O table at the sector level and information about the sectors the customer and supplier are in. A gravity model is then used to estimate the link weights, where a firm's size is given by its total assets. To assess the reconstruction of the binary topology, they compare to the Hungarian network, whereas to assess the performance of the method regarding the weights reconstruction, they use the Economic Systemic Risk Index and compare with results obtained for Hungary by Diem et al. (2022). They find similarities between results obtained for Hungary and their reconstructed network. Second, Ialongo et al. (2022) develop the stripe-corrected gravity model, which builds on the degree-corrected gravity model (Cimini et al., 2015b) by adding constraints on the in-strength of each industry. They test their method on two transaction data made available by two Dutch banks. The assessment is carried out on the degree and strength distributions, degree-strength correlations and average nearest neighbour strength.

3 Network reconstruction

We begin this section by explaining the notation we use and by defining the production network at the firm level. We then concisely describe the method we use. Afterwards, we explain the metrics we employ to assess the performance of the reconstruction method.

3.1 Firm-level input-output tables

This section describes firm-level production networks and gives an example of the supply chain network of publicly listed firms we aim to reconstruct (Section 4.1). We then explain I-O tables and outline key differences between I-O tables at the sector and firm level.

The production or supply chain network is composed of N firms (nodes) and links between firms indicate yearly trading relationships. Links may be weighted, where each weight w_{ij} represents firm j's intermediate input expenditure on goods produced by i. We label the weighted adjacency $\boldsymbol{W}$ and $\boldsymbol{A}$ the binary adjacency matrix. Figure 1 shows the binary (left) and weighted (right) adjacency matrices depicting the empirical data collected by FactSet. Both matrices have on the i-th row the customers of firm i, while column j lists the suppliers of the j-th firm. The unknown weights are labelled as question marks. Given $\boldsymbol{A}$ (and other aggregate information about firms that we outline below), we aim to reconstruct $\boldsymbol{W}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{W} = \begin{bmatrix} 0 & ? & 0 & ? \\ 0 & 0 & ? & 0 \\ ? & 0 & 0 & ? \\ ? & ? & 0 & 0 \end{bmatrix}$$

Figure 1: Example of the data we aim to reconstruct. **Left:** Binary directed adjacency matrix. **Right:** Weighted directed adjacency matrix.

For each firm, we can define its (total) intermediate expenditure and sales as, respectively, the column and row sums of the weighted adjacency matrix. The column and row sums are also called the weighted in- and out-degrees or in- and out-strengths; they are given by

$$\boldsymbol{s}^{\text{in}} = \boldsymbol{W}^{\top}\mathbb{1}, \text{ and} \tag{1}$$

$$\boldsymbol{s}^{\text{out}} = \boldsymbol{W}\mathbb{1}, \tag{2}$$

where $\mathbb{1}$ is a vector of ones of appropriate size.

The supply chain network just described captures only a part of the economic activity, namely firm-to-firm trades. Transactions with other economic actors (e.g., households) are captured in an input-output table, usually at the sector level. One can also define an input-output table at the firm level, however with notable differences. For a more in-depth discussion, we refer to Appendix A.4.

In I-O studies of production networks, the weighted adjacency matrix is usually normalised using firms' total costs instead of their in-strengths. A firm's total costs are the costs of intermediate inputs plus value-added, which is itself composed of labour costs, depreciation, amortisation and profit (see Appendix A.2.3). The weights so normalised are called the *technical coefficients* and represent the percentage of inputs firm j buys from firm i. The technical coefficients are given by

$$T_{ij} = \frac{W_{ij}}{\sum_i W_{ij} + y_j} \,, \tag{3}$$

where y_j is value-added of firm j.

Similar to the technical coefficients, one can define the *allocation coefficients*, $\boldsymbol{B}$. B_{ij} tells us the percentage of output firm i sells to firm j. Letting f_i be the amount of final demand satisfied by firm i, the allocation coefficient is defined as

$$B_{ij} = \frac{W_{ij}}{\sum_j W_{ij} + f_i} \,. \tag{4}$$

3.2 Method

To reconstruct the weighted network given the binary topology, we use the conditional maximum entropy ensemble reconstruction method developed by Parisi et al. (2020). We do not use any of the previously developed methods for reconstructing firm-level networks because either they are too computationally expensive (Welburn et al., 2020; Hillman et al., 2021), demand too many data inputs that we do not have (Hooijmaaijers and Buiten, 2019) or would imply constraining the firm-level network with sector-level I-O tables (Inoue and Todo, 2019; Ialongo et al., 2022). We disregard the latter methodologies because we think using sector-level data in the way proposed by Inoue and Todo (2019) or in the spirit of Ialongo et al. (2022) could bias the reconstruction in unwanted ways given the underlying differences in accounting standards between national I-O tables and firms' financial statements (see Appendix A.4).

Parisi et al. (2020) develop a maximum entropy method that reconstructs an ensemble of likely weighted networks given some prior information about the binary ensemble and aggregate information about each node. The procedure is flexible in that it allows to use of an observed binary topology or to infer it in a previous step and account for the additional uncertainty. They proposed two methods that use different constraints. One method constrains the in- and out-strength sequences, while the other one constrains the expected link weights (and the in- and out-strengths indirectly). We choose the second method for two reasons. First, it is computationally more efficient since it involves solving

m (the number of links) decoupled equations, while the method constraining the in- and out-strengths require solving $2N$ coupled equations, where N is the number of firms. Second, the authors show that it predicts the weights better compared to the model constraining the in- and out-strengths.

The method consists of two steps. The first step constrains the in- and out-strengths (total intermediate expenditure and sales) of each firm and determines the values of the expected link weights used as constraints in the second step. As discussed in Section 3, MaxEnt reconstructs the weights best among the other methods (Anand et al., 2018; Lebacher et al., 2021); thus, MaxEnt is used in the first step. The second step allows us to account for any prior information about the binary topology and generates an ensemble of weighted networks compatible with this information and with the constraints on the expected link weights.

As discussed, MaxEnt assumes a fully connected network; however, the conditional maximum entropy method assumes no self-loops. Additionally, in our case, we know the binary topology. To redistribute the weights corresponding to $A_{ij} = 0, \forall\ (i,j) \notin \mathcal{E}$, where $\mathcal{E}$ is the edge set, we employ the IPF algorithm. The IPF algorithm redistributes the weights in an iterative procedure until the constraints on the in- and out-strengths are met (see Appendix B.1). Parisi et al.'s (2020) method turns the IPF algorithm into a probabilistic method and allows calculating confidence intervals around each reconstructed weight. We give a brief outline of the method below and describe it in detail in Appendix B.1.

First step. The method derives values for the expected link weights to enforce as constraints in the second step. The values of the weights are derived by solving the MaxEnt problem, which maximises an entropy-like functional subject to constraints on intermediate sales and costs of each firm. The value of each weight is given by

$$W_{ij}^{\mathrm{ME}} = \frac{s_i^{\mathrm{out}*} s_j^{\mathrm{in}*}}{W^{\mathrm{tot}*}} \ ,$$

where $W^{\mathrm{tot}*} = \sum_i s_i^{\mathrm{out}*} = \sum_j s_j^{\mathrm{in}*}$ is the total weight of the empirical network.

Second step. The method maximises the conditional entropy defined over the probability density function of the weighted networks compatible with the prior on the binary ensemble and subject to constraints on the expected weights. Solving the conditional maximum entropy problem yields that the probability of observing a weight $W_{ij} > 0$ given that there is a link between i and j is of exponential form with parameter λ_{ij} (which also corresponds to the Lagrange multiplier):

$$Q_{ij}(W_{ij} \mid A_{ij} = 1) = \lambda_{ij} e^{-\lambda_{ij} W_{ij}},\ W_{ij} > 0. \tag{5}$$

To find the values of the λ_{ij}'s, one maximises the log-likelihood function, which leads to the first order conditions

$$\langle W_{ij} \rangle = \frac{p_{ij}}{\lambda_{ij}} \ , \ \forall\ i \neq j. \tag{6}$$

Since for each link the expected weight $\langle W_{ij} \rangle = W_{ij}^{\mathrm{ME}}$, the Lagrange multipliers are given by

$$\lambda_{ij}^* = p_{ij} \frac{W^{\mathrm{tot}*}}{s_i^{\mathrm{out}*} s_j^{\mathrm{in}*}} \ , \ \forall\ i \neq j. \tag{7}$$

We set $p_{ij} = 1, \forall (i,j) \in \mathcal{E}$.

Confidence interval on the expected edge weight. For each expected weight, the confidence interval is $[w^-, w^+]$. The lower bound is given by

$$w^- = -\frac{\ln[e^{-1} + q^-]}{\lambda_{ij}^*} \ , \tag{8}$$

where q^- is a desired confidence level and the upper bound is given by

$$w^+ = -\frac{\ln[e^{-1} - q^+]}{\lambda_{ij}^*} \ . \tag{9}$$

We set $q^+ = q^- = 0.25$. We refer to Appendix E in Parisi et al. (2020) for the derivation.

3.3 Assessing the reconstruction

We assess how well the reconstruction method can recover the empirical network at three scales. First, we look at microscale quantities: weights, and technical and allocation coefficients. Second, we evaluate the reconstruction of higher-order properties (i.e., multipliers): the output multipliers and the influence vector. Third, we turn to macroscale properties and use a general equilibrium I-O model to study how the propagation of shocks through the network affects GDP volatility. We conclude this section by defining the statistical indicators used to compare the empirical and reconstructed quantities.

3.3.1 Weights

As discussed in Section 3.1, technical and allocation coefficients (Equation 3 and 4, respectively) are normalised weights frequently used in economic models. Therefore, we assess the reconstruction of both coefficients and the weights: their exact values and their distributions.

3.3.2 Higher-order properties

We look at two higher-order properties widespread in the economic literature: the output multipliers and the influence vector. These are two centrality measures that quantify firms' contributions to economy-wide fluctuations.

The output multipliers. The output multipliers are derived from the Leontief model (Miller and Blair, 2009) and are defined as

$$\mathcal{O} \equiv (\boldsymbol{I} - \boldsymbol{T}^{\top})^{-1}\mathbb{1} \ , \tag{10}$$

where $\boldsymbol{I}$ is the identity matrix and $\mathbb{1}$ a vector of ones, both of appropriate size. The output multiplier captures the upstream propagation channel of an exogenous shock to a firm's final demand and its economy-wide impacts (where final demand increases by one monetary unit; Miller and Blair, 2009). It can also be seen as the average length of a firm's production chain (McNerney et al., 2022; Fally, 2012; Miller and Temurshoev, 2017). The higher the output multiplier is, the longer the production chain is on average and the greater the impact of a change in a firm's final demand is on the whole economy.

The influence vector. The influence vector is derived from the Cobb-Douglas model proposed by Acemoglu et al. (2012). The influence vector is defined as

$$\boldsymbol{v} \equiv \frac{\alpha}{N}[\boldsymbol{I} - (1-\alpha)\boldsymbol{\Omega}]^{-1}\mathbb{1} \ , \tag{11}$$

where $\boldsymbol{\Omega}$ is the matrix of input shares with $\omega_{ij} \in [0,1]$ being the share of input i used in j's production process ($\sum_i \omega_{ij} = 1$), $\alpha \in (0,1]$ is the share of labour and N is the number of firms.

Contrary to the output multipliers, the influence vector captures the downstream propagation channel of TFP shocks and it gauges the contribution of firms to fluctuations in aggregate GDP. Positive TFP shocks can be thought of as firms' innovating their production processes and becoming more efficient in using their inputs. The influence vector is equivalent to a Reverse Weighted PageRank with a damping factor equal to $(1-\alpha)$.

3.3.3 Macroscale properties

As highlighted by recent crises and natural disasters, assessing how different shocks affect the economy is of paramount importance to be better prepared in preventing or alleviating crises. Therefore, after assessing the weights and the multipliers, we study how supply-side shocks at the firm level propagate through the network to downstream firms, ultimately leading to fluctuations in aggregate GDP.

Supply-side shocks and aggregate volatility. We model supply shocks as shocks to firms' TFP. We use the model developed by Acemoglu et al. (2012) and refer to the paper for a derivation. In the competitive equilibrium, aggregate volatility is given by

$$\sigma_{\Delta y} = \sqrt{\sum_i \mathrm{Var}(\Delta\epsilon_i) v_i^2} \ , \tag{12}$$

where v_i is the influence of firms i (Equation 11), $\Delta\epsilon_i$ is the TFP shock of firm i and ϵ_i is an i.i.d. random variable with mean zero and bounded variance. Equation 12 shows that productivity shocks at the firm level affect aggregate value-added through the production network. Initially, the shock propagates to the customers of the affected firm and, subsequently, propagates downstream to the customers' customers and so on, potentially spreading through the whole supply chain network.

Note that the model gives a static representation of the economy in equilibrium and firms' input shares are exogenous; therefore, the network structure and hence the influence vector are constant over time. We use Equation 12 to assess the impact of firms' TFP shocks on GDP volatility using either the empirical or the reconstructed weighted network given the TFP shocks.

Simulating TFP shocks. To estimate TFP shocks, we cannot use an econometric technique (e.g., Magerman et al., 2016) because it requires knowledge of several variables that we do not observe for Ecuador. Besides requiring all the variables describing a firm's production function, the estimation of TFP requires a time series of these variables. Since our goal is not to empirically validate the economic model but to assess the discrepancy in the predicted GDP volatility when the reconstructed influence vector is used instead of the empirical one, we simulate firm-level TFP shocks. For FactSet, we do not perform this test since it would entail estimating TFP, which is outside of the scope of this paper.

We simulate TFP shocks from a normal distribution with mean zero and standard deviation of 6. We chose a zero mean in line with the empirical mean reported for the Belgian production network by Magerman et al. (2016). Since they do not report the standard deviation, we set it to 6 so that GDP volatility, calculated using the model with the true network, matches the observed country-level volatility.[3] We set the TFP shock of the proxy sector equal to the median of the TFP shocks of firms that were excluded from our test network. We simulate 10 time periods. Figure 2 shows the distribution of the simulated TFP growth rates in Panel (a) and the distribution of their standard deviation in Panel (b). Once we simulate the TFP volatilities, we use Equation 12 to predict the fluctuation in aggregate GDP using either the empirical or the reconstructed influence vector. We then compare the empirical volatility with that predicted by the reconstruction.

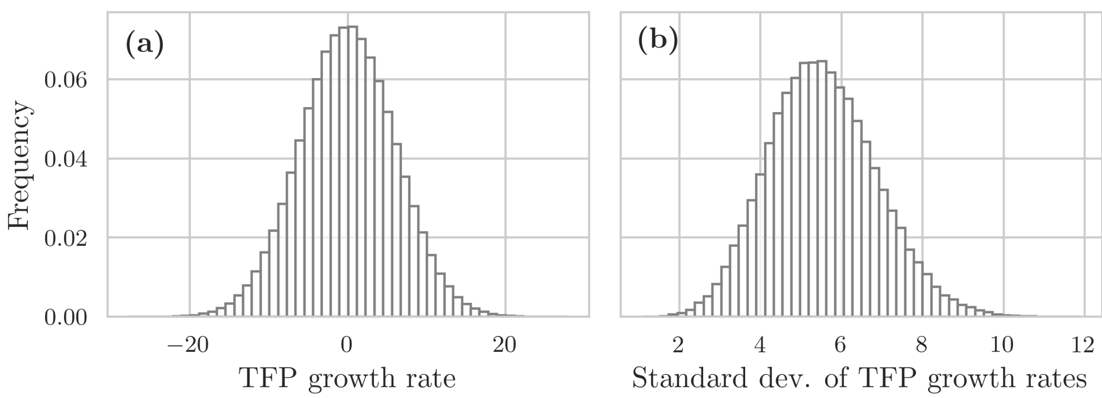

Figure 2: Distribution of TFP **(a)** growth rates and **(b)** their standard deviation. The growth rates are pooled over the 10 simulated years. We binned the data into 50 equally-spaced bins.

3.3.4 Statistical indicators

To compare how well the reconstruction method can recover the quantities defined in Section 3.3.1, 3.3.2 and 3.3.3, we employ metrics that are standard in the literature: the L_1-error, the root-mean-squared error (RMSE), the mean and median absolute error (MAE and MedAE, respectively) and the cosine similarity.

[3]To calculate the GDP volatility in Ecuador, we use data from the IMF; available at https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sId=1479331931186. We use nominal GDP since our data are not adjusted for inflation. Since we simulate 10 years, we calculate GDP volatility for the period 2005-2015, which is 6.35%.

We tried different parametrisations of the normal distribution, which do not match the volatility in the empirical data, and results do not change, at least qualitatively.

The L_1-error assesses the degree to which constraints on intermediate sales and costs are violated. It is defined as

$$L_1 = \sum_i |s_i^{\text{in}} - s_i^{\text{in}^*}| + \sum_i |s_i^{\text{out}} - s_i^{\text{out}^*}| \; ,$$

s_i^{in} is firm i's in-strength in the reconstructed network and s_i^{out} its out-strength in the reconstructed network; quantities with a * refer to observed, empirical values.

In what follows, we define the error measures using the technical coefficients, but they similarly apply to any other quantity of interest. We do not use the RMSE, the MAE and the MedAE to assess the raw weights because their distribution has heavy tails. For the technical and allocation coefficients, and the multipliers, we further normalise the metrics to allow their comparison across variables that have different scales. We rescale by ϕ, the difference between the maximum and minimum value (excluding the zeros) of the empirical quantity of interest. Our rescaled measures compare the variation in the residuals to the range of the empirical data. For instance, a normalised RMSE of 0.1 means that the variation in the residuals is 10% of the range of variation of the empirical data. The lower the normalised error metric is, the better the reconstruction is. The normalised root-mean-square error is given by

$$\text{RMSE} = \frac{1}{\phi}\sqrt{\frac{1}{m}\sum_{ij}(T_{ij} - T_{ij}^*)^2} \; ,$$

where m is the number of links. The normalised mean absolute error is given by

$$\text{MAE} = \frac{1}{\phi m}\sum_{i,j}|T_{ij} - T_{ij}^*| \; .$$

The normalised median absolute error is defined as

$$\text{MedAE} = \frac{\text{Median}\Big(|\boldsymbol{T} - \boldsymbol{T}^*|\Big)}{\phi} \; .$$

The cosine similarity is defined as

$$\vartheta = \frac{\sum_{ij} T_{ij}^* T_{ij}}{\sqrt{\sum_{ij} T_{ij}^{*2}}\sqrt{\sum_{ij} T_{ij}^2}} \; .$$

4 Data

FactSet is the global production network that we aim to reconstruct and for which we do not know the values of the monetary transactions. Therefore, we test the reconstruction method on the administrative dataset of Ecuador, for which we know the monetary values of the transactions. While Ecuador is not a global dataset, it is an ideal case study as it (1) collects information on supply chain relationships for all firms in the formal economy and (2) firms have to report the value of the transaction so we can assess the quality of the reconstruction method. As explained in Section 4.2, we engineer our test network so that it is comparable (as much as possible) to FactSet, meaning that we artificially remove firms and links that are more likely to be missing in FactSet.

4.1 FactSet

We use three primary data sources provided by FactSet: Fundamentals, Supply Chain Relationships and Supply Chain Shipping Transactions.[4] FactSet covers mainly listed firms around the world. The supply chain relationships of these companies are collected through two primary sources: company filings required by US Federal Accounting Standards (Supply Chain Relationships) and import and export declarations at ports from the US Bill of Lading (Supply Chain Shipping Transactions).[5] FactSet also collects information on supply chain relationships from investor presentations, company

[4]The datasets were downloaded in April 2020.

[5]The Statement of Financial Accounting Standards No. 131 requires publicly traded firms on US stock exchanges to report customers that account for 10% or more of their annual revenues, formally called *major customers*.

websites and press releases. Due to the nature of the data collection process, coverage is biased toward companies listed on US stock exchanges, large firms and large transactions. For a more detailed description of FactSet, see Appendix A.2.

We aggregate customer-supplier relations within a fiscal year to ensure time consistency between the formation of supplier-customer relations and financial statements.[6] We further aggregate all three datasets at the parent company level. For each company, we also have information on the sector (NACE Rev.2 codes at the 4-digit level) and the country where the company's headquarters are located.

We use several variables from companies' income statements (FactSet Fundamentals): revenues, the cost of goods sold, labour expenses, earnings before interest and taxes (EBIT), depreciation and amortisation. We convert all the variables to USD using the currency conversion tables provided by FactSet. We define value-added as the sum of labour expenses, EBIT, amortisation and depreciation (Appendix A.2.3). Some firms do not disclose their labour costs and include them in the costs of goods sold; we estimate these firms' labour expenses (see Appendix A.2.2).

For simplicity, we limit ourselves to the 2014 network, which we call henceforth "FactSet". We keep firms with positive sales, intermediate expenses and value-added, and with non-negative labour costs (see Appendix A.2.4). We exclude firms in financial and insurance, extraterritorial organisations and bodies and activities of households as employers. The number of firms in the 2014 cleaned dataset is 5,442; these are involved in 15,916 trading relations. The average degree is 2.9.

Evaluation of coverage and proxy node. In 2014, FactSet captures around 16.4% of world gross output as reported in the WIOD (see Appendix A.3.4). To capture the rest of the economic activity that we do not capture in FactSet, we introduce a "proxy" node in the network to which all firms are connected. We construct the proxy node's variables (gross output, intermediate sales and expenditure, etc.) using the WIOD aggregated at the world level.

A good approach would be to integrate FactSet with the WIOD at the country and sector level. However, due to differences between national accounting standards and firms' financial statements, we had to aggregate the WIOD at the world level. We refer to Appendix A.4 for a detailed discussion of how we integrated the two datasets.

4.2 Ecuador

Ecuador collects customer-supplier relations through VAT filings, which are mandatory for firms and natural persons. We do not have access to firms' financial statements, so we do not know their revenues, labour costs or profit, but we know the sectors' firms are in (ISIC Rev. 4 codes). The Ecuador dataset was provided by Ecuador's government to one of the authors; we refer to Appendix A.1 for more information.

Constructing the test network. While Ecuador's dataset has comprehensive coverage, FactSet does not. Therefore, we construct a test network that mimics the missing firms and links in FactSet. To mimic the missing firms in FactSet, we keep the same number of firms in Ecuador that we have in FactSet. We choose to keep the largest firms (in terms of out-strength) since we observed predominantly large firms in FactSet. We also require firms to have positive in-strengths, meaning that firms need to buy some inputs from the other firms in the subgraph.[7] The resulting network, which we call "test network" (third column in Table 1), has a much higher average degree compared to FactSet, meaning that the average Ecuadorian firm is connected to many more firms than the average firm in FactSet.

To mimic the missing links in FactSet, we eliminate links at random in the "test network" until we match FactSet's average degree. Since FactSet's supply chain relations mostly cover customers that account for 10% or more of a firm's annual revenues, we delete links with a smaller weight with a higher probability: we set the link deletion probability to be inversely proportional to the link weight $p_{ij} \propto 1/W_{ij}$. We do this procedure 50 times and reconstruct each of the 50 randomised networks. The summary statistics for these networks, which we call "trimmed test network", are shown in the last column of Table 1. In deleting links to match FactSet's average degree, 96% of the links among firms

[6] The fiscal year goes from June to May, meaning that if a company's fiscal year end-month falls between January and May, the fiscal year is the current calendar year minus one; otherwise, it is the current calendar year.

[7] 4 firms are further dropped because they do not have any links with the other firms in the sampled subgraph.

in the test network are deleted. Consequently, our results can be interpreted as an approximate lower bound on the quality of the reconstruction.

We aggregate the firms and transactions left out of the test network in one proxy node representing the rest of the economy. As done for FactSet, we establish an incoming and outgoing link between each firm and the proxy node.

Summary statistics	Full network	Test network	Trimmed test network
N. nodes	84,978	5,440	5,440
N. edges	3,439,975	432,910	15,776
Average degree	40.5	79.6	2.9

Table 1: Summary statistics for the 2014 Ecuador network. The first column "Full network" is the network composed of all the firms except those in sectors: finance, insurance, activities of households as employers, activities of extraterritorial organisations and bodies, and those that have no sectoral code. The third column, "Test network", refers to our test network, which is composed of the top 5,440 firms with the largest total intermediate sales and positive intermediate expenses. The last column "Trimmed test network" refers to the network built from the "Test network" by further deleting links at random to match FactSet's average degree. The summary statistics of the test and trimmed test network do not include the proxy node.

Inferring missing data. For Ecuador, we do not have information on final demand, revenues and the variables that compose value-added (i.e., labour costs, depreciation, amortisation and profits). To carry out the analyses described in Section 3.3, we need final demand and value-added of each firm. Therefore, we simulate final demand and value-added using the 2014 I-O table of Ecuador at the sector level.[8] Consider value-added (a similar procedure is done for final demand), for each sector s, we calculate the ratio of value-added to intermediate expenditure $\nu_s = y_s/s_s^{\text{in}}$. Assuming that a firm's ratio is the same as that of the sector the firm is in, a firm's value-added is given by $y_i = \nu_s \cdot s_i^{\text{in}}$.

5 Results

This section presents and discusses the results for Ecuador and FactSet. It starts with microscale properties, proceeds to higher-order and macroscale properties, and concludes with results for different numbers of unknown links.

5.1 Link weights

In this section, we show the results of the reconstruction method for the weights, and the technical and allocation coefficients. We start by discussing the results for our test network, Ecuador, and then show the results for FactSet, for which we do not know the ground truth.

We compare weights (normalised or not) for firms only and not those of the proxy node since the proxy node can be thought of as a sink node and does not meaningfully represent either a firm or a sector. For parsimony, we show plots for one of the 50 randomised reconstructions (always the same one throughout the paper) since they all yield virtually identical results; the same holds for the summary statistics. Regarding the statistical indicators described in Section 3.3, we compute them for each of the 50 randomised test networks and report the average value of each metric across the 50 randomised networks.

5.1.1 Ecuador

The constraints on the intermediate sales and costs are always satisfied (L_1-error $= 10^{-4}$). The reconstruction of individual weights is rather poor, as shown in Figure 3a; perfect prediction is achieved when points lie on the 45-degree line (grey dashed line). The reconstruction method tends to underpredict weights of high values and overpredict weights with intermediate or low values, although there

[8] The sector-level I-O table is available at https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Catalogo/CuentasNacionales/Anuales/Dolares/MenuMatrizInsumoProducto.htm.

is significant dispersion. (We show the histogram of the relative prediction errors and the empirical weights against their prediction errors in Figure C.2.) On average, 47% of the weights fall in the 50% confidence interval (CI).

Although the reconstruction cannot recover individual weights particularly well, the weight distribution is recovered quite well. The weight distribution has heavy tails in both the empirical and the reconstructed networks (Figure 3b; see Figure C.1a for a comparison of the weight distribution across the 50 randomised networks, the empirical test network and the full network). As expected from maximum entropy methods, the expected weight distribution is less heterogeneous than the empirical distribution. Bacilieri et al. (2023) find that the weight distribution is likely to follow a power-law with an exponent that varies between 1.0 and 1.4 depending on the country, year and estimation method. Therefore, we check whether we can recover a similar power-law exponent. We can recover it pretty well (see Figure C.1b). The power-law exponent is 1.1 for the empirical weight distribution and 1.3 for the reconstructed one.[9] Although the exponent of the reconstructed distribution is slightly higher, it still implies a divergent second moment.

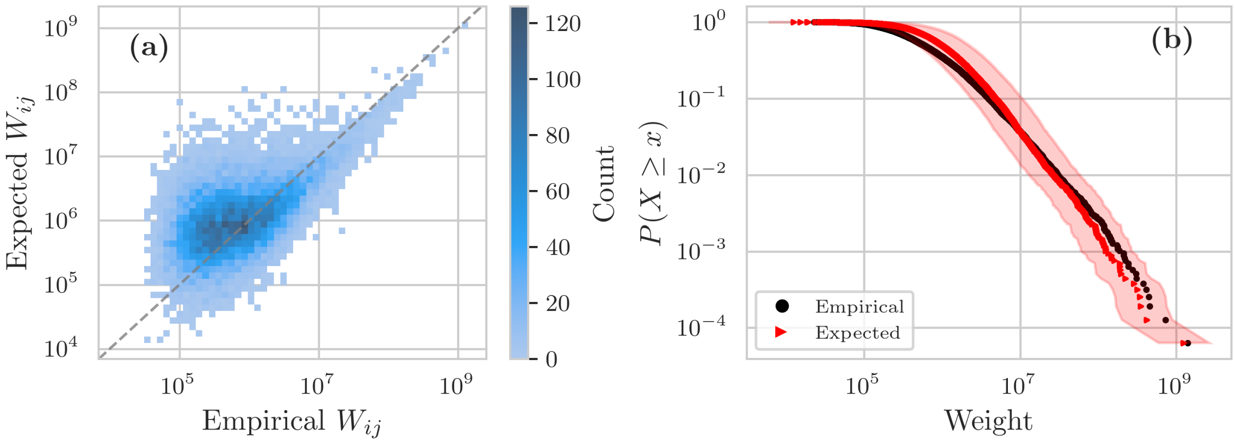

Figure 3: **(a)** 2D histogram of the empirical (x-axis) and expected (y-axis) weights for Ecuador. We divide both axes into 50 log-spaced bins and then count the number of data points falling in each square. Perfect prediction is achieved when points lie on the identity line (dashed grey line). **(b)** CCDF of the empirical (black dots) and expected (red triangles) weights for Ecuador. The shaded area represents the 50% confidence bounds. We compute the CCDF as $\bar{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \geq x)$, where $\mathbf{1}$ is the indicator function. Values are in USD.

In Table 2, we do not report the RMSE, MAE or MedAE for the weights since there is too much variation to obtain a meaningful metric; additionally, the weights likely have a diverging second moment. Therefore, we report only the cosine similarity, which is 0.93.

Type	RMSE	MAE	MedAE	Cosine similarity
Weight	–	–	–	0.928
				(0.006)
Technical	0.081	0.041	0.013	0.723
	(0.001)	(0.000)	(0.000)	(0.004)
Allocation	0.105	0.054	0.019	0.758
	(0.001)	(0.000)	(0.000)	(0.003)

Table 2: Statistical indicators for the weights, and the technical and allocation coefficients for Ecuador. As defined in the main text, RMSE denotes the root mean squared error, MAE the mean absolute error and MedAE the median absolute error. For each metric, we show its mean value across the 50 randomised reconstructions. Below the mean value, the standard deviation in parenthesis. We excluded the proxy node from the calculations.

Figure 4a and 4c show, respectively, the empirical technical and allocation coefficients on the x-

[9]To fit a power-law distribution to our data, we use the method of Clauset et al. (2009) since (1) it yields a single exponent estimate and (2) it is the most widely used estimator in the literature. See Bacilieri et al. (2023) for an in-depth discussion. Bacilieri et al. (2023) use also the estimators developed by Voitalov et al. (2019), which are based on extreme value theory. We abstain from such an analysis in this paper.

axis and their expected values on the y-axis. As for the weights, the reconstruction method does not perform particularly well in recovering either of the coefficients. Although less pronounced for the coefficients than for the weights, the method tends to overpredict coefficients of small values and underpredict coefficients with high values, again with considerable dispersion. This is further highlighted in Figure 4b and 4d, showing the empirical and reconstructed CCDF of the technical and allocation coefficients, respectively. Figure 4b and 4d further show that we can reconstruct the technical coefficients better, which also have smaller error metrics (Table 2) and for which we can recover the moments more accurately (Table 3). However, the cosine similarity is slightly higher for the allocation coefficients: 0.76 compared to 0.72.

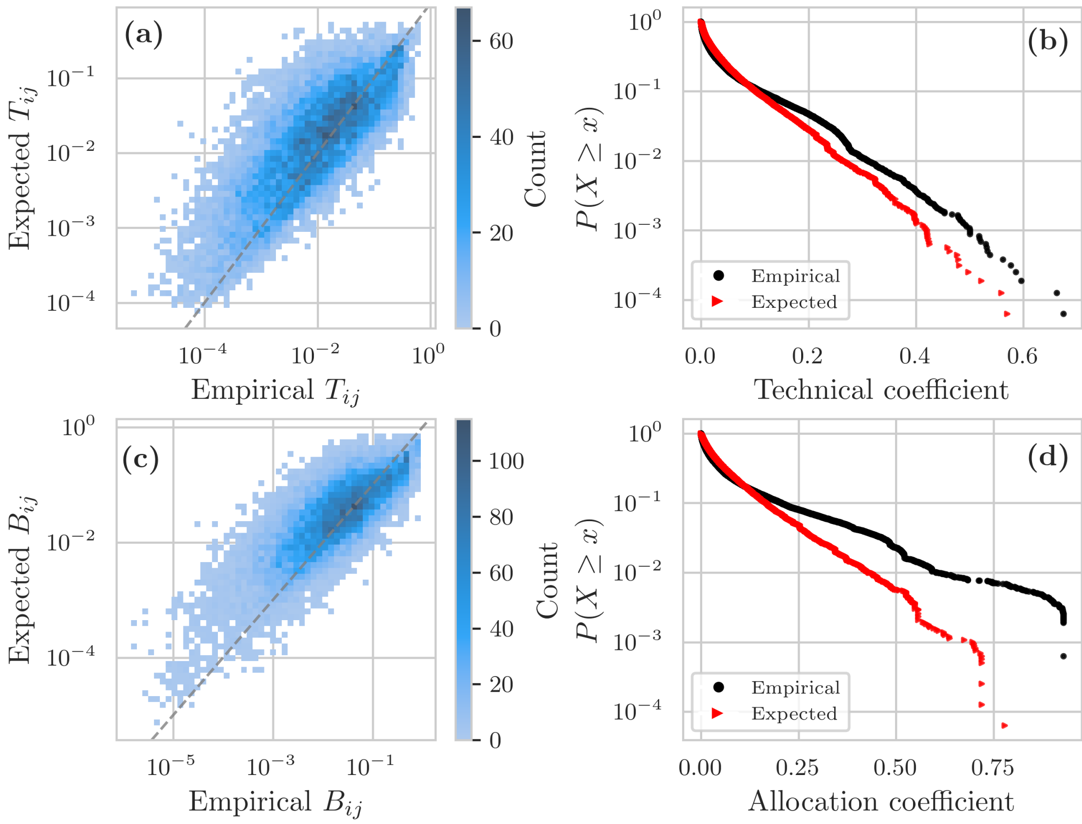

Figure 4: **(a)**, **(c)** 2D histogram of the empirical (x-axis) and the expected (y-axis) technical and allocation coefficients, respectively, for Ecuador. We use 50 log-spaced bins and then count the number of data points falling in each square. **(b)**, **(d)** CCDF of the empirical (black dots) and expected (red triangles) technical and allocation coefficients, respectively, in semi-log scale for Ecuador.

	Technical coefficients		Allocation coefficients	
	Empirical	Expected	Empirical	Expected
Mean	0.037	0.038	0.071	0.065
Median	0.010	0.016	0.019	0.031
Standard dev.	0.067	0.057	0.132	0.088

Table 3: Summary statistics of the technical and allocation coefficients for Ecuador. For each coefficient, the first column reports summary statistics for the empirical coefficients and the second column for the reconstructed ones. We show results for one of the reconstructions; all three quantities have virtually the same summary statistics across the 50 randomised empirical and reconstructed networks. We excluded the proxy node from the calculations.

5.1.2 FactSet

The constraints on the total intermediate expenditure and sales are satisfied (L_1-error $= 6 \times 10^{-4}$). Figure 5a shows the reconstructed weight distribution, which visually appears to have heavy tails. As done for Ecuador, we fit a power-law distribution. The estimated power-law exponent is lower than that recovered for Ecuador (1.0 for FactSet and 1.3 for Ecuador, see Figure C.3) and on the lower end of what is observed for other supply chain networks (Bacilieri et al., 2023). The reconstructed technical coefficients (Figure 5b) and allocation coefficients (Figure 5c) have a narrow range of variation and tend to be much smaller than those reconstructed for Ecuador. For FactSet, the technical coefficients reach a maximum value of approximately 0.008, while for the expected Ecuadorian network, they can be as high as 0.570. Similarly, the allocation coefficients are not higher than ∼0.007 in FactSet, while in Ecuador, they reach a maximum value of 0.778.

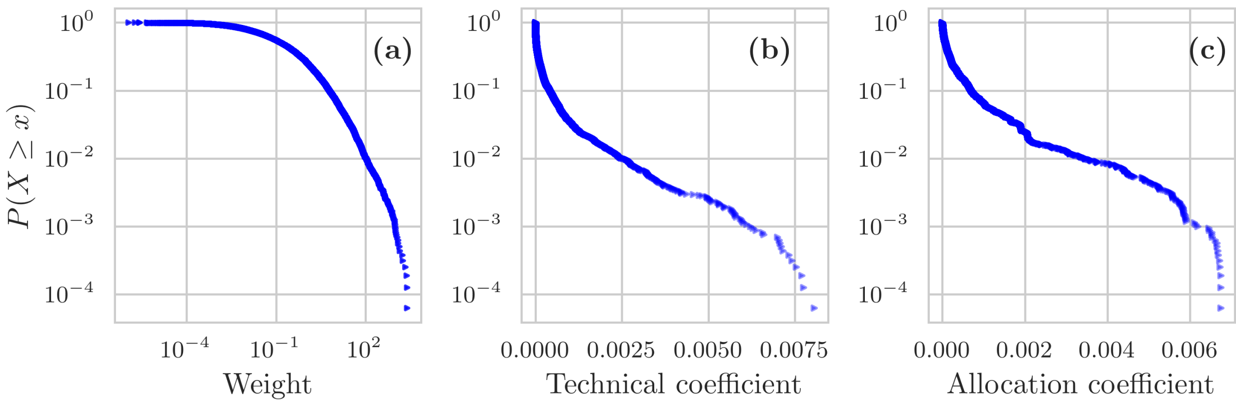

Figure 5: **(a)** CCDF of the expected weights, values in thousand USD. **(b)**, **(c)** CCDF of, respectively, the expected technical and allocation coefficients in semi-log scale for FactSet.

Given the data collection method of customer-supplier relations in FactSet, there is a bias towards observing links with customers that account for 10% or more of a firm's annual revenues. Therefore, we would expect the CCDF of the allocation coefficients to have most of the mass around, or at the very least include this 10% threshold. Instead, the maximum value is around 0.7%, well below this threshold. The average expected allocation coefficient is 3×10^{-4} while the median is 8×10^{-5} (Table 4), both of which are three to four orders of magnitude smaller than what we would have expected given the data collection method.

	Expected	
	Technical coefficient	Allocation coefficient
Mean	2×10^{-4}	3×10^{-4}
Median	3×10^{-5}	8×10^{-5}
Standard dev.	5×10^{-4}	6×10^{-4}

Table 4: Summary statistics of the technical and allocation coefficients for FactSet.

All three quantities have a narrow range of variation and are smaller than expected because total intermediate sales and expenditure of the proxy node (which represent the constraints that need to be satisfied in the weights allocation) are much bigger than those of the other firms. The proxy node accounts for 80% of intermediate expenditure and 75% of intermediate sales. To compare, in Ecuador, the proxy node accounts for 29% and 14%, respectively. Therefore, in FactSet, firms make most of their trades with the proxy node (Figure 6). The maximum transaction value for transactions between firms and the proxy node is \$407 million. But, the overall maximum is between the proxy node and itself, where it is \$48 billion, 3 orders of magnitude bigger than the maximum value among firms as well as among firms and the proxy node. Consequently, the coefficients representing trades among firms are much smaller than those representing the trades between firms and the proxy node, as shown in Figure 6.

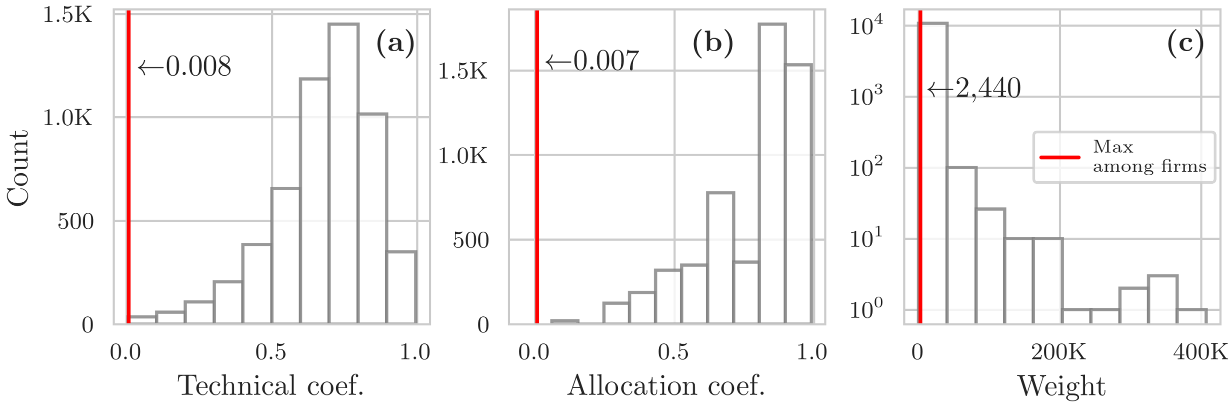

Figure 6: Trades between firms and the proxy node. **(a)** Percentage of inputs that firms buy from the proxy node (i.e, their technical coefficients with the proxy node). **(b)** Percentage of sales that firms make to the proxy node (i.e., their allocation coefficients with the proxy node). **(c)** Weights among firms and the proxy node (in both directions); values in thousand USD. We do not show quantities regarding trades of the proxy node with itself. The red line marks the maximum value of the quantity among firms only.

5.1.3 Discussion

Since the other studies that reconstruct firm-level production networks do not assess how their network reconstruction method performs on reconstructing weights, technical and allocation coefficients (Inoue and Todo, 2019; Welburn et al., 2020; Hooijmaaijers and Buiten, 2019; Reisch et al., 2021; Ialongo et al., 2022; Hillman et al., 2021), we cannot compare our results with these studies. Moreover, although several studies assess network reconstruction methods on the international trade network (ITN) or financial networks, few assess the reconstruction of the weights. Most of the studies look at higher-order network properties or dynamic indicators of systemic risk, which we discuss in Section 5.2.

Table 5 shows a comparison of our results with those of the literature. Parisi et al. (2020) test the reconstruction method on the ITN and the Electronic Market for Interbank Deposits (e-MID). They find a similar percentage of empirical weights that fall in the 50% CI. For the e-MID network, depending on the year, between 35% and 55% of the empirical weights fall in the 50% CI, while around 30% of the weights fall in the 50% CI for the ITN. They find that the empirical and expected link weights have a Pearson correlation of 0.50 for the e-MID and 0.75 for the ITN. The only other study we could find providing a comparison metric for link weights is Ramadiah et al. (2020). The authors test several reconstruction methods on Japan's bipartite bank-firm credit network. For the most disaggregated network, their reconstruction yields a cosine similarity of around 0.68 for the MaxEnt method (for bi-partite networks) and of 0.63 for the configuration fitness model with weights allocated using the IPF algorithm; these are lower than what we find for the weights but similar to our results for the technical and allocation coefficients.

While the power-law exponent of FactSet's weight distribution is in the ranges of what had been found for other supply chain networks (Bacilieri et al., 2023), the reconstructed weights and coefficients among firms are much smaller than expected. There are two main factors that deteriorate the quality of the reconstruction and act mainly through the constraints on intermediate sales and expenditure: the data cleaning procedure and the data imputations. On the one hand, the data cleaning procedure implies that we had to exclude many firms from the network (see Appendix A.2.4); this leads to (1) a higher share of the proxy node in the economy and (2) excluding many existing supply-chain relations. On the other hand, the data imputations concerning final demand and labour costs affect the values of firms' intermediate sales and expenditures. Because firms report the cost of goods sold, which often includes labour costs, we do not always know intermediate costs exactly. Similarly, firms disclose their revenues (intermediate sales plus sales to final demand), so for all firms, we do not know their intermediate sales exactly. As shown in Appendix A.2.2, labour costs tend to be overestimated. Although we cannot test whether final demand is over or underestimated, we think we are very likely overestimating it for the majority of firms. The main reason for overestimating final demand is the use of national I-O tables at the sector level, which have a very different treatment of the wholesale and retail sectors compared to firm-level data. National accounts treat wholesale and retail

Data set	Method	Quantity	Pct in 50% CI	Measure	Score	Source
Ecuador SC	CReM	Weight	47%	Cosine	0.93	This paper
Ecuador SC	CReM	Technical coefficient		Cosine	0.72	This paper
Ecuador SC	CReM	Allocation coefficient		Cosine	0.76	This paper
ITN	CReM	Weight	30%	Pearson	0.75	Parisi et al. (2020)
e-MID	CReM	Weight	35-55%	Pearson	0.50	Parisi et al. (2020)
Japan's bank-firm credit	MaxEnt	Weight		Cosine	0.68	Ramadiah et al. (2020)
Japan's bank-firm credit	BFiCM + IPF	Weight		Cosine	0.63	Ramadiah et al. (2020)

Table 5: Microscale quantities: comparison of our results with the literature. In column "Data set", "Ecuador SC" stands for Ecuador supply chain network, "ITN" for international trade network and e-MID for electronic market for interbank deposits. In column "Method", CReM stands for the conditional reconstruction method that we employ in this paper (Parisi et al., 2020). Besides the MaxEnt method, Ramadiah et al. (2020) use a two-step procedure. In the first step, the binary topology is reconstructed using a bipartite fitness-induced configuration model (BFiCM) (Squartini et al., 2017), which extends the fitness-induced configuration model of Cimini et al. (2015b) to the bipartite case. In the second step, weights are allocated using the iterative proportional fitting algorithm. For Ramadiah et al. (2020), we show results for the most disaggregated network and only for the two methods with the highest cosine similarity.

as "pass-through" sectors, accounting only for the trade margin they make and re-distribute the rest of their output among the other industries.[10] The final demand of these other industries thus becomes (fictitiously) higher. In firm-level data quite the opposite happens, with many firms selling to retail and wholesale firms that then sell to final demand. We refer to Appendix A.4 and Bacilieri et al. (2023) for a longer discussion on differences between national accounts and firm-level data.

The data cleaning and data imputation problems just discussed imply that the constraints on intermediate sales and expenditure of the proxy sector are much bigger than those of other firms. Therefore, to satisfy the constraints posed in the maximum entropy procedure, the weights allocated to the proxy node need to be much bigger than those among firms, meaning that firms buy most of their inputs and sell most of their output to the proxy node. This deteriorates the reconstruction of the weights, and of the technical and allocation coefficients among firms.

Lastly, the findings for both Ecuador and FactSet suggest that additional mechanisms, not captured by the constraints we pose, may be at work that are essential to the formation of network weights. Further unravelling what these mechanisms are could improve the reconstruction.

5.2 Higher-order and macroscale properties

First, this section discusses the results for the output multipliers and the influence vector; we start with Ecuador and then FactSet. Second, we show the results for aggregate volatility. We do not assess aggregate volatility for FactSet as that would entail estimating TFP using an econometric technique such as that outlined in Magerman et al. (2016), which is outside of the scope of this paper. We conclude this section with a discussion.

5.2.1 Multipliers

Ecuador. The reconstruction method performs well at reproducing the empirical output multipliers but not that well at reconstructing the influence vector. Figure 7a and 7b show the empirical (x-axis) and the expected (y-axis) output multipliers and influence vector, respectively. For the output multipliers, points cluster fairly tightly around the identity line (dashed grey line), while for the influence vector most points are located at the bottom left corner, suggesting that the influence is consistently overestimated. However, there are a few exceptions, which tend to be firms with higher influence (Figure C.7). These findings are further confirmed in Figure 7c and 7d, showing the CCDF

[10] If wholesale is trading service goods, then all of its output is distributed to other products.

of the (empirical and reconstructed) output multipliers and influence vector, respectively. It also highlights that the minimum of the empirical influence vector is around one order of magnitude smaller than the reconstructed one. The cosine similarity of the output multipliers is higher than that of the influence vector (0.99 and 0.56, respectively); however, the other error metrics are less clear cut (Table 7). The reconstruction method can recover the first two moments and the median of the output multipliers (Table 6). For the influence vector, we can recover only the standard deviation.

In Figure 7a clusters tend to form. Clusters arise because, for the output multipliers, we simulated firms' value-added using sector-level data. The interaction of firms' intermediate expenditures, sectoral ν_s's and the binary topology produces those clusters; see Appendix C.3.3 for more details.

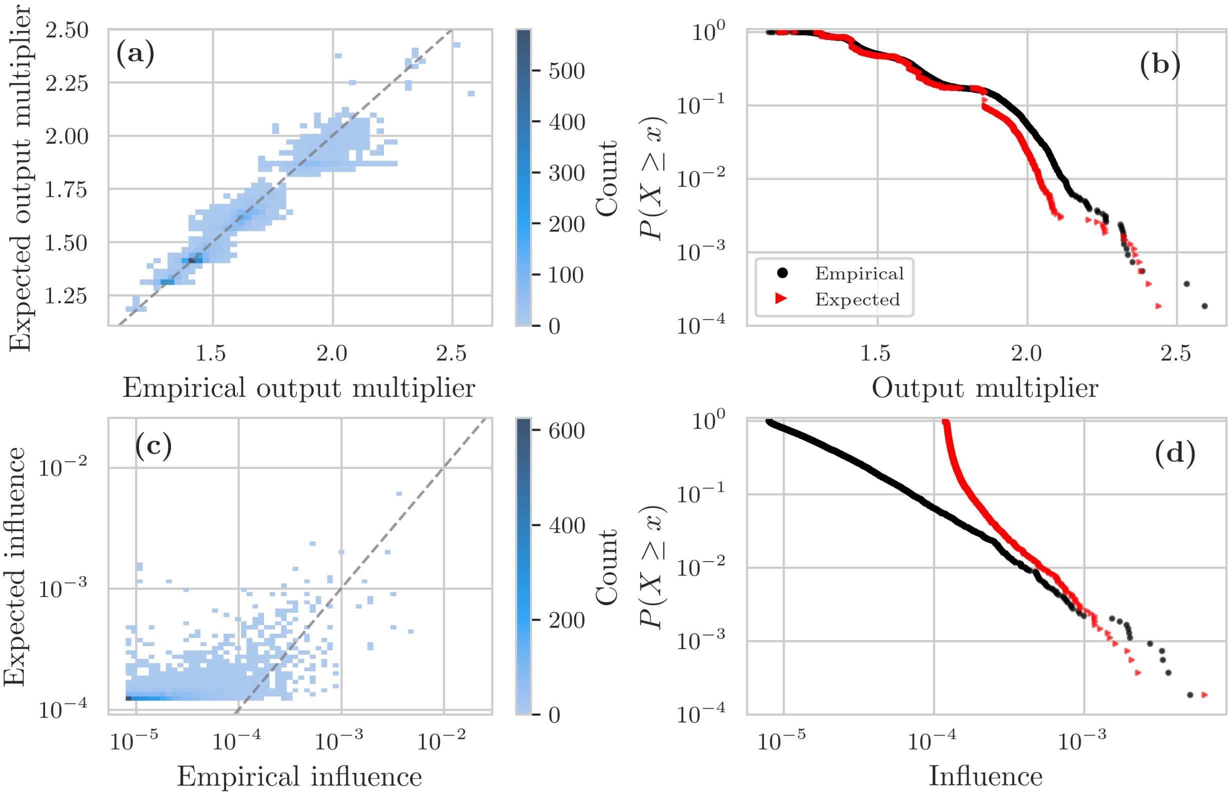

Figure 7: **(a)**, **(c)** 2D histogram of the empirical (x-axis) and reconstructed (y-axis) output multipliers and influence vector, respectively, for Ecuador. Perfect prediction is achieved when points lie on the 45-degree line (dashed grey line). We use 50 log-spaced bins for both axes and count the number of points falling in each square. **(b)**, **(d)** CCDF of the output multipliers and influence vector, respectively, for Ecuador. Black dots refer to the empirical CCDF and red triangles to the CCDF of the expected network. For calculating the empirical multipliers, we use the full network and consider the multipliers of firms in our test network only.

	Output multiplier		Influence vector	
	Empirical	Expected	Empirical	Expected
Mean	1.423	1.553	4×10^{-5}	1×10^{-4}
Median	1.420	1.475	2×10^{-5}	1×10^{-4}
Standard dev.	0.247	0.202	1×10^{-4}	1×10^{-4}

Table 6: Summary of the output multipliers and influence vector for Ecuador. For each multiplier, the first column reports summary statistics for the multipliers calculated using the empirical network while the second column for the multipliers calculated using the reconstructed network. We show results for one of the reconstructions; all three quantities have virtually the same summary statistics across the 50 randomised empirical and reconstructed networks. We excluded the proxy node from the calculations of the expected multipliers. For calculating the empirical multipliers, we use the full network and consider the multipliers of firms in our test network only.

Type	RMSE	MAE	MedAE	Cosine similarity
Output multipliers	0.036	0.024	0.015	0.999
	(3×10^{-4})	(1×10^{-4})	(2×10^{-4})	(8×10^{-6})
Influence vector	0.033	0.024	0.022	0.560
	(2×10^{-4})	(2×10^{-5})	(1×10^{-5})	(3×10^{-3})

Table 7: Statistical indicators for the output multipliers and influence vector for Ecuador As defined in the main text, RMSE denotes the root mean squared error, MAE the mean absolute error and MedAE the median absolute error. For each multiplier, we show its mean value across the 50 randomised reconstructions. Below the mean value, the standard deviation in parenthesis. We excluded the proxy node from the calculations of the expected multipliers. For calculating the empirical multipliers, we use the full network and consider the multipliers of firms in our test network only.

FactSet. Figure 8a and 8b show the CCDF of, respectively, the expected output multipliers and the expected influence vector for FactSet. While the CCDF of the influence vector displays heavy tails, that of the output multipliers does not. As done for the weight distribution, we compare with empirical findings of other networks where the influence vector is found to have heavy tails and likely follows a power-law with a divergent second moment (Bacilieri et al., 2023). The estimated power-law exponent is equal to 1.9, higher than what found for Belgium (1.12, Magerman et al., 2016), Hungary and Ecuador (around 1.3-1.5 and 1.2-1.4, respectively, Bacilieri et al., 2023). We show the CCDF and its power-law fit in Appendix C.3.1.[11] The output multipliers have a much higher median, first and second moment in FactSet (Table 8) compared to Ecuador (Table 6), while the influence vector has lower moments.

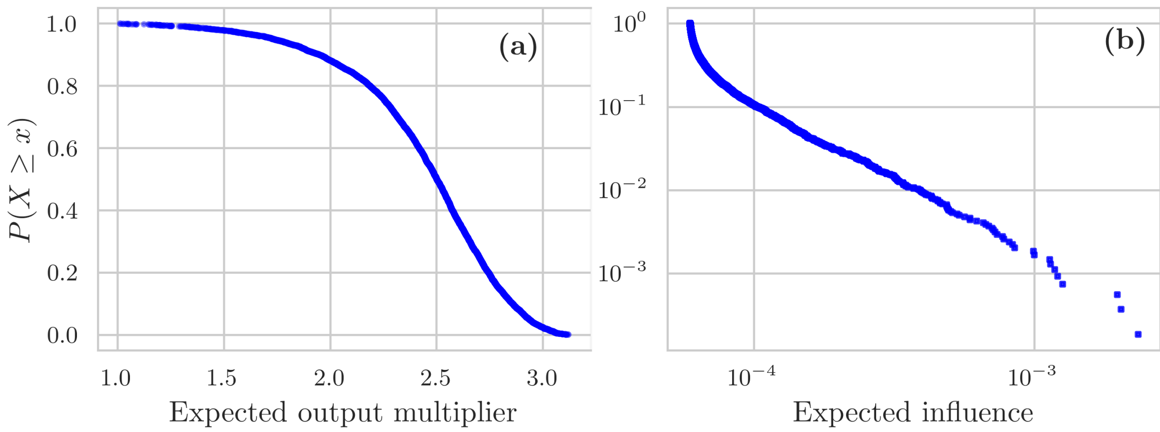

Figure 8: CCDF of the expected **(a)** output multipliers and **(b)** influence vector for FactSet. The multipliers of the proxy node are excluded.

	Expected	
	Output multiplier	Influence vector
Mean	2.443	8×10^{-5}
Median	2.501	6×10^{-5}
Standard dev.	0.364	8×10^{-5}

Table 8: Summary statistics of the output multipliers and the influence vector for FactSet. The multipliers of the proxy node are excluded.

[11] We also fit a power-law distribution to Ecuador's influence vector and recover an exponent of 2.0. However, the fit was rather poor. We show results for Ecuador in Appendix C.3.1.

5.2.2 Supply-side shocks and aggregate volatility

We measure how firm-level TFP shocks affect macroeconomic output by propagating through the supply chain network using Equation 12. We simulate TFP shocks as explained in Section 3.3.3. We then use the empirical influence vector to calculate the empirical volatility and the reconstructed influence vector to calculate the predicted volatility. The volatility predicted by our reconstructed network is much higher than the empirical one. Using the empirical network, GDP volatility is 6.4%, while the mean across the 50 reconstructions is 102.3% with a standard deviation of 0.002 (Table 9).

To understand why we predict such a high GDP volatility, we do a variance decomposition analysis and look at the role of the proxy node in explaining aggregate volatility:

$$\sigma^2_{\Delta y} = \sum_{i \neq \kappa} \mathrm{Var}(\Delta\epsilon_i) v_i^2 + \mathrm{Var}(\Delta\epsilon_\kappa) v_\kappa^2 \,,$$

where the first term on the RHS is the contribution to the variance of GDP growth of the firms in the test network and the second term is the contribution of the proxy node, indexed by κ. We further look at shares

$$\Lambda_S = \frac{\sum_{i \in S} \mathrm{Var}(\Delta\epsilon_i) v_i^2}{\mathrm{Var}(\Delta\epsilon_\kappa) v_\kappa^2 + \sum_{j \neq \kappa} \mathrm{Var}(\Delta\epsilon_j) v_j^2} \,.$$

where S is the set of observations for which we are computing the share of the variance. We use the variance to ensure that shares sum up to 1. The proxy node contributes to 99.3% of the variance, while all the other firms to 0.7%. If we calculate GDP volatility without including the proxy node, the predicted GDP volatility drops to 8.8%.

		Reconstructed		Benchmark	
	Empirical	Proxy node	No proxy node	Proxy node	No proxy node
GDP volatility	6.4%	102.3% (0.0020)	8.8% (0.0010)	94% (0.0030)	7.0% (0.0002)

Table 9: Predicted aggregate GDP volatility using the empirical, the reconstructed and the benchmark influence vector (Equation 12). For the reconstructed and benchmark, "Proxy node" reports the predicted aggregate volatility calculated including the proxy node, while "No proxy node" excludes the proxy node from the calculations. Values for the reconstructed and the benchmark show the average taken over each of the 50 randomised networks. Below the mean value, we show the standard deviation in parenthesis.

Benchmark. To benchmark our results, we first calculate the influence vector assuming that each firm buys the same proportion of inputs from its suppliers (i.e., we modify $\mathbf{\Omega}$ in Equation 11) and then compute aggregate volatility using Equation 12. Assigning homogeneous input shares for each firm still satisfies the constraints on the intermediate costs, but the constraints on intermediate sales are not guaranteed to be satisfied. Our benchmark yields a volatility of 94% when the proxy node is included and 7.0% when excluded (Table 9). The volatility predicted by the benchmark is 0.6 percentage points higher than the empirical volatility, while the volatility of the reconstruction is 2.4 percentage points higher than the empirical volatility.

5.2.3 Discussion

As noticed in Section 5.1.3, we cannot compare with previous results of other studies reconstructing firm-level production networks. For financial networks and the ITN, different reconstruction methods perform reasonably well in reconstructing higher-order network properties such as the weighted clustering coefficient (e.g., Mastrandrea et al., 2014; Cimini et al., 2015b,a; Parisi et al., 2020). Table 10 reports some of the findings in the literature. For financial networks, Ramadiah et al. (2020) find that all the reconstruction methods they employ underestimate the level of systemic risk, except for a small region of the parameter space. Anand et al. (2015) report similar findings for MaxEnt, but find that the minimum density method overestimates systemic risk. Di Gangi et al. (2018) find that the cross-entropy capital asset pricing model can reproduce very well systemic risk while the other ensemble methods overestimate or underestimate it depending on the shock scenario. Differently, individual banks' systemic risk and indirect vulnerability, which could be seen as akin to the multipliers we test, are consistently underestimated across all the methods they assess. The degree-corrected gravity model can reproduce DebtRank (also akin to the multipliers), with a Person correlation equal to 1 (Cimini et al., 2015b).

Dataset	Method	Finding	Source
Ecuador SC	CReM	Overestimates influence vector, cosine sim. = 0.56	This paper
Ecuador SC	CReM	Under/Overestimates output multipliers, cosine sim. = 1.	This paper
e-Mid	DcGM	DebtRank: correlation = 1	Cimini et al. (2015b)
ITN	DcGM	DebtRank: corrrelation = 1	Cimini et al. (2015b)
US bank-asset	CE CAPM, Max. entropy CAPM, BWCM, BECM	Underestimate banks' measures of systemic risk	Di Gangi et al. (2018)
Ecuador SC	CReM	Overestimates aggregate volatility	This paper
German banks	MaxEnt	Underestimates systemic risk	Anand et al. (2015)
German banks	Minimum density	Overestimates systemic risk	Anand et al. (2015)
US bank-asset	Cross-entropy CAP model	Reproduces well aggregate vulnerability	Di Gangi et al. (2018)
US bank-asset	Max. entropy CAP model, BWCM, BECM	Over/underestimate aggregate vulnerability depending on the shock scenario	Di Gangi et al. (2018)
Japan bank-firm credit	MaxEnt, Min. density, CF + IPF, BFiCM + IPF	Underestimate the average probability of default	Ramadiah et al. (2020)

Table 10: Higher-order and macroscale quantities: comparison of our results with the literature. In column "Dataset", "Ecuador SC" stands for Ecuador supply chain network, "ITN" for international trade network and "e-MID" for electronic market for interbank deposits. In column "Method", CReM stands for the conditional reconstruction method that we use in this paper (Parisi et al., 2020). "DcGM" stands for degree-corrected gravity model (Cimini et al., 2015b), "CE CAPM" for cross-entropy capital asset pricing model (it is a deterministic method), "Max. entropy CAPM" for maximum entropy capital asset pricing model, "BWCM" for bipartite weighted configuration model and "BECM" for bipartite enhanced configuration model (all of the last three methods are ensemble methods). "Min. density" is the minimum density method developed by Anand et al. (2015). "CF" is the bipartite maximum entropy configuration model and "BFiCM" is the bipartite fitness-induced configuration model (Squartini et al., 2017); in both cases, weights are allocated in a second step using the iterative proportional fitting algorithm.

While the vast majority of the studies find that systemic risk is underestimated, for Ecuador, we find that the conditional maximum entropy reconstruction method we employ overestimates aggregate volatility, similar to what Anand et al. (2018) find for the minimum density method. First, while including the proxy node in the calculations of aggregate volatility is supposed to capture the volatility of the firms that were excluded from the test network, there is a fundamental difference between the role the proxy node has in the network and that of the firms that we excluded. The proxy node is connected to all the other firms in the test network. In contrast, the firms excluded from the test network are less well connected in the empirical network. Additionally, the proxy node has the biggest influence, which equals 0.19. Therefore, when the proxy node is shocked, it immediately passes 19% of the shock to all the other firms in the network. To compare, the firm with the biggest influence in the empirical network has an influence of 0.005, with the biggest influence among the firms excluded from the test network being much lower and equal to 0.0002. Noting also that the firms excluded from the test network contribute to a mere 7% of aggregate volatility while the firms included in the test network to 93%, it makes sense to exclude the proxy node from the calculation of aggregate volatility as it is not a good aggregate representation of the rest of the economy, at least in this context. In fact, while the inclusion of the proxy node degrades the prediction of aggregate volatility, it enhances the reconstruction of the multipliers; see Appendix C.3.4.

Second, aggregate volatility is still overestimated even when the proxy node is excluded because the influence vector tends to be overestimated. Notice that the benchmark can predict aggregate volatility more accurately since it tends to overestimate influences lightly less than the reconstruction (see Appendix C.3.5). The reconstructed influence vector is overestimated for a combination of two factors, one of which is related to the binary topology and the other to the weighted topology. We can see how the binary topology affects the influence vector by looking at the out-degrees, which are highly affected by the selection of firms and the link deletion. In creating the test network, firms lose more customers than suppliers (see Appendix C.4.1). Already at the first step (selecting the firms to keep in our test network), the maximum number of customers (out-degree) decreases almost 7 folds, while the maximum number of suppliers only 3 folds. Since the influence vector calculates the weighted sum of the number of walks from firm i to firm j (so following the outgoing edges starting at i) for walks of various lengths and the out-degrees have been considerably truncated, it cannot capture walks of longer lengths. Instead, the output multipliers, relying on the incoming edges and thus the in-degrees, are that not affected. The weighted topology comes into play because the reconstruction method tends to overestimate weighted quantities for intermediate and small values, which are numerous. Although higher (and other) weights can be underestimated, these are not abundant enough and the magnitude of the underestimation is not big enough compared to the quantity and the size of the overestimation of low and intermediate weights.

For FactSet, a significant result is that the estimated power-law exponent of the distribution of the influence vector implies a divergent second moment. Acemoglu et al. (2012) show, at a theoretical level, that the influence vector affects aggregate volatility, so it is important that we can recover an exponent more or less in line with empirical observations.[12]

5.3 Different numbers of unknown links

We now discuss how the reconstruction method performs when the number of unknown links in the test network changes; we keep the number of firms constant. We delete 0%, 10%, 20% and so on up to 90% of the links; we also show results for the reconstruction matching the mean degree of FactSet, which has 96% of unknown links. For each percentage of unknown links, we simulate 50 randomised networks and investigate the number of weights that fall into the 50% CI, and the median absolute error and the cosine similarity for microscale and higher-order quantities. We also assess aggregate volatility.

Microscale and higher-order quantities. Figure 9 shows the different metrics used for assessing the quality of the reconstruction of the weights, and the technical and allocation coefficients, and the multipliers. For the weights, we show the percentage of weights that fall in the 50% CI (Figure 9a) and the cosine similarity (Figure 9b). For all the other quantities, we show the median absolute error (left column) and the cosine similarity (right column). The overall trend is that as the number of unknown links increases, so do the metrics.

As the number of unknown links increases, the cosine similarity increases, with the two multipliers being the only exceptions. Such a counter-intuitive finding arises from the link deletion mechanism. We are deleting links with a smaller weight with a higher probability, hence as the number of unknown links increases, the empirical weights associated with the known links are less heterogeneous. Since maximum entropy methods allocate weights as uniformly as possible given the constraints, the higher the number of unknowns, the better it can reconstruct the weights associated with the known links. Consequently, also more weights fall into the 50% confidence interval.

For the allocation coefficients, the cosine similarity is considerably lower than that of all the other quantities and it jumps from 0.05 to 0.76 when the number of unknown links increases from 90% to 96%. This jump is associated with a considerable decrease in the number of customers firms have when the number of unknown links increases from 90% to 96% (see Appendix C.4.1). Therefore, the allocation coefficients, which gauge the percentage of total output firm i sells to firm j, are highly affected by a drastic decrease in the number of customers (see Appendix C.4.2). However, we would have expected this to have a negative effect. Instead, it enhances the predictions. On the contrary, the technical coefficients, which measure the percentage of inputs that j buys from i, do not experience

[12] Acemoglu et al.'s (2012) theoretical result is exemplified in Equation 12, which shows that aggregated volatility scales with the Euclidean norm of the influence vector. Luca's diversification argument implies that aggregate volatility decays as $N^{1/2}$. However, if the CCDF of the influence vector is Pareto distributed with parameter $\gamma \in (1, 2)$, the diversification argument no longer holds and aggregate volatility decays much more slowly, as $N^{1/(\gamma-1)}$ (Carvalho and Tahbaz-Salehi, 2019).

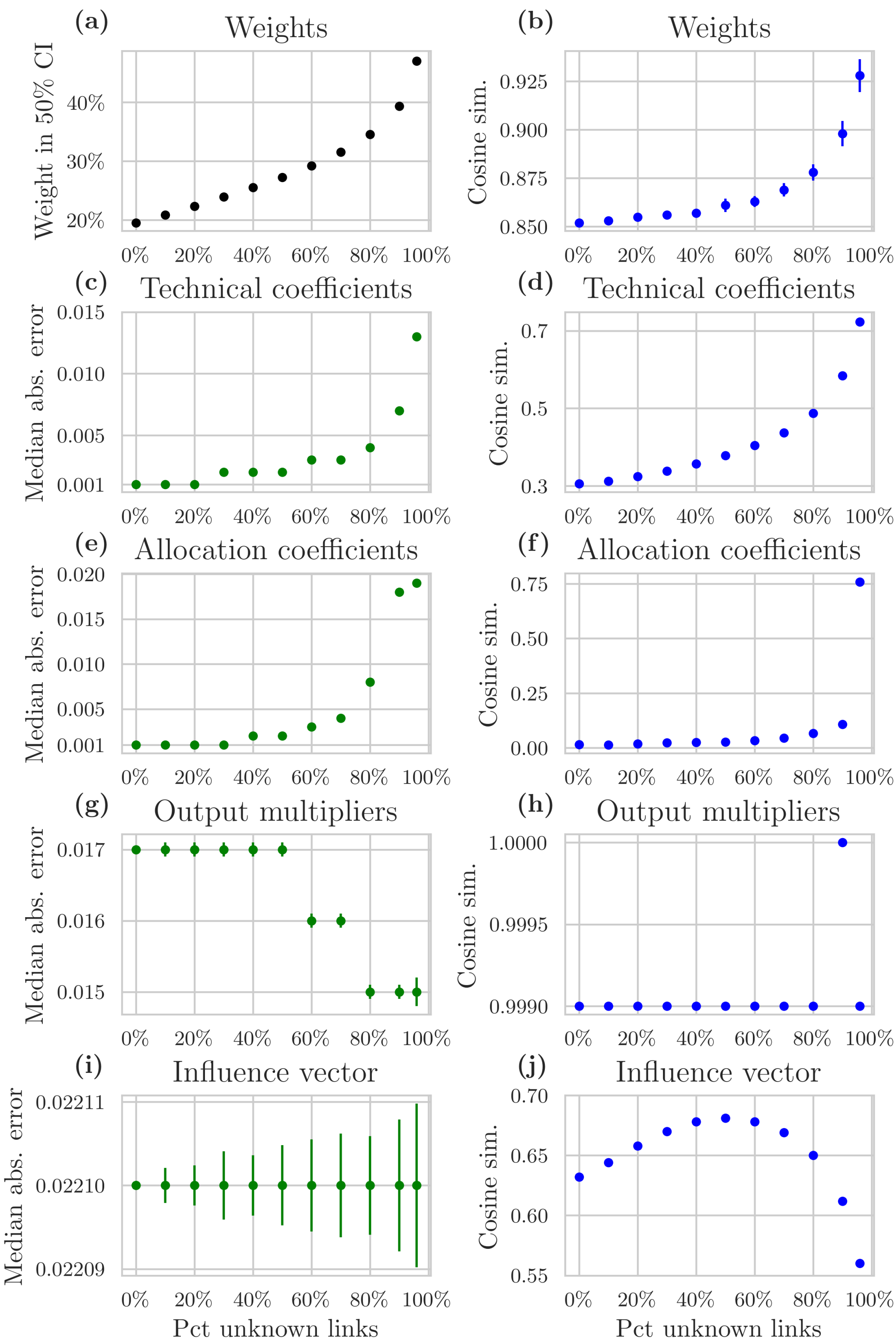

Figure 9: Error metrics and similarity measures as the number of unknown links increases. **(a)** Percentage of empirical weights in the 50% confidence interval. **(b)**, **(d)**, **(f)**, **(h)**, **(j)** cosine similarity for, respectively, weights, technical and allocation coefficients, output multipliers and influence vector. **(c)**, **(e)**, **(g)**, **(i)**, Mean absolute error for, respectively, technical and allocation coefficients, output multipliers and influence vector. Bars show the standard deviation across the 50 randomised networks.

such a jump because the same drastic change does not happen for the number of suppliers. Moreover, firms tend to have more customers than suppliers (see Appendix C.4.1, and Bacilieri et al., 2023, for an extension to production networks of other countries and over time), making it easier to guess the weights correctly from the supply side and thus reconstruct the technical coefficients better than the allocation coefficients throughout.

The output multipliers always have the same cosine similarity. Since the output multipliers are derived from the technical coefficient matrix and depend on the incoming edges, hence the number of suppliers firms have, they are not that affected by the number of unknown links. Instead, the influence vector relies on the number of customers firms have (out-degree), which we saw being more affected by the deletion of firms and links, so the influence vector has a much lower cosine similarity than the output multipliers. It is not so clear why the cosine similarity has an inverted U-shaped curve, with the maximum value of 0.68 reached when 50% of the links are unknown. It is however the case that the cosine similarity increases from 0.63 when all links are known to 0.68 (50% unknown links) and then decreases again until it drops to 0.56 when 96% of the links are unknown. It might be that these are small fluctuations of no particular value and that the reconstruction of the influence vector starts deteriorating when more than 50% of the links are unknown because the out-degrees start being too affected by the deletion of the links.

The median absolute error slightly increases for the technical and allocation coefficients because higher errors (in absolute value) tend to occur for higher weights. The median absolute error is constant for the multipliers.

Aggregate volatility. Figure 10 shows the predicted aggregated volatility (calculated excluding the proxy node) as the number of unknown links increases for the reconstruction (red dots), the benchmark (green squares) and the empirical volatility (black dashed line). In line with the results for the case of 96% of unknown links, the reconstruction always predicts a much higher volatility, as does the benchmark. Although the benchmark's predictions are closer to the empirical volatility. One would think that the higher the number of known links, the better we get at predicting aggregate volatility. However, this is not the case. In fact, the fewer links we know, the better we can predict GDP volatility. This is because the more links we know, the more we overestimate the influences (especially bigger influences, see Appendix C.4.2).

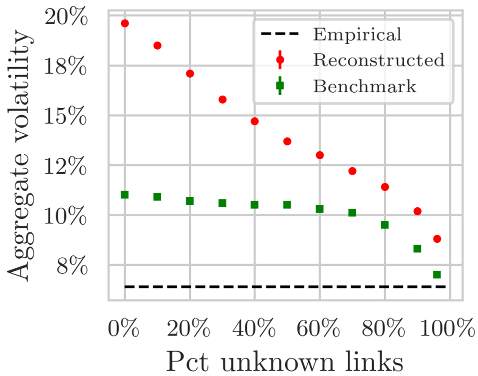

Figure 10: Aggregate volatility as the number of unknown links increases using the reconstructed network (red dots) and the benchmark (green squares). The black dashed line marks the empirical volatility. Bars show the standard deviation across the 50 randomised networks. We show the volatility calculated excluding the proxy node.

6 Conclusions

There is widespread interest in modelling the global economy from the bottom up. There is also widespread agreement that supplier-customer relations are an essential feature of such modelling efforts. However, data are scarce, have missing information and are not easily accessible. In this paper, we have made a couple of first steps in bringing this agenda forward.

First, we provided the first rigorous assessment of a network reconstruction method (Parisi et al., 2020) on the administrative dataset of Ecuador. We focused on reconstructing the weighted network given the binary topology when many links and firms are missing. An interesting finding is that the quality of the reconstruction of different quantities seems to depend on network features that are particularly sensitive to the sampling strategy of firms and links, something that future research should explore further. Second, we assessed whether a global dataset of listed firms, where many links and firms are missing, can be enhanced by merging it with sector-level data. We then used this "augmented" dataset for inferring the link weights using a conditional maximum entropy method (Parisi et al., 2020). Our results show that further work needs to be done, especially in reconciling firms' financial accounts with national accounts, which is essential for better reconstructing the weighted production network, in particular when many firms are missing.

In our study, we assumed to know the binary topology (although partially) to cover a use case that could help reconstruct commercial datasets. A natural next step would be to predict links and then reconstruct weights, possibly with different degrees of knowledge of the production network. It is of particular value to understand the performance of reconstruction methods when one does not know the binary topology, as it could unlock the study of economies for which no such data is available. The reconstruction method we employed can easily accommodate the prediction of the binary topology, either partially or in its entirety.

Our assessment of macroscale quantities was rather limited and restricted to a standard general equilibrium I-O model (Acemoglu et al., 2012). Therefore, there is much research to be done on different models (and scenarios), especially agent-based models, where we believe that the accuracy of the reconstruction of microscale quantities matters more for the model's outcomes than for general equilibrium models.

While many reconstruction methods are available, the high number of nodes and links in firm-level networks renders almost all of them infeasible. Future research is thus necessary to develop different reconstruction methods suited for large-scale firm-level networks. It is also important to conduct similar analyses on other firm-level datasets for which the ground truth network is available, as ours was only one of the first initial steps.

Aknowledgments

We thank François Lafond, Jose Moran, R. Maria del Rio-Chanona, Anton Pichler, Luca Mugo and Fabian Dablander for useful comments. We thank Doyne J. Farmer, Jangho Yang, Giancarlo Antonucci and the Complexity group at the Institute for New Economic Thinking at the Oxford Martin School for inspiring discussions. We have benefited from many comments from audiences at Cambridge IfM and CSH Vienna. We also thank the Servicio de Rentas Internas (SRI) and its Centro de Estudios Fiscales that provided the raw data of Ecuador for research purposes. This work was supported by the Oxford Martin Programme on the Post-Carbon Transition, the Institute for New Economic Thinking at the Oxford Martin School and Baillie Gifford. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract no. 2019-1902010003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorised to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. FactSet had the opportunity to review the paper.

Author contributions

A. Bacilieri suggested the research topic, designed the work, implemented the reconstruction method and associated tests, analysed the data and wrote the paper. Pablo Astudillo-Estevez provided the cleaned Ecuador dataset and relevant information about it. They also declare that there were no conflicts of interest.

Research data

We do not have permission to share the two firm-level datasets. For FactSet because it is a commercial dataset, while for Ecuador because it is a confidential, administrative dataset with restricted access.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016.

Anand, K., Craig, B., and von Peter, G. (2015). Filling in the blanks: Interbank linkages and systemic risk. *Quantitative Finance*, 15:625–636.

Anand, K., van Lelyveld, I., Banai, Á., Friedrich, S., Garratt, R., Hałaj, G., Fique, J., Hansen, I., Jaramillo, S. M., Lee, H., et al. (2018). The missing links: A global study on uncovering financial network structures from partial data. *Journal of Financial Stability*, 35:107–119.

Astudillo-Estevez, P. A. (2021). *Towards a Post-Oil Economy: A Complexity Approach to Understanding Natural Resource Dependency and Economic Diversification in Ecuador*. PhD thesis, University of Oxford.

Atalay, E., Hortacsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202.

Bacilieri, A., Borsos, A., Astudillo-Estevez, P., and Lafond, F. (2023). Firm-level production networks: what do we (really) know? *Production networks and planetary boundaries: challenges and opportunities for Integrated Assessment Models* (Annex A). DPhil thesis, University of Oxford.

Baral, P. and Fique, J. (2012). Estimation of bilateral connections in a network: Copula vs. maximum entropy. In CIRANO https://cirano.qc.ca/conferences/public/pdf/networks2012/02-BARAL-FIQUE-Estimation_of_Bilateral_Exposures-A_Copula_Approach.pdf.

Barrot, J.-N. and Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., and Moxnes, A. (2019). The origins of firm heterogeneity: A production network approach. Technical report, National Bureau of Economic Research.

Boehm, J. and Sonntag, J. (2022). Vertical integration and foreclosure: Evidence from production network data. *Management Science*.

Brintrup, A., Wichmann, P., Woodall, P., McFarlane, D., Nicks, E., and Krechel, W. (2018). Predicting hidden links in supply networks. *Complexity*, 2018:1–12.

Carvalho, V. M. (2014). From micro to macro via production networks. *Journal of Economic Perspectives*, 28(4):23–48.

Carvalho, V. M., Nirei, M., Saito, Y. U., and Tahbaz-Salehi, A. (2021). Supply chain disruptions: Evidence from the great east japan earthquake. *The Quarterly Journal of Economics*, 136(2):1255–1321.

Carvalho, V. M. and Tahbaz-Salehi, A. (2019). Production networks: A primer. *Annual Review of Economics*, 11:635–663.

Cimini, G., Mastrandrea, R., and Squartini, T. (2021). *Reconstructing networks*. Cambridge University Press.

Cimini, G., Squartini, T., Gabrielli, A., and Garlaschelli, D. (2015a). Estimating topological properties of weighted networks from limited information. *Physical Review E*, 92(4):040802.

Cimini, G., Squartini, T., Garlaschelli, D., and Gabrielli, A. (2015b). Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports*, 5:15758.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

Di Gangi, D., Lillo, F., and Pirino, D. (2018). Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. *Journal of Economic dynamics and control*, 94:117–141.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2022). Quantifying firm-level economic systemic risk from nation-wide supply networks. *Scientific Reports*, 12(1):1–13.

Diem, C., Borsos, A., Reisch, T., Kertész, J., and Thurner, S. (2023). Estimating the loss of economic predictability from aggregating firm-level production networks. arXiv preprint arXiv:2302.11451.

Fally, T. (2012). Production staging: measurement and facts. University of Colorado Boulder, Boulder, Colorado.

Gandy, A. and Veraart, L. A. (2017). A Bayesian methodology for systemic risk assessment in financial networks. *Management Science*, 63(12):4428–4446.

Golan, A., Judge, G., and Robinson, S. (1994). Recovering information from incomplete or partial multisectoral economic data. *The Review of Economics and Statistics*, pages 541–549.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2020). Firm volatility in granular networks. *Journal of Political Economy*, 128(11):4097–4162.

Hillman, R., Barnes, S., Wharf, G., and MacDonald, D. (2021). A new firm-level model of corporate sector interactions and fragility: the Corporate Agent-Based (CAB) Model. Economics Department Working Papers No. 1675, OECD. Available at: https://www.oecd-ilibrary.org/economics/a-new-firm-level-model-of-corporate-sector-interactions-and-fragility-the-corporate-agent-based-cab-model_e9de0097-en.

Hooijmaaijers, S. and Buiten, G. (2019). A methodology for estimating the dutch interfirm trade network, including a breakdown by commodity. Technical report, Statistics Netherlands.

Ialongo, L. N., de Valk, C., Marchese, E., Jansen, F., Zmarrou, H., Squartini, T., and Garlaschelli, D. (2022). Reconstructing firm-level interactions in the dutch input–output network from production constraints. *Scientific reports*, 12(1):1–12.

Inoue, H. and Todo, Y. (2019). Firm-level propagation of shocks through supply-chain networks. *Nature Sustainability*, 2(9):841–847.

Kosasih, E. E. and Brintrup, A. (2022). A machine learning approach for predicting hidden links in supply chain with graph neural networks. *International Journal of Production Research*, 60(17):5380–5393.

Lahr, M. and De Mesnard, L. (2004). Biproportional techniques in input-output analysis: table updating and structural analysis. *Economic Systems Research*, 16(2):115–134.

Lebacher, M., Cook, S., and Klein, N. (2021). In search of lost edges: a case study on reconstructing financial networks. *Journal of Network Theory in Finance*.

Lenzen, M., Gallego, B., and Wood, R. (2009). Matrix balancing under conflicting information. *Economic Systems Research*, 21(1):23–44.

Magerman, G., De Bruyne, K., Dhyne, E., and Van Hove, J. (2016). Heterogeneous firms and the micro origins of aggregate fluctuations. NBB Working Paper No. 312, National Bank of Belgium, Brussels.

Mastrandrea, R., Squartini, T., Fagiolo, G., and Garlaschelli, D. (2014). Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics*, 16(4):043022.

Mastromatteo, I., Zarinelli, E., and Marsili, M. (2012). Reconstruction of financial networks for robust estimation of systemic risk. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(03):P03011.

Mazzarisi, P. and Lillo, F. (2017). Methods for reconstructing interbank networks from limited information: A comparison. In *Econophysics and Sociophysics: Recent Progress and Future Directions*, pages 201–215. Springer.

McDougall, R. A. (1999). Entropy theory and RAS are friends. GTAP Working Paper No. 300, Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V. M., and Farmer, J. D. (2022). How production networks amplify economic growth. *Proceedings of the National Academy of Sciences*, 119(1).

Miller, R. E. and Blair, P. D. (2009). *Input-output analysis: foundations and extensions*. Cambridge University Press.

Miller, R. E. and Temurshoev, U. (2017). Output upstreamness and input downstreamness of industries/countries in world production. *International Regional Science Review*, 40(5):443–475.

Moussa, A. (2011). *Contagion and systemic risk in financial networks*. PhD thesis, Columbia University.

Mungo, L., Lafond, F., Astudillo-Estévez, P., and Farmer, J. D. (2022). Reconstructing production networks using machine learning. INET Oxford Working Paper No. 2022-02, Institute for New Economic Thinking at the Oxford Martin School. Available at: https://www.inet.ox.ac.uk/publications/no-2022-02-reconstructing-production-networks-using-machine-learning/.

Pankratz, N. M. and Schiller, C. M. (2019). Climate change and adaptation in global supply-chain networks. Proceedings of Paris December 2019 Finance Meeting EUROFIDAI - ESSEC, European Corporate Governance Institute – Finance Working Paper No. 775/2021. Available at SSRN: https://ssrn.com/abstract=3475416orhttp://dx.doi.org/10.2139/ssrn.3475416.

Parisi, F., Squartini, T., and Garlaschelli, D. (2020). A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks. *New Journal of Physics*, 22(5):053053.

Pichler, A. and Farmer, J. D. (2021). Simultaneous supply and demand constraints in input–output networks: The case of covid-19 in germany, italy, and spain. *Economic Systems Research*, pages 1–21.

Pöyhönen, P. (1963). A tentative model for the volume of trade between countries. *Weltwirtschaftliches Archiv*, pages 93–100.

Ramadiah, A., Caccioli, F., and Fricke, D. (2020). Reconstructing and stress testing credit networks. *Journal of Economic Dynamics and Control*, 111:103817.

Reisch, T., Heiler, G., Diem, C., and Thurner, S. (2021). Inferring supply networks from mobile phone data to estimate the resilience of a national economy. *arXiv preprint arXiv:2110.05625*.

Robinson, S., Cattaneo, A., and El-Said, M. (2001). Updating and estimating a social accounting matrix using cross entropy methods. *Economic Systems Research*, 13(1):47–64.

Squartini, T., Almog, A., Caldarelli, G., Van Lelyveld, I., Garlaschelli, D., and Cimini, G. (2017). Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks. *Physical Review E*, 96(3):032315.

Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A., and Garlaschelli, D. (2018). Reconstruction methods for networks: the case of economic and financial systems. *Physics Reports*, 757:1–47.

Squartini, T. and Garlaschelli, D. (2014). Jan Tinbergen's legacy for economic networks: from the gravity model to quantum statistics. In *Econophysics of Agent-Based Models*, pages 161–186. Springer.

Stolowy, H. and Lebas, M. (2006). *Financial accounting and reporting: a global perspective*. Cengage Learning EMEA.

Taschereau-Dumouchel, M. (2020). Cascades and fluctuations in an economy with an endogenous production network. Available at SSRN: https://ssrn.com/abstract=3115854orhttp://dx.doi.org/10.2139/ssrn.3115854.

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R., and De Vries, G. J. (2015). An illustrated user guide to the world input–output database: the case of global automotive production. *Review of International Economics*, 23(3):575–605.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

United Nations (2010). *System of National Accounts 2008*. European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, World Bank.

United Nations (2018). *Handbook on supply, use and input-output tables with extensions and applications*. United Nations, New York, US.

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. *Journal of Financial Stability*, 7(3):111–125.

Voitalov, I., van der Hoorn, P., van der Hofstad, R., and Krioukov, D. (2019). Scale-free networks well done. *Physical Review Research*, 1(3):033034.

Vom Lehn, C. and Winberry, T. (2022). The investment network, sectoral comovement, and the changing us business cycle. *The Quarterly Journal of Economics*, 137(1):387–433.

Wang, Y., Li, J., Wu, D., and Anupindi, R. (2021). When ignorance is not bliss: An empirical analysis of subtier supply network structure on firm risk. *Management Science*, 67(4):2029–2048.

Welburn, J. W., Strong, A., Nekoul, F. E., Grana, J., Marcinek, K., Osoba, O., Koirala, N., and Setodji, C. M. (2020). Systemic Risk in the Broad Economy: Interfirm Networks and Shocks in the US Economy. Santa Monica, CA: RAND Corporation, 2020. Available at: https://www.rand.org/pubs/research_reports/RR4185.html.

Wells, S. J. (2004). Financial interlinkages in the United Kingdom's interbank market and the risk of contagion. Bank of England Working Paper No. 230, Bank of England. Available at SSRN: https://ssrn.com/abstract=641288 or http://dx.doi.org/10.2139/ssrn.641288.

Wu, J. (2016). *Firm performance and risk in supply chain networks*. PhD thesis, The University of Chicago.

Yang, J., Heinrich, T., Winkler, J., Lafond, F., Koutroumpis, P., and Farmer, J. (2019). Measuring productivity dispersion: a parametric approach using the Lévy alpha-stable distribution. INET Working Paper No. 2019-14, Institue of New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, UK.

Appendices

A Data

A.1 Ecuador

Data source. The production network of Ecuador is collected through VAT filings and it is provided by the Internal Revenues Service (IRS) of Ecuador (Astudillo-Estevez, 2021). It comprises fine-grained information about all the legal (firms) and natural persons registered in the country between 2007 and 2015. We consider a dataset with 328,640 unique firms over the whole time period. In principle, it includes information on every transaction between all the entities, aggregated per year. For individual firms, it provides information on the industrial classification: ISIC Code at the 4-digit level, Rev. 4,[13] the taxpayer classification (public sector, private sector, IGO or NGO) and the fiscal address at the municipal level.

Firms need to report both their suppliers and customers. Sometimes the value of the transaction reported by the customer and by the supplier can differ. The IRS takes care of cleaning possible mismatches and we do not know which of the reported relationships are kept. The IRS is particularly concerned with detecting possible frauds related to transactions with large firms, which they identify using the weighted degree centrality. In the first couple of years of the data collection, numbers were reported manually, so the latest years are more reliable.

The dataset includes transactions between registered entities and some foreign companies that are not registered in Ecuador. Since the focus of the analysis is on the domestic product chain and because their information is incomplete, all these transactions were excluded. The dataset also contains self-loops, which represent transfers among establishments of the same firm. These transactions are not taxed and are purely for accounting purposes within the firm. We also remove these from all analyses.

Finally, we replaced one implausible value for a transaction (of the order 10^{12}) by its value in the previous year (of the order 10^{6}).

Sectoral composition. Figure A.1 shows the sectoral composition of the Ecuadorian economy according to national I-O tables (black bars) and our firm-level dataset aggregated at the sector level (green bars), for the year 2015. Since we do not have access to firm-level final demand or total revenues, we compare network sales to the sum of intermediate sales and GFCF in national accounts. We use a concordance table to translate National I-O tables (CICN codes) into the ISIC system.[14] The most important discrepancies are for Construction, which is vastly under-represented in the network, and Wholesale and retail trade, which, as expected, is vastly over-represented in the network.

[13] ISIC stands for International Standard Industrial Classification of all economic activities. It is a standard classification of economic activities where entities are classified according to the main activity they carry out.

[14] The crosswalk is available at https://contenido.bce.fin.ec/documentos/PublicacionesNotas/Catalogo/CuentasNacionales/ClasProdSCN_12042013.xlsx, downloaded on 1 December 2019. We use the crosswalk that goes from 69 to 13 industries.

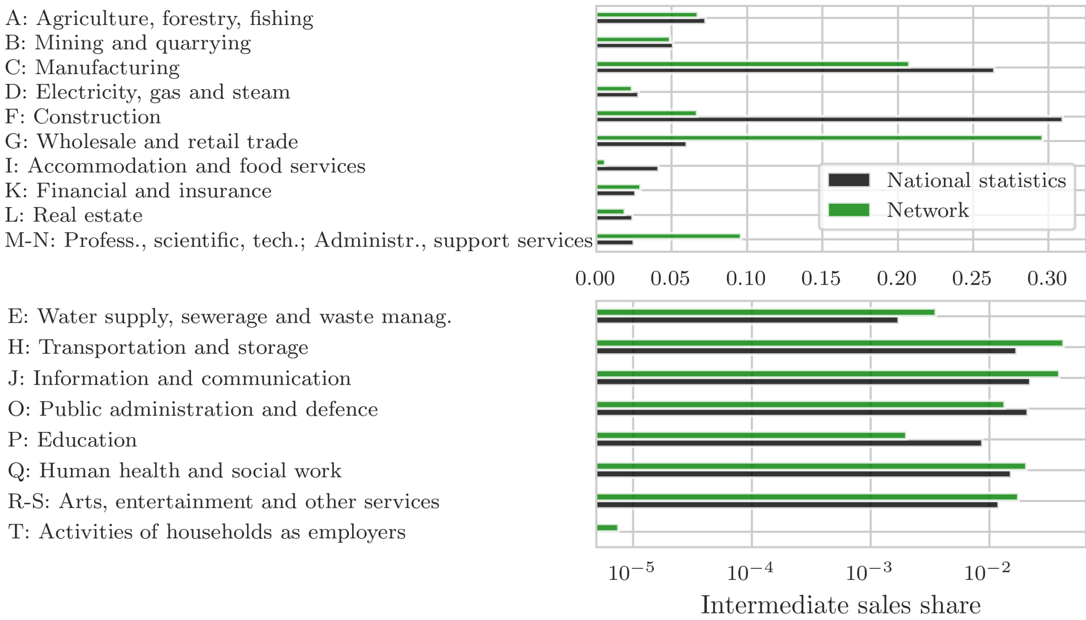

Figure A.1: Sectoral composition (intermediate sales shares) in the national I-O table at the sector level (black) and in our firm-level dataset (green) in 2015. Sectoral codes are ISIC codes at the 1-digit level (Rev. 4); sectors M and N, and R and S are grouped together to align the codes at the firm and sector level. The top panel shows intermediate sales shares in national accounts that are bigger than 0.023 and the bottom panel those that are smaller than or equal to 0.023.

A.2 FactSet: description of the dataset

We use three primary data sources provided by FactSet: Fundamentals, Supply Chain Relationships and Supply Chain Shipping Transactions. FactSet covers mostly listed firms around the world. We downloaded the datasets in April 2020.

The Supply Chain Relationships data come in part from companies' filings required by US Federal Accounting Standards[15] and in part from information on supply chain relationships released in investor presentations, company websites and press releases. Supply Chain Shipping Transactions records shipment declarations at ports from the US Bill of Lading. FactSet collects this information from the US Customs and Border Protection. Fundamentals provides firms' financial statements.

The supply chain dataset goes from 2003 to the present date, while the shipment dataset starts in 2013 and goes until the present date (we downloaded these two datasets on 11 February 2021). Due to the nature of the data collection process, coverage is biased towards companies listed on US stock exchanges, large firms and large transactions. We keep the dataset from 2014 to 2020. We do not use years prior to 2014 because in 2013 FactSet changed the data collection methodology, enhancing the quality of the dataset.

In the raw data, links report the year, month, day and hour. The start and end dates correspond to when the record was first published and when the ending was announced. We consider that a relationship exists in a given year if it exists at any point during that year. For each company, we also have information on the sector (NACE Rev.2 codes at the 4-digit level) and the country where the company's headquarters are located.

The monetary values of customer-supplier transactions are rarely recorded. When recorded, the money flows are reported as a revenue percentage earned by the seller from a specific customer. However, it is unknown to what disclosed revenue figure that percentage refers. For example, it could be the quarterly or the annual income statement, but also some interim revenue disclosure not tied to any specific financial statement. Similarly, in the shipment dataset, the cumulative value of the goods shipped, when disclosed, is in most cases above the seller's revenues and/or the customers' costs since valuation methods do not necessarily match balance sheet information. Therefore, we dismiss this information when reconstructing the money flows and use only the binary topology. There are other,

[15]The Statement of Financial Accounting Standards No. 131 requires publicly traded firms on US stock exchanges to report customers that account for 10% or more of their annual revenues, formally called *major customers*.

less comprehensive datasets, such as Compustat, which report the revenue percentage earned by the seller. These could be potentially merged to enhance the reconstruction. We leave this for future work.

We consider industrial firms only, hence we exclude firms in the financial and insurance sector, and firms classified as extraterritorial organisations. We aggregate customer-supplier relations within a year. We use the fiscal year instead of the calendar year to ensure time consistency between the formation of supplier-customer relations and financial statements. The fiscal year goes from June to May, meaning that if a company's fiscal year end-month falls between January and May, the fiscal year is the current calendar year minus one, otherwise it is the current calendar year.

We further aggregate all three FactSet datasets at the parent company level, meaning that we use consolidated income statements and attribute subsidiaries' supplier-customer relations to the parent company. We delete self-loops (i.e., supply chain relations among the parent and its subsidiaries) as these stem from intra-group sales that cancel out in consolidated income statements. We rely on the latest available information on a company's ownership structure since it is impossible to know the evolution of companies' ownership structures (mergers, acquisitions, buy-backs, etc.). For each company, we also have information on the sector (NACE Rev.2 codes at the 4-digit level) and the country where the company's headquarters are located.

A.2.1 Methods for writing the income statement

There are two main methods companies can follow to write their income statements: the *function of expense* method or the *nature of expense* method – some companies adopt a hybrid method. As the names suggest, the nature of expense method lists expenses based on their nature, while the function of expense method lists expenses based on their function. Thus, the nature of expense method breaks down expenses based on the inputs used to perform the business activity, e.g., materials, delivery charges, changes in inventory, rent, labour expenses and employee benefits. The function of expense method allocates expenses based on the activity for which the expense arises, i.e., "if the expense did not contribute to the creation of the [good] or service that is the underlying source of sales revenue, they are not part of the cost of goods sold." (Stolowy and Lebas, 2006, p. 208).

A.2.2 Labour expenses

Depending on the method followed in writing the income statement (explained in Appendix A.2.1), companies may disclose or not their labour expenses as a separate line item. If a company follows the *nature of expense* method, labour costs and the cost of intermediate inputs are broken down into two separate items on the income statement. However, if a company follows the *function of expense* methods, labour and intermediate inputs expenses are lumped together in the cost of goods sold. Moreover, even if they follow a function of expenses method, some companies may disclose their labour costs in a footnote.

FactSet collects the labour expenses disclosed in the footnote but without flagging the method followed by the company in writing its income statement. Therefore, there is no systematic way to know whether a company follows the nature of expense or the function of expense method. The only reasonable assumption is that if labour expenses are not observed in the dataset, it is because the company followed the nature of expenses method and thus the cost of goods sold correctly represents intermediate inputs costs. Consequently, it is impossible to extract labour costs from the cost of goods sold only when necessary.

Estimating missing labour expenses. For constructing the I-O table, we must discern labour and intermediate inputs expenses as they correspond to different parts of the I-O table. Labour costs are part of value-added, while intermediate input expenses correspond to in-strengths, i.e., the column sums of the weighted adjacency matrix (also referred to as inter-industry transaction matrix in sector-level data). Having an exact quantification of intermediate expenses is necessary to reconstruct the firm-to-firm money flows (i.e., network weights).

To estimate labour expenses for those firms that do not report them, we use the cost of goods sold share of labour expenses disclosed by other firms in the same sector. For each sector, we calculate the sector-level cost of goods sold share of labour expenses using four different estimation strategies, all of which use a rolling window approach (explained below). In this paper, we use the method that minimises the root mean squared error, the average absolute error and the median absolute error. On the one hand, for the cost of goods sold share of labour expenses, methods 2a and 3 (Equation A.3 and A.4, respectively) achieve the minimum RMSE and MAE while method 2b (Equation A.3) achieves the lowest MedAE (Table A.2). On the other hand, for at the estimated labour costs and intermediate costs, it is method 2b (Equation A.3) that has the minimum error (Table A.3). Therefore, we use method 2b (Equation A.3) in the paper.

We estimate labour expenses only from 2013 onward and start the rolling window in 2011 because before 2011 the number of firms in each sector was too low. For firms that disclose their labour expenses only for a few years, we extrapolate the missing values of the cost of goods sold share of labour expenses from those that we observe: (1) if we observe labour expenses up to time t and the missing values are only from $t+1$ onward, we use the last value of the cost of goods sold share of labour expenses and propagate it forward; (2) if we observe labour expenses from t onward but do not observe labour expenses form 2013 till t, we backpropagate the first available observation; and (3) if we do not observe labour costs in the middle, meaning that we do observe labour costs from 2013 to t and from $t+\tau$ until the end, we estimate the cost of goods sold share of labour expenses with linear interpolation. Table A.1 reports the percentage of firms that disclose labour costs for each sector average over time (second column) and its standard deviation (last column). Figure A.2 shows the density of firms' cost of goods sold share of labour expenses for each sector (NACE Rev. 2 at the 2-digit level) over time.

NACE division description	Av. pct firms	Standard dev.
Accommodation	0.618	0.038
Activities of head offices; management consultancy activities	0.528	0.022
Activities of membership organizations	0.750	0.274
Advertising and market research	0.575	0.026
Air transport	0.882	0.042
Architectural and engineering activities; technical testing and analysis	0.732	0.024
Civil engineering	0.663	0.015
Computer programming, consultancy and related activities	0.530	0.022
Construction of buildings	0.608	0.021
Creative, arts and entertainment activities	0.944	0.167
Crop and animal production, hunting and related service activities	0.653	0.052
Education	0.554	0.052
Electricity, gas, steam and air conditioning supply	0.639	0.018
Employment activities	0.541	0.029
Extraction of crude petroleum and natural gas	0.489	0.040
Fishing and aquaculture	0.707	0.068
Food and beverage service activities	0.430	0.020
Forestry and logging	0.722	0.105
Gambling and betting activities	0.758	0.035
Human health activities	0.692	0.016
Information service activities	0.450	0.028
Land transport and transport via pipelines	0.562	0.060
Legal and accounting activities	0.669	0.093
Libraries, archives, museums and other cultural activities	0.750	0.289
Manufacture of basic metals	0.677	0.023
Manufacture of basic pharmaceutical products and pharmaceutical preparations	0.610	0.026
Manufacture of beverages	0.685	0.035
Manufacture of chemicals and chemical products	0.603	0.031
Manufacture of coke and refined petroleum products	0.739	0.023
Manufacture of computer, electronic and optical products	0.633	0.027
Manufacture of electrical equipment	0.680	0.050
Manufacture of fabricated metal products, except machinery and equipment	0.634	0.022
Manufacture of food products	0.680	0.014

Manufacture of furniture	0.604	0.028
Manufacture of leather and related products	0.709	0.022
Manufacture of machinery and equipment n.e.c.	0.575	0.047
Manufacture of motor vehicles, trailers and semi-trailers	0.629	0.043
Manufacture of other non-metallic mineral products	0.657	0.021
Manufacture of other transport equipment	0.639	0.036
Manufacture of paper and paper products	0.661	0.015
Manufacture of rubber and plastics products	0.665	0.024
Manufacture of textiles	0.776	0.027
Manufacture of tobacco products	0.698	0.062
Manufacture of wearing apparel	0.643	0.037
Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	0.656	0.047
Mining of coal and lignite	0.649	0.053
Mining of metal ores	0.719	0.037
Mining support service activities	0.594	0.061
Motion picture, video and television programme production, sound recording and music publishing activities	0.679	0.025
Office administrative, office support and other business support activities	0.647	0.029
Other manufacturing	0.548	0.025
Other mining and quarrying	0.711	0.029
Other personal service activities	0.548	0.048
Other professional, scientific and technical activities	0.612	0.038
Postal and courier activities	0.813	0.089
Printing and reproduction of recorded media	0.720	0.007
Programming and broadcasting activities	0.586	0.047
Public administration and defence; compulsory social security	0.593	0.206
Publishing activities	0.549	0.026
Real estate activities	0.493	0.044
Remediation activities and other waste management services	0.717	0.076
Rental and leasing activities	0.515	0.033
Repair and installation of machinery and equipment	0.681	0.022
Repair of computers and personal and household goods	1	0
Residential care activities	0.588	0.041
Retail trade, except of motor vehicles and motorcycles	0.500	0.021
Scientific research and development	0.569	0.038
Security and investigation activities	0.688	0.033
Services to buildings and landscape activities	0.508	0.045
Sewerage	0.483	0.059
Social work activities without accommodation	0.680	0.096
Specialized construction activities	0.592	0.037
Sports activities and amusement and recreation activities	0.689	0.044
Telecommunications	0.690	0.019
Travel agency, tour operator, reservation service and related activities	0.621	0.045
Veterinary activities	0.972	0.083
Warehousing and support activities for transportation	0.690	0.031
Waste collection, treatment and disposal activities; materials recovery	0.651	0.034
Water collection, treatment and supply	0.603	0.057
Water transport	0.577	0.030
Wholesale and retail trade and repair of motor vehicles and motorcycles	0.658	0.019
Wholesale trade, except of motor vehicles and motorcycles	0.575	0.013

Table A.1: Time average of the percentage of firms that disclose labour costs per sector.

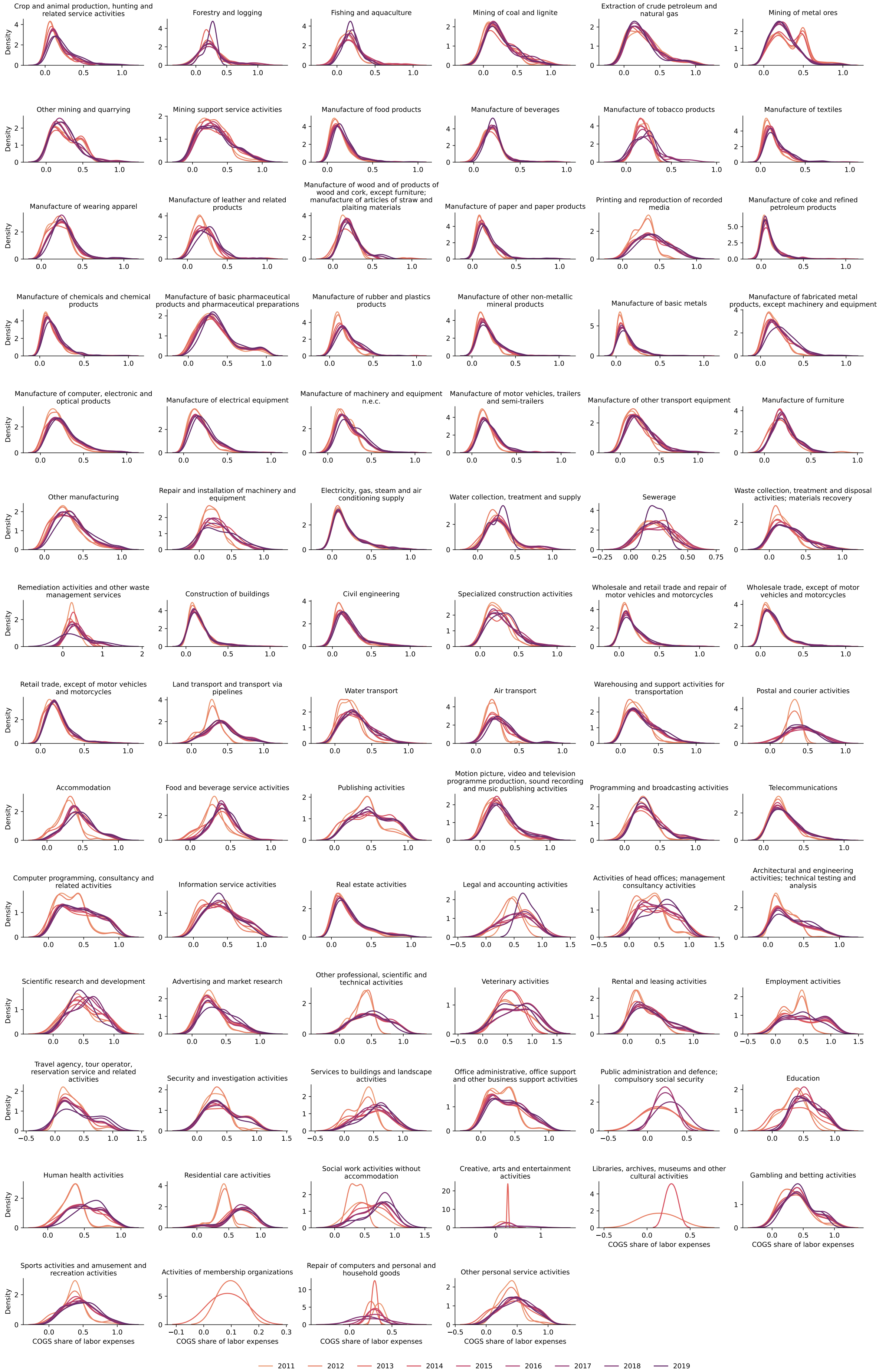

Figure A.2: Density of the cost of goods sold share of labour expenses for firms in different sectors and over time (2011-2019). The titles show the NACE Rev. 2 sectoral classification at the 2-digit level.

Let $w_{i,t}$ be labour expenses of firm i at time t and $g_{i,t}$ its cost of goods sold excluding depreciation and amortisation (for firms employing a *by nature* method in writing their income statements), then the cost of goods sold share of labour expenses of firm i at time t is defined as

$$\alpha_{i,t} = \frac{w_{i,t}}{g_{i,t} + w_{i,t}} \ . \tag{A.1}$$

For some firms, we observed labour expenses for some years but not for other years.

Estimation method 1. Firstly, we calculate a firm's average cost of goods sold share of labour expenses over a three-year period starting from 2013. For example, $\alpha_{i,2013}$ is computed using a rolling window going from 2011 to 2013, while for $\alpha_{i,2014}$ the rolling window goes from 2012 to 2014.[16] A firm's average cost of goods sold share of labour expenses is given by

$$\bar{\alpha}_{i,t} = \frac{1}{3} \sum_{\tau=t-2}^{t} \alpha_{i,\tau} \ .$$

Afterwards, for each year t, we compute sector s' cost of goods sold share of labour expenses by averaging over the α_i's of the firms in sector s at time t:

$$\tilde{\alpha}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{S}} \bar{\alpha}_{i,t} \ , \tag{A.2}$$

where $N_{s,t}$ is the number of firms in sector s at time t and $\mathcal{S}$ is the set of firms in sector s.

Estimation method 2. We calculate a sector-level time-varying cost of goods sold share of labour expenses in the following two ways.

2.a We compute a sector's cost of good sold share of labour expenses for each time step t by averaging over the α_i's of firms in sector s:

$$\alpha^{\mathrm{a}}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{S}} \alpha_{i,t} \ .$$

2.b We calculate a sector's cost of goods sold share of labour expenses by summing the labour expenses of firms in sector s at time t and dividing these by the sum of firms' labour and intermediate costs:

$$\alpha^{\mathrm{b}}_{s,t} = \frac{\sum_{i \in \mathcal{S}} w_{i,t}}{\sum_{i \in \mathcal{S}} w_{i,t} + \sum_{i \in \mathcal{S}} g_{i,t}} \ .$$

After using either of the two methods above (2.a or 2.b), to compute the sector-level cost of goods sold share of labour expenses at time t, we employ a three-year rolling window to calculate the time average:

$$\bar{\alpha}^{k}_{s,t} = \frac{1}{3} \sum_{\tau=t-2}^{t} \alpha^{k}_{s,\tau} \ , \tag{A.3}$$

for $k = \{\mathrm{a}, \mathrm{b}\}$, indicating whether we used method 2.a or 2.b in the first step.

Estimation method 3. We start from $\alpha^{a}_{s,t}$ calculated using method 2.a and then compute a weighted average using a three-year rolling window:

$$\bar{\alpha}^{*}_{s,t} = \sum_{\tau=t-2}^{t} \theta_{s,\tau} \alpha^{a}_{s,\tau} \ ,$$

the $\theta_{s,t}$'s are weights that sum up to one. $\theta_{s,t}$ is the share of sector s at time t in the total expenditure of the sector on labour and intermediate inputs during the three years. $\theta_{s,t}$ is defined as:

[16] For some firms, picked at random, we look at how their α's change over time; changes are negligible.

$$\theta_{s,t} = \frac{\sum_{i \in \mathcal{S}} g_{i,\tau} + w_{i,\tau}}{\sum_{\tau=t-2}^{t} \sum_{i \in \mathcal{S}} g_{i,\tau} + w_{i,\tau}} . \tag{A.4}$$

Assessing the methods. Figure A.3 shows the empirical against the predicted cost of goods sold share of labour expenses for our four estimation methods. They all seem to perform very similarly except for method 2b, which has the lowest average absolute error but the highest root mean squared error and median absolute error. Among the others, methods 2a and 3 (Equation A.3 and A.4, respectively) have the same and lowest error metrics (Table A.2).

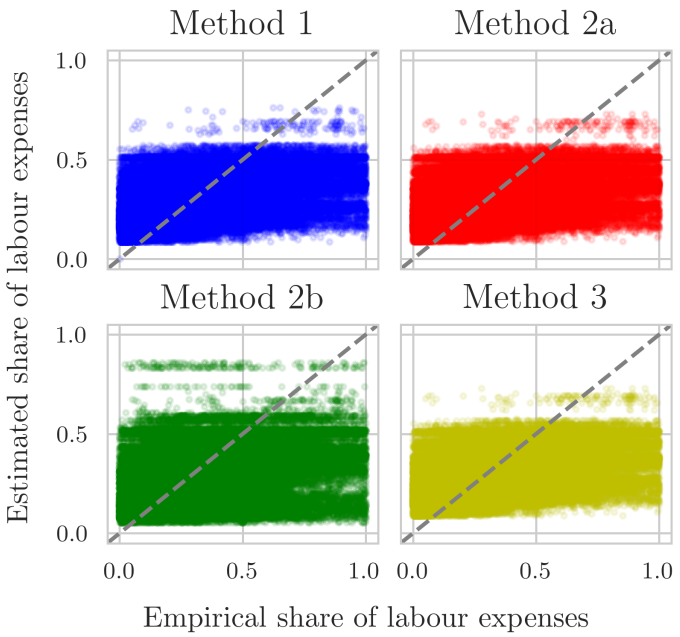

Figure A.3: Empirical (x-axis) and predicted (y-axis) cost of goods sold share of labour expenses for our four estimation methods.

	RMSE	MAE	MedAE
Method 1	0.184	0.138	0.106
Method 2a	0.184	0.137	0.105
Method 2b	0.201	0.144	0.098
Method 3	0.184	0.137	0.105

Table A.2: Error metrics for the cost of goods sold share of labour expenses for the four estimation methods we devised. RMSE denotes the root mean squared error, MAE the mean absolute error and MedAE the median absolute error.

Figure A.4 shows the empirical against the predicted labour expenses (top) and intermediate expenses (bottom) for our four estimation methods. They all seem to perform very similarly. Method 2b (Equation A.3) has the lowest error metrics (Table A.3). The error metrics are the same for the labour and intermediate expenses because they mirror each other. Had we not taken the absolute values, for instance, the error metrics would have opposite sign.

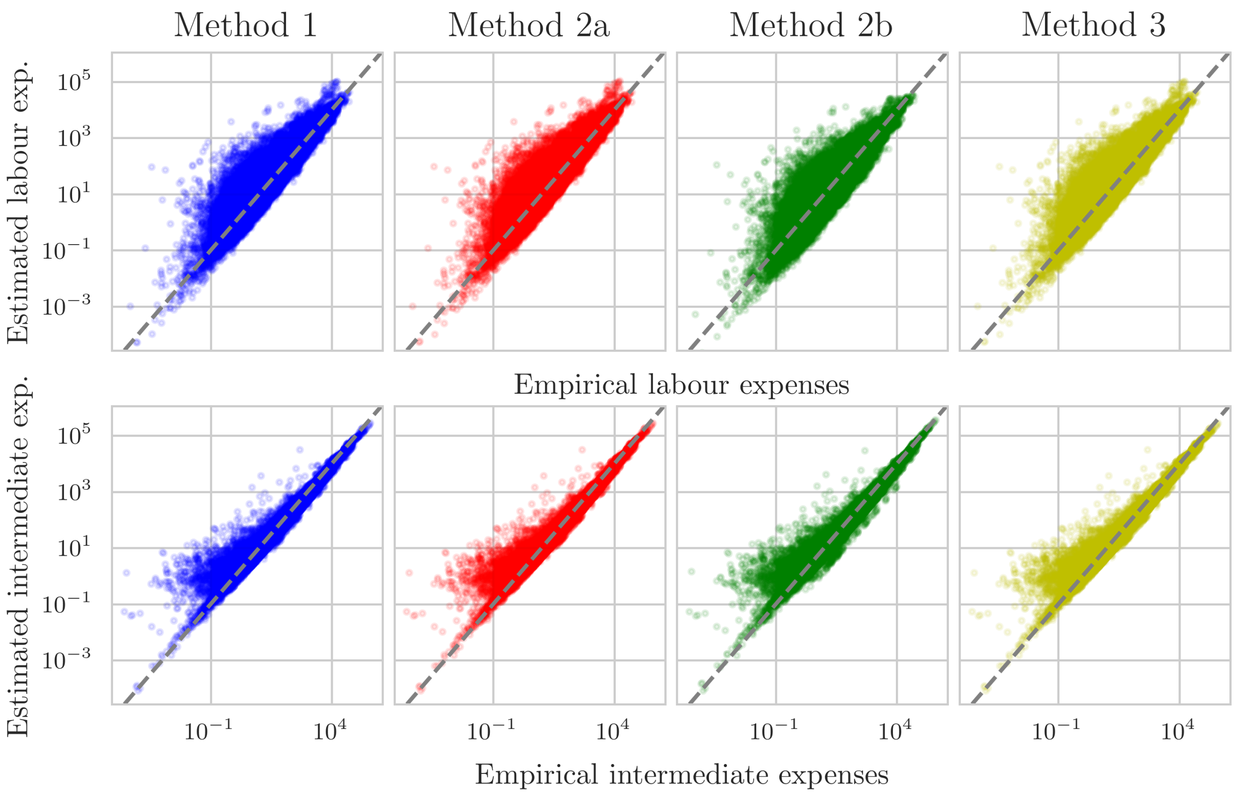

Figure A.4: Empirical (x-axis) and predicted (y-axis) labour expenses (top) and intermediate expenses (bottom) for our four estimation methods.

	Labour exp.			Intermediate exp.		
	RMSE	MAE	MedAE	RMSE	MAE	MedAE
Method 1	1,153.6	126.0	8.5	1,153.6	126.0	8.5
Method 2a	1,163.0	126.1	8.5	1,163.0	126.1	8.5
Method 2b	497.0	90.2	8.4	497.0	90.2	8.4
Method 3	1,162.0	126.1	8.5	1,162.0	126.1	8.5

Table A.3: Error metrics for labour and intermediate expenses for the four estimation methods we devised. RMSE denotes the root mean squared error, MAE the mean absolute error and MedAE the median absolute error.

A.2.3 Value-added

Value-added can be calculated using income statements. It is defined in different ways in the literature. For instance, Yang et al. (2019) calculate it as the sum of deflated wages and deflated earnings before interest and taxes. Magerman et al. (2016) calculate value-added as the difference between sales and material costs, which is standard in the production network literature (at the firm level). We adopt a third strategy based on Miller and Blair's (2009) book on input-output analysis, which also corresponds to the definition in the WIOD (which we use to construct the proxy node).

Following Miller and Blair (2009), valued-added is the wage bill (labour costs) plus EBITDA (earnings before interest and taxes). Their definition, which follows national and international accounting standards, also includes amortisation and depreciation. We do not observe the wage bill for all companies since labour expenses are disclosed depending on the method followed in writing the income statement. Therefore, for some companies, we estimate labour costs as explained in Section A.2.2.

A.2.4 Cleaning financial statements for the construction of the I-O table

We kept firms with positive sales, intermediate expenses and value-added, and non-negative labour costs (all firms with zero labour costs have disclosed them), EBITDA and amortisation and depreciation.

Sanity check. We checked that firms in the dataset respect the accounting identity:

$$\text{sales} = \text{intermediate\_expenditure} + \text{value\_added} + \text{other\_costs}.$$

Since we did not account for other costs, we expected the residual (other_costs) to be non-negative. However, in 2014, for 30% of the firms, the residual is negative, meaning that sales were smaller than the sum of the variables on the RHS (excluding other_costs).

Most firms with a negative residual disclosed their labour expenses: is it possible that we double-counted labour expenses? This would mean that a firm adopts a *by function* method to write the income statement but reports labour costs in a footnote, which would lead FactSet to record the labour expenses. If this were the case, labour costs would already be in the cost of goods sold, leading to counting this expense twice. Thus, we subtract the disclosed labour costs from intermediate expenses for these firms. Is this reasonable? We checked whether this subtraction caused any firm to have negative intermediate expenses. It did, but only for 0.5% of the observations for which we did the procedure.

A positive consequence of the exercise just described is that it allows us to identify firms that likely adopt a *by function* method to write the income statement. We propagate this information back in our estimation of the cost of goods sold share of labour expenses in order to adjust the denominator in Equation A.1, which for these firms would otherwise be double counting labour costs. After doing this procedure only 1.2% of the firms in 2014 do not respect the accounting identity. We drop these firms. The cleaning procedure led to all firms in the sample having a strictly positive value-added.

A.3 FactSet: evaluation of coverage

First, we show the types of firms that are in FactSet. Second, we assess the coverage of global economic activity covered by FactSet by comparing it to the WIOD. We compare the countries covered and the sectoral composition in FactSet against that of the WIOD. Subsequently, we quantify the percentage of world gross output captured by FactSet and then evaluate the growth rates of gross output.

A.3.1 Firm types

Figure A.5 shows the percentage of firms for each firm type as defined by FactSet for the year 2014 (there are 5,442 firms). In the dataset, there are some big private companies that also file financial statements and companies that are extinct by April 2020, when the dataset was downloaded.

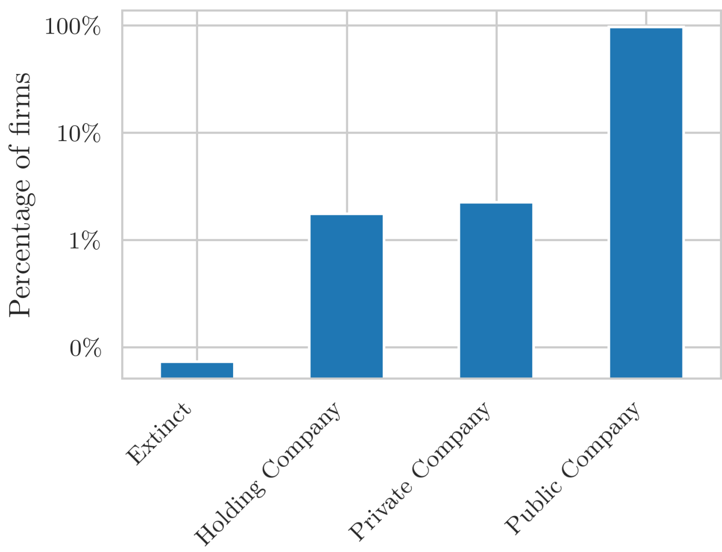

Figure A.5: Percentage of firms per type in FactSet in 2014. The firm type is defined by FactSet.

A.3.2 Countries covered in WIOD and FactSet

Table A.4 shows the countries that are in the WIOD and the countries where firms in FactSet have their headquarters for the period 2013–2019. The WIOD has a node named "rest of the world" (RoW), which captures all the other countries not in the list. The number of overlapping countries is 41. The Czech Republic and Estonia are covered in the WIOD but not in FactSet.

WIOD (43 countries)	FactSet (87 countries)
Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, India, Indonesia, Ireland, Italy, Japan, Korea (the Republic of), Latvia, Lithuania, Luxembourg, Malta, Mexico, Netherlands (the), Norway, Poland, Portugal, Romania, Russian Federation (the), Slovakia, Slovenia, Spain, Sweden, Switzerland, Taiwan (Province of China), Turkey, United Kingdom of Great Britain and Northern Ireland, United States of America (the).	Argentina, Australia, Austria, Bahamas (the), Bahrain, Bangladesh, Belgium, Bermuda, Brazil, Bulgaria, Canada, Cayman Islands (the), Chile, China, Colombia, Costa Rica, Croatia, Cyprus, Côte d'Ivoire,, Denmark, Egypt, Faroe Islands (the), Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Korea (the Republic of), Kuwait, Latvia, Lithuania, Luxembourg, Macao, Malaysia, Malta, Marshall Islands (the), Mauritius, Mexico, Monaco, Morocco, Netherlands (the), New Zealand, Nigeria, Norway, Oman, Pakistan, Panama, Peru, Philippines (the), Poland, Portugal, Qatar, Republic of North Macedonia, Romania, Russian Federation (the), Saudi Arabia, Singapore, Slovakia, Slovenia, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan (Province of China), Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Arab Emirates (the), United Kingdom of Great Britain and Northern Ireland (the), United States of America (the), Venezuela (Bolivarian Republic of), Vietnam, Zambia

Table A.4: Countries covered by the WIOD and FactSet.

A.3.3 Sectoral composition

Figure A.6 shows the sectoral composition in the WIOD (black bars) and in FactSet aggregated at the sector level (green bars) using firms' revenues. We assess the sectoral composition using the sectors' gross output shares (or revenues for firms). We group ISIC Rev.4 codes at the 1-digit level into eight higher-level classes shown in Table A.5.

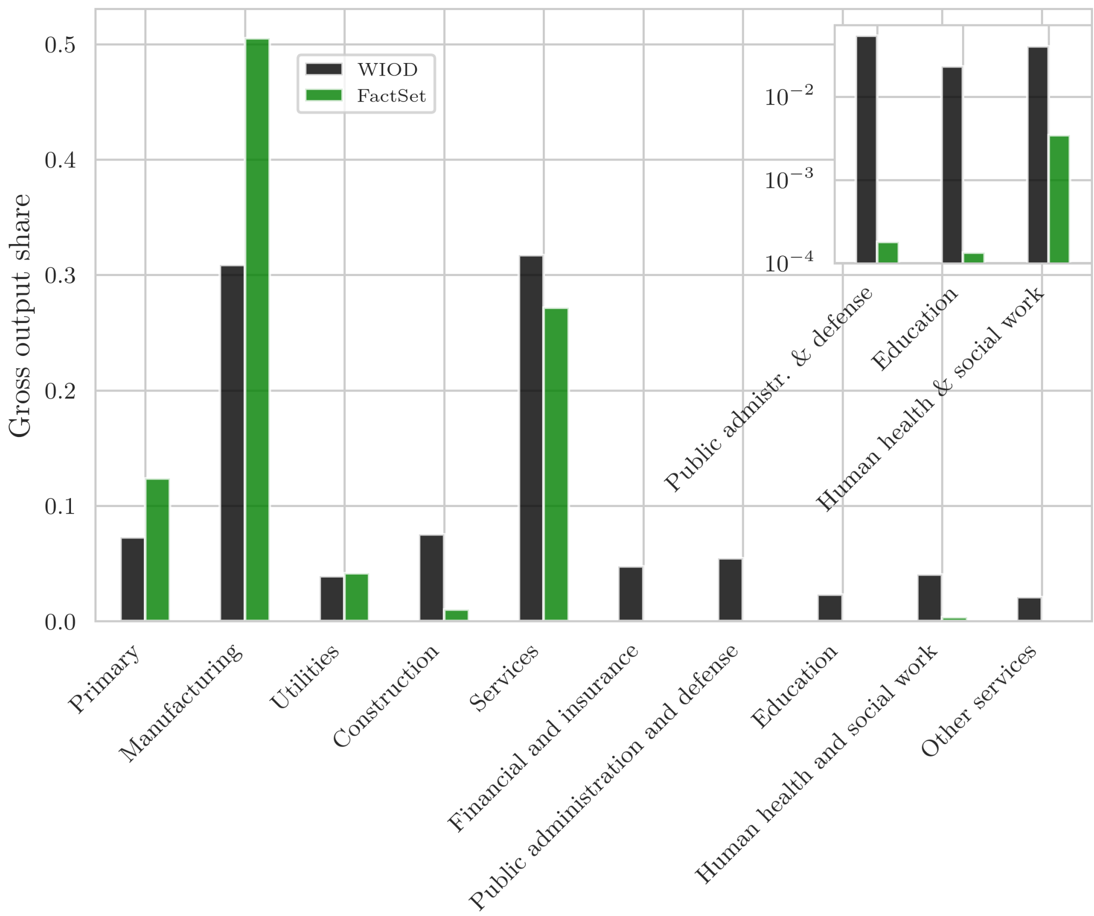

Figure A.6: Sectoral composition (gross output shares) of the WIOD (black) and FactSet (green) in 2014. For FactSet, we group firms into sectors and use firms' revenues. Sectors are defined as in Table A.5, where we aggregate ISIC Rev. 4 codes at the 1-digit level into macro classes.

ISIC	Description	Macro class
A	Agriculture, forestry and fishing	Primary
B	Mining and quarrying	Primary
C	Manufacturing	Manufacturing
D	Electricity, gas, steam and air conditioning supply	Utility
E	Water supply; sewerage, waste management and remediation activities	Utility
F	Construction	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	Services
H	Transportation and storage	Services
I	Accommodation and food service activities	Services
J	Information and communication	Services
K	Financial and insurance activities	Financial and insurance
L	Real estate activities	Services
M	Professional, scientific and technical activities	Services
N	Administrative and support service activities	Services
O	Public administration and defence; compulsory social security	Public administration and defence
P	Education	Education
Q	Human health and social work activities	Human health and social work
R	Arts, entertainment and recreation	Other services
S	Other service activities	Other services
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	Not included
U	Activities of extraterritorial organisations and bodies	Not included

Table A.5: Description of ISIC Rev. 4 codes at the 1-digit (left column) and our macro-level classes (right column).

A.3.4 Evaluation of gross output

To understand how well FactSet captures global economic activity, we compare gross output levels and growth rates in FactSet to national accounting data (WIOD). We show the evaluation from 2013 to 2018, although we only use 2014.

To compare with world economic activity, we use gross output from the WIOD. The last year for which the WIOD is available is 2014; therefore, we forecast world gross output from 2015 to 2018.[17] We assess both the levels and growth rates of gross output.

Levels. A comparison of the time series of aggregate firms' revenues and world gross output is shown in Figure A.7a. FactSet captures, on average, 16.4% of world gross output over time. The yearly percentage of world gross output captured by FactSet is given by

$$\phi_t = \frac{\sum_i q_{i,t}}{q_t}, \tag{A.5}$$

where q_t is world gross output at time t and $q_{i,t}$ is firm i's sales at time t. When forecasting world gross output, besides our central estimate, we also calculate a best and worst case (see Footnote 17); these yield a lower bound of 15.1% and an upper bound of 18.3% on the central estimate.

[17] We forecast world gross output from 2015 until 2018 using GDP from the World Bank as follows. We take the ratio of gross output to GDP, which is known to be fairly stable over time, and assume that after 2014 it stays constant. This gives us gross output q_t as a function of GDP y_t and the gross output/GDP ratio ζ_t, thus $q_t = \zeta_{2014} \cdot y_t$ for all $t = 2015, \ldots, 2020$. Data for the World Bank time series are available at https://data.worldbank.org/indicator/CM.MKT.LDOM.NO; data were downloaded on 27 July 2020.

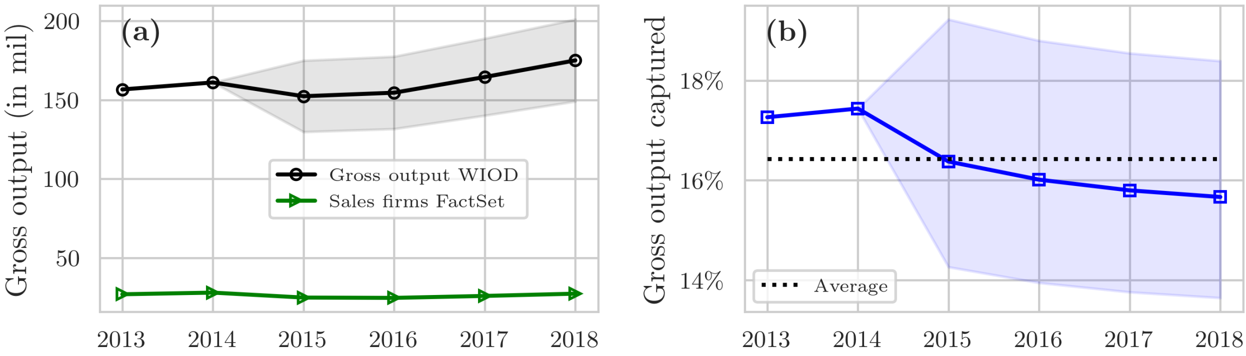

Figure A.7: **(a)** World gross output in the WIOD (black dotted line) and cumulative revenues of firms in FactSet (green line with triangles). Values in millions of US dollars. We show error bars for the years for which we forecast gross output (2015–2018); see Footnote 17 for a description of the forecasting methodology. **(b)** Percentage captured by FactSet of world gross output (WIOD, Equation A.5). The shaded blue area shows error bounds related to the gross output forecasts. The dashed black line marks the time average.

Growth rates. For the growth rate of gross output, we carry out the same assessment described above for the levels of gross output. Figure A.8a shows the growth rate of gross output in our aggregated FactSet dataset (green line with triangles) and in the WIOD (black line with circles). Over time, FactSet's growth rates differ from world growth rates by 2.3 percentage points on average. To assess how much the growth rates differ, we used the following metric

$$\phi_t = |g(\tilde{q}_t) - g(q_t)|, \tag{A.6}$$

where $g(q_t)$ denotes the growth rate of world gross output at time t and the tilde indicates our aggregate variable constructed from firms' revenues.

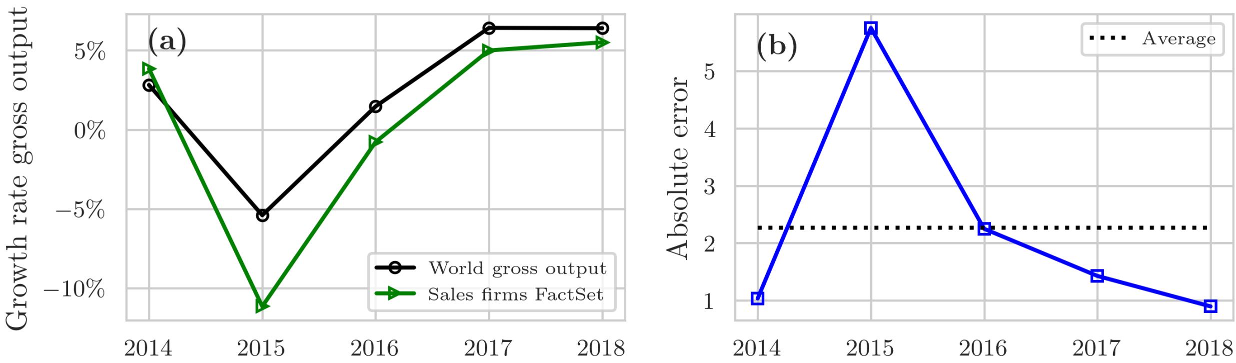

Figure A.8: **(a)** Growth rate of world gross output (WIOD, black line with circles) and the growth rates of firms' revenues in FactSet (green line with triangles). As in Figure A.7, we forecast world gross output from 2015 to 2018. **(b)** Absolute error between the growth rate of world gross output (WIOD) and the growth rate of firms' revenues (FactSet). The y-axis shows percentage points and the dotted black line shows the time average.

A.4 Merging FactSet with sector-level I-O tables (the WIOD)

Since FactSet does not cover the entire economy, we supplement it with one proxy node constructed from input-output data at the sector level; we use the WIOD. I-O tables at the sector level capture economic linkages between industries in a country and sometimes also across countries. These tables provide information on the monetary flows, related to input expenditure and consumption, from each industry to each of the other industries, itself and other agents in the economic system such as households and the government sector.

Figure A.9 shows the structure of the WIOD, which is standard in aggregate I-O datasets based mostly on national accounts. On the rows, there are the supplying industries grouped by country and on the columns are the customers. Thus, each column portrays an industry production "recipe" and the row its customer base. The central block, which involves only trades among industries, is called the inter-industry transaction matrix or intermediate consumption. On its right, there is final demand (i.e., final uses), which comprises both consumption and investment demand. At the bottom, there is value-added. The 2016 version of the WIOD covers 43 countries and 56 sectors from 2000 to 2014.

			Use by country-industries							**Final use by countries**			**Total use**
			Country 1			…	Country M			Country 1	…	Country M	
			Industry 1	…	Industry N	…	Industry 1	…	Industry N		…		
Supply from country-industries	Country 1	Industry 1											
		…											
		Industry N											
	……												
	Country M	Industry 1											
		…											
		Industry N											
Value added by labour and capital													
Gross output													

Figure A.9: Example of a multi-country I-O table; here we show the World Input-Output Database. *Source*: Timmer et al. (2015, p.577).

We construct the I-O table using FactSet and supplement it with a synthetic node constructed using the WIOD. The synthetic, or proxy, node is connected to all the firms (both incoming and outgoing links). The synthetic node accounts for the production, sales and value-added of firms that are not in our network (i.e., the rest of the global economy). While the WIOD is at the country-sector level, we aggregate it into only one node for two reasons. Firstly, the firms in our supply chain data are multinational companies aggregated at the group level, as such their production is scattered around the world, making it difficult to attribute their sales, expenditure and value-added to a specific country. Secondly, while we tried to keep the sector dimension – which would have allowed us to have one proxy node per sector – differences between national accounting standards (WIOD) and financial accounting standards (FactSet) prevented us from accurately constructing an I-O table at the firm-sector level. We discuss the differences between national accounts and firms' financial accounts in Appendix A.4.1 and A.4.2.

As shown in Figure A.9, an I-O table is composed of three main parts: the central block (inter-firm or inter-industry transaction matrix), final demand and value-added. The structure of the I-O table implies that there are four main variables to construct for and harmonise with our firm-level data when introducing the proxy sector (we discuss this in more detail in Appendix A.4.2). These four variables are total and intermediate sales and costs, value-added and final demand. We build these variables for the proxy sector by subtracting from the WIOD totals the observed firms' cumulative values (for the firms in the cleaned FactSet dataset). While these four variables are relatively easy to identify in sector-level I-O tables, labour expenses and final demand pose challenges in their exact quantification at the firm level. Before delving into our procedure for harmonising firms' financial statements with national accounts (Appendix A.4.2), we discuss the differences between firm-level data and national I-O tables (Appendix A.4.1).

A.4.1 Differences between firm-level networks and national I-O tables

We provide a non-technical discussion of the differences between firm-level data and the Supply-Use (SU) or I-O tables framework from national accounts. For a detailed handbook on the compilation of SU tables, see United Nations (2018).

Investment vs intermediate consumption. I-O tables are central to national accounts as they make it possible to compute GDP (i.e., total value-added) by subtracting intermediate consumption from gross output (i.e., total sales). Because net investment spending (i.e., gross fixed capital formation) is part of final demand, national accounts record transactions for intermediate consumption in the inter-industry transaction matrix, while transactions for investment goods and services are in a separate column. By contrast, firm-to-firm transaction networks include both intermediate and investment transactions indiscriminately. In practice, this can lead to a substantial bias: the total transactions observed within the firm network should in principle be higher than those in the inter-industry transaction matrix. If investment is 25% of GDP,[18] and intermediate transactions are about as large as GDP, the network transactions should be 25% higher than in the I-O table. This bias should be highly heterogenous across industries: Vom Lehn and Winberry (2022) reconstruct the investment network for 37 industries in the US, showing that a few industries represent a very high share of investment goods: construction, machinery, professional and technical services, and motor vehicles.

Wholesale trade, retail trade and transport. In national accounts, the convention is that wholesale, retail and transport should be better thought of as "pass-through sectors" rather than producing and consuming in a similar way as the other sectors do. More precisely, national accounts consider that the output of these industries is not their total sales but the *margins* they apply over the goods they buy and sell or transport. When industry j buys goods produced by industry i through a wholesaler k, I-O tables are thus able to record the flow of goods directly between industries i and j. The total cost paid by industry j is then split into the sales proceeds for industry i and a "trade margin" received by k. Another way to think about this is to consider that the wholesaler is a service provider – its true inputs are, say, labour, electricity and real estate, not the goods that it buys only for reselling. In sharp contrast, in firm-to-firm transaction data, we would observe the wholesaler buying the goods and reselling them and we would not observe a transaction between industry i and j. Therefore, we expect that the total sales of wholesale trade, retail trade and transport would be much higher in firm data than in the I-O tables (roughly 5 times higher if margins are 20%). Furthermore, we also expect the structure of the inter-industry matrix to be substantially different.

Financial institutions and financial services. The measurement of financial services in national accounts is complex. Additionally, financial institutions usually obey specific accounting rules and regulations. As a result, it is customary to remove financial firms when analysing firm-level datasets. Although we have opted to keep all the firms in our analysis, we expect that firm-level network datasets based on VAT, surveys or payment systems would show quite different flows in and out of financial firms. Reconciling this with national accounts should proceed on a case-by-case basis.

Unit of analysis and industrial classification. To construct I-O tables, national statistical offices conduct surveys at the establishment level rather than the firm level. Establishments are preferred to aggregate production units into sectors because they tend to have a more homogeneous production. However, the unit in micro-level datasets is typically a firm; this can cause significant issues as multi-establishment firms are common.[19] Having multi-establishment firms implies that to aggregate firm-level data into proper I-O tables, one needs to split a firm's output into various industries and then decide how to split the firm's inputs into the various output, a well-known problem for constructing symmetric input-output tables from SU tables (Miller and Blair, 2009).

Prices and volume measures. National I-O tables use different concepts of prices. Roughly speaking and omitting imports and exports, on the resources side (output + net taxes + transport and trade margins), a unit of output is evaluated at the basic price; that is, the price that the producer can obtain, omitting taxes that go back directly to the government and including subsidies received. On the uses side (intermediate consumption + final demand), prices are market prices; that is, including VAT, and transport and trade margins. In firm-level data, it is more likely that the prices paid are market prices. In addition, national accounts elaborate price indices that can be used to deflate the nominal values of transactions, while firm-level datasets often cannot distinguish between prices and quantities.

[18] This is roughly the ratio at the world level according to World Bank data (accessible here https://data.worldbank.org/indicator/NE.GDI.TOTL.ZS).

[19] Another issue is that different classification systems may be used. In the Ecuadorian dataset, the sectoral codes used in the national I-O tables differ from the ISIC classification codes used in the firm-level dataset. A one-to-one mapping from one classification system to the other is available only for the highest level of aggregation. For FactSet, the firm-level dataset and the I-O table use the same sectoral codes – i.e., ISIC Codes Rev. 4.

Timing of transactions. In principle, both national and financial accounts are compiled on an accrual basis; that is, they record "flows at the time economic value is created, transformed, exchanged, transferred or extinguished. This means that flows that imply a change of ownership are entered when the change occurs, services are recorded when provided, output at the time products are created and intermediate consumption when materials and supplies are being used." (United Nations, 2010, 3.166 p. 57). This can create substantial inconsistencies with firm-level datasets, particularly those created from direct money flows rather than from financial accounts, due to the prevalence of trade credit.

International Trade. Multinational firms typically file their accounts (and taxes) in various countries so that national accounts can ultimately try to separate the contributions of foreign firms domestically and domestic firms abroad. When using firm-level data, the ability to reconstruct tables close to national accounts would depend on the ability to access detailed financial accounts.

Taxes and government sector. In most countries, the public sector represents a large share of GDP. National accounts can represent this rather accurately, but it is more difficult for firm-level datasets, as the presence of transactions would depend on the legal nature of the entity and confidentiality issues (e.g., for defence spending).

Informal sector. In most countries, national accounts make an estimate of the value of the informal economy, which is unlikely to have a counterpart in network datasets.

A.4.2 Accounting identities in the WIOD

Table A.6 describes the variables and associated notation used.

Notation	Description
q_i	firm i's total sales
q_s	sector s' gross output
d_i	firm i's intermediate sales
d_s	sector s' intermediate sales
x_i	firm i's intermediate expenditure
x_s	sector s' intermediate expenditure
f_i	firm i's final demand
f_s	sector s' final demand
k_s	sector s' gross fixed capital formation
Δn_s	changes in inventories and valuables of sector s
w_i	firm i's labour costs
y_i	firm i's value-added
y_s	sector s' value-added
τ	taxes minus subsidies on products of sector s
e_s	CIF and FOB adjustments on exports of sector s
p_s	direct purchases abroad by residents and purchases on the domestic territory by non-residents for sector s
u_s	international transport margins of sector s
$Z_{r,s}$	the amount of intermediate inputs sector s buys from sector r

Table A.6: Notation and terminology.

Within a year, the WIOD table is at the country-sector level. Even though we aggregate over countries and sectors, it is helpful to understand the accounting identities underlying it. We show accounting identities by omitting the country and time index, but identities hold for country-sector tables and for one aggregated sector in the same way. Although for one sector, the inter-industry matrix becomes somewhat meaningless.

From the use side, the accounting identity used in the WIOD is

$$q_s = \sum_r Z_{sr} + f_s + k_s + \Delta n_s \,, \tag{A.7}$$

which states that gross output of sector s is the sum of intermediate sales of sector s, its final demand f_s, gross fixed capital formation k_s and changes in inventories Δn_s.

From the expenditure side, the accounting identity is

$$q_s = \sum_r Z_{rs} + \tau_s + e_s + p_s + y_s + u_s \ , \tag{A.8}$$

which states that gross output of sector s equals total expenditure on intermediate inputs, plus taxes minus subsidies τ_s, plus cost, insurance and freight (CIF) and free on board (FOB) adjustments on exports e_s, value-added y_s and international transport margins u_s.

To construct the proxy sector, the variable definitions in the sector-level I-O table need to match the definition of the variables at the firm level taken from firms' financial statements. We discuss how we harmonise the two datasets in Section A.4.3 and A.4.4.

A.4.3 Expenditure-side harmonisation

In this section, we define and describe how we harmonised two main variables that make up the I-O table from the expenditure side: intermediate expenditure and value-added. We do not account for direct purchases abroad by residents and purchases on the domestic territory by non-residents since these are zero for industries.

Expenditure on intermediate inputs. From the expenditure side, we define *sectoral intermediate expenditure* to include also CIF and FOB costs, and international transport margins since these are costs that if firms incur into are in the cost of goods sold:

$$x_s = \sum_r Z_{rs} + e_s + u_s \ . \tag{A.9}$$

Firms' intermediate consumption is defined as expenditure on intermediate inputs and services as stated in their invoices and reported as a cumulative figure in their income statements in the cost of goods sold. The cost of goods sold may include labour costs and thus needs to be cleaned for some firms depending on the method they use to write their income statement (see Appendix A.2.1 and A.2.2).

Value-added. Sectoral *gross value-added* is defined by the System of National Accounts (SNA) as labour and capital compensation, consumption of fixed capital and taxes net of subsidies (United Nations, 2010). The WIOD provides taxes net of subsidies as a separate variable, which thus needs to be added to the definition of value-added.[20] Sectoral gross value-added is given by

$$y_s^{\text{gross}} = \tau_s + y_s \ . \tag{A.10}$$

Firms do not disclose the subsidies they received as a separate line item; therefore, we did not include subsidies and simply added taxes (a definition of firms' value-added is provided in Appendix A.2.3).

A.4.4 Use-side harmonisation

In this section, we define and describe how we harmonise the four variables that make up the I-O table from the use side, namely final demand, intermediate sales, gross fixed capital formation and changes in inventories.

While it is more straightforward to reconcile prices on the expenditure side, it is less so on the use side. On the use side, prices are at the purchaser's price; that is, the amount paid by the purchaser

[20] The variable capital compensation that makes up value-added in the WIOD comprises both profit and consumption of fixed capital. It is a residual variable after subtracting labour compensation from value-added. Moreover "it is the gross compensation for capital, including profits and depreciation allowances. Because of its derivation as a residual, it reflects the remuneration for capital in the broadest sense. This does not include only traditional reproducible assets such as machinery and buildings but also includes non-reproducible assets. Examples are mineral resources and land, intangible assets (such as R&D knowledge stocks, software, databases, brand names and organisational capital) and financial capital" (Timmer et al., 2015, p. 601).

plus trade margins (wholesale and retail), transport margins (if invoiced by the producer) and non-deductible VAT minus deductible VAT. However, at the firm level output (i.e., revenues) is the amount received by the producer for the good or service sold minus VAT (deductible and non) and subsidies. While transport and trade margins may be in the firm's revenues, VAT and subsidies are not. We cannot take care of these mismatches because we lack the data to do so.

Final demand. Sectoral *final demand* is defined as in the WIOD and it comprises the consumption of households, the government and non-profit organisations.

Firms do not disclose sales to final demand nor sales to other firms; hence revenues include sales to other firms in the network and sales to final demand. To infer how much a firm sells to final demand, we use the share of final demand satisfied by the sector the firm is in. More formally, let q_s be gross output of sector s, f_s be final demand of sector s and q_i be firm i's revenues (or total sales), then final demand of firm i is given by

$$f_i = q_i \frac{f_s}{q_s} \ . \tag{A.11}$$

Gross fixed capital formation. Gross fixed capital formation (GFCF) is defined in the SNA as "the value of [...] acquisitions less disposals of fixed assets" (United Nations, 2010); it is a measure of investment. In the WIOD, GFCF corresponds to the part of a sector's output that ends up as an investment. It also includes some intangible assets when they are part of the SNA, but not all intangibles are covered.

In firm-level data, firms making capital goods disclose their customers and sales of capital goods are accounted for in the disclosed revenues. However, we cannot distinguish these types of transactions. Therefore, we apply a correction factor to a firm's revenues similar to Magerman et al. (2016) so that expenditure on intermediate inputs and intermediate sales have a consistent definition. As for final demand, we estimated a firm's GFCG as follows

$$k_i = q_i \frac{k_s}{q_s} \ , \tag{A.12}$$

where k is gross fixed capital formation either of firm i or sector s. A firm's expenditure on capital goods is not included in the cost of goods sold, thus no adjustment is needed on the expenditure side.

In the WIOD, the ratio k_s/q_s is negative for ISIC code E37-E39 " sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services". We assume that firms in this sector do not provide investment goods and hence have zero GFCF.

Intermediate sales. Sectoral *intermediate sales* are sales of intermediate inputs and services to other sectors in the economy. We use the same definition for firms, with the caveat that to get intermediate sales, we have to subtract from revenues the sales to final demand and GFCF.

Changes in inventories. We exclude changes in inventory since, in the WIOD, it is a column used for adjustments. Suppose we were to include changes in inventories. In that case, it should go with intermediate expenditure since the variable to which inventory changes correspond to at the firm level is the cost of goods sold, which is defined as beginning inventory minus ending inventory plus purchases during the period.

A.5 Sectoral codes

Table A.7 and A.8 describe, respectively, NACE Rev.2 codes at the section level (1-digit) and ISIC Rev. 4 codes at the 2-digit level.

ISIC code	Description
A01	Crop and animal production, hunting and related service activities
A02	Forestry and logging
A03	Fishing and aquaculture

B	Mining and quarrying
C10-C12	Manufacture of food products, beverages and tobacco products
C13-C15	Manufacture of textiles, wearing apparel and leather products
C16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
C17	Manufacture of paper and paper products
C18	Printing and reproduction of recorded media
C19	Manufacture of coke and refined petroleum products
C20	Manufacture of chemicals and chemical products
C21	Manufacture of basic pharmaceutical products and pharmaceutical preparations
C22	Manufacture of rubber and plastic products
C23	Manufacture of other non-metallic mineral products
C24	Manufacture of basic metals
C25	Manufacture of fabricated metal products, except machinery and equipment
C26	Manufacture of computer, electronic and optical products
C27	Manufacture of electrical equipment
C28	Manufacture of machinery and equipment n.e.c.
C29	Manufacture of motor vehicles, trailers and semi-trailers
C30	Manufacture of other transport equipment
C31-C32	Manufacture of furniture; other manufacturing
C33	Repair and installation of machinery and equipment
D35	Electricity, gas, steam and air conditioning supply
E36	Water collection, treatment and supply
E37-E39	Sewerage; waste collection, treatment and disposal activities; materials recovery; remediation activities and other waste management services
F	Construction
G45	Wholesale and retail trade and repair of motor vehicles and motorcycles
G46	Wholesale trade, except of motor vehicles and motorcycles
G47	Retail trade, except of motor vehicles and motorcycles
H49	Land transport and transport via pipelines
H50	Water transport
H51	Air transport
H52	Warehousing and support activities for transportation
H53	Postal and courier activities
I	Accommodation and food service activities
J58	Publishing activities
J59-J60	Motion picture, video and television programme production, sound recording and music publishing activities; programming and broadcasting activities
J61	Telecommunications
J62-J63	Computer programming, consultancy and related activities; information service activities
K64	Financial service activities, except insurance and pension funding
K65	Insurance, reinsurance and pension funding, except compulsory social security
K66	Activities auxiliary to financial services and insurance activities
L68	Real estate activities
M69-M70	Legal and accounting activities; activities of head offices; management consultancy activities
M71	Architectural and engineering activities; technical testing and analysis
M72	Scientific research and development
M73	Advertising and market research
M74-M75	Other professional, scientific and technical activities; veterinary activities
N	Administrative and support service activities
O84	Public administration and defence; compulsory social security
P85	Education
Q	Human health and social work activities
R-S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies

Table A.8: Description of ISIC Rev. 4 at the 2-digit level.

Section (1-digit)	Description	Divisions (2-digit)
A	Agriculture, forestry and fishing	01 – 03
B	Mining and quarrying	05 – 09
C	Manufacturing	10 – 33
D	Electricity, gas, steam and air conditioning supply	35
E	Water supply; sewerage, waste management and remediation activities	36 – 39
F	Construction	41 – 43
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	45 – 47
H	Transportation and storage	49 – 53
I	Accommodation and food service activities	55 – 56
J	Information and communication	58 – 63
K	Financial and insurance activities	64 – 66
L	Real estate activities	68
M	Professional, scientific and technical activities	69 – 75
N	Administrative and support service activities	77 – 82
O	Public administration and defence; compulsory social security	84
P	Education	85
Q	Human health and social work activities	86 – 88
R	Arts, entertainment and recreation	90 – 93
S	Other service activities	94 – 96
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	97 – 98
U	Activities of extraterritorial organisations and bodies	99

Table A.7: Description of NACE Rev.2 at the 1-digit (first column) and 2-digit level (last column).

B Derivations

B.1 The conditional maximum entropy reconstruction method of Parisi et al. (2020)

The conditional maximum entropy method developed by Parisi et al. (2020) reconstructs an ensemble of likely weighted networks that are consistent with prior information about the binary topology and that satisfy (in expectation) a set of constraints on observed network quantities, such as the in- and out-strengths. To reconstruct the ensemble, the method infers the probability distribution over the weighted network configurations by maximising the conditional (on the binary topology) Shannon entropy subject to the imposed constraints. Within this generalised procedure Parisi et al. (2020) examine two cases. One constrains the in- and out-strengths, the other consists of a two-step procedure. For reasons that are explained below, we chose the second case, which we explain in the remainder of this section.

In the first step a deterministic maximum entropy procedure, known as *MaxEnt* in the literature, is used to derive the weights. MaxEnt derives the weights by maximising an entropy-like function subject to constraints on the in- and out-strengths. The weights are then used in the second step, which maximised the conditional entropy of the weighted networks given the binary topology and subject to constraints on the expected value of each link weight. The MaxEnt values for the weights are used as constraints since MaxEnt has been found to be the best performing method in reconstructing weights (Anand et al., 2015; Lebacher et al., 2021).[21] This second case yields a computational benefit over the other case they examine because it requires solving m (the number of links) decoupled equations. Instead, the first case requires solving $2N$ coupled equations, where N is the number of nodes (see Parisi et al., 2020, for the derivation).

[21]Instead of the MaxEnt values, one could choose any value, including observed weights. The observed values then need to be subtracted from the in- and out-strengths.

First step. MaxEnt is a deterministic maximum entropy problem, which maximises an entropy-like function subject to constraints on the in- and out-strengths (firms' intermediate sales and expenditure) of each firm:

$$\begin{aligned} \underset{\{W_{ij}\}}{\text{maximise}} \quad & S(\boldsymbol{W}) = -\sum_{ij} W_{ij} \log W_{ij} \\ \text{subject to} \quad & \sum_j W_{ij} = s_i^{\text{out}^*} \;\; i = 1, \dots, N \\ & \sum_j W_{ji} = s_i^{\text{in}^*} \;\; i = 1, \dots, N \;, \end{aligned} \tag{B.1}$$

which yields

$$W_{ij}^{\text{ME}} = \frac{s_i^{\text{out}^*} s_j^{\text{in}^*}}{W^{\text{tot}^*}} \;, \tag{B.2}$$

where W_{ij} is the money flow from firm j to firm i, $s_i^{\text{out}^*}$ are the observed total intermediate sales (out-strength) of firm i, $s_i^{\text{in}^*}$ are the observed total intermediate costs (in-strength) of firm i and $W^{\text{tot}^*} = \sum_i s_i^{\text{out}^*} = \sum_i s_i^{\text{in}^*}$ is the total weight of the empirical network.

Second step. Let $\boldsymbol{W} \in \mathbb{W}$ be a weighted adjacency matrix, which is a realisation of the random variable $\mathcal{W}$. Similarly, let $\boldsymbol{A} \in \mathbb{A}$ be a binary adjacency matrix, which is a realisation of the random variable $\mathcal{A}$. The method maximises the conditional entropy $S(\mathcal{W} \mid \mathcal{A})$ defined over the probability density function of the weighted network configurations compatible with the binary topology and satisfying the constraints on the expected weights found by MaxEnt (Equation B.2):

$$\begin{aligned} \underset{\{Q(\boldsymbol{W}|\boldsymbol{A})\}}{\text{maximise}} \quad & S(\mathcal{W} \mid \mathcal{A}) = -\sum_{\boldsymbol{A} \in \mathbb{A}} P(\boldsymbol{A}) \int_{\mathbb{W}_{\boldsymbol{A}}} Q(\boldsymbol{W} \mid \boldsymbol{A}) \log Q(\boldsymbol{W} \mid \boldsymbol{A}) \mathrm{d}\boldsymbol{W} \\ \text{subject to} \quad & \langle W_{ij} \rangle = \sum_{\boldsymbol{A} \in \mathbb{A}} P(\boldsymbol{A}) \int_{\mathbb{W}_{\boldsymbol{A}}} W_{ij}(\boldsymbol{W}) Q(\boldsymbol{W} \mid \boldsymbol{A}) \mathrm{d}\boldsymbol{W} = W_{ij}^{\text{ME}} \;\; \forall (i,j) \in \mathcal{E} \\ & \int_{\mathbb{W}_{\boldsymbol{A}}} Q(\boldsymbol{W} \mid \boldsymbol{A}) \mathrm{d}\boldsymbol{W} = 1, \; \forall \boldsymbol{A} \in \mathbb{A} \;, \end{aligned} \tag{B.3}$$

where $\mathbb{W}_{\boldsymbol{A}}$ is the set of weighted configurations compatible with the binary topology, $Q(\boldsymbol{W} \mid \boldsymbol{A})$ is the conditional probability of generating a weighted network $\boldsymbol{W}$ consistent with adjacency matrix $\boldsymbol{A}$ and $\mathcal{E}$ is the edge set. The information on the binary network enters in probabilistic terms through $P(\boldsymbol{A})$, the probability of observing the binary topology $\boldsymbol{A}$. $P(\boldsymbol{A})$ could be estimated in a prior step using any suitable method. Given our global dataset, we use the observed topological structure $\mathcal{A} = \boldsymbol{A}^*$, thus $P(\boldsymbol{A}^*) = 1$ and $P(\boldsymbol{A}) = 0$, $\forall \boldsymbol{A} \in \mathbb{A}, \boldsymbol{A} \neq \boldsymbol{A}^*$.

Solving the above maximisation problem yields the Hamiltonian $H(\boldsymbol{W}) = \sum_{i \neq j} \lambda_{ij} W_{ij}$ and the graph probability $Q(\boldsymbol{W}|\boldsymbol{A}) = \prod_{i \neq j} Q_{ij}(W_{ij} \mid A_{ij})$ with

$$Q_{ij}(W_{ij} \mid A_{ij}) = \begin{cases} \lambda_{ij} e^{-\lambda_{ij} W_{ij}} & W_{ij} > 0 \text{ , if } \; A_{ij} = 1 \;, \\ 0 & \text{otherwise.} \end{cases} \tag{B.4}$$

Therefore, $Q_{ij}(W_{ij}|A_{ij} = 1)$ is of exponential form with parameter λ_{ij}, the Lagrange multiplier associated with the constraint on $\langle W_{ij} \rangle$. The generalised likelihood for the given set of constraints is

$$g(\boldsymbol{\lambda}) = -\sum_{i \neq j} \lambda_{ij} W_{ij}^{\text{ME}} + \sum_{i \neq j} \sum_{\boldsymbol{A} \in \mathbb{A}} P(\boldsymbol{A}) A_{ij} \log \lambda_{ij} \;, \tag{B.5}$$

with first order conditions given by

$$\langle W_{ij} \rangle = \frac{p_{ij}}{\lambda_{ij}} = W_{ij}^{\text{ME}}, \; \forall i \neq j, \tag{B.6}$$

where $p_{ij} = \sum_{\boldsymbol{A} \in \mathbb{A}} P(\boldsymbol{A}) A_{ij}$ is the probability that a link going from i to j exists. Substituting the MaxEnt prescription (Equation B.2) in the first order condition yields

$$\lambda_{ij}^* = p_{ij} \frac{W^{\text{tot}^*}}{s_i^{\text{out}^*} s_j^{\text{in}^*}} \;. \tag{B.7}$$

Since we consider a deterministic binary topology, the Lagrange multipliers simplify to

$$\lambda_{ij}^* = \frac{W^{\text{tot}^*}}{s_i^{\text{out}^*} s_j^{\text{in}^*}} \quad \text{for } (i,j) \in \mathcal{E} \ . \tag{B.8}$$

The MexEnt procedure assumes that the network is fully connected; however, the conditional maximum entropy method assumes no self-loops. Additionally, we know where the links are. Therefore, we need to redistribute the weights corresponding to $A_{ij} = 0, \forall (i,j) \notin \mathcal{E}$. To do so, the IPF algorithm is used. It redistributes the weights in an iterative procedure such that at the n-th iteration the weight is given by

$$W_{ij}^n = s_i^{\text{out}^*} \frac{W_{ij}^{(n-1)}}{\sum_{k \neq i} W_{ki}^{(n-1)}} \quad \text{and} \quad W_{ij}^{(n+1)} = s_j^{\text{in}^*} \frac{W_{ij}^n}{\sum_{k \neq j} W_{kj}^n} \ . \tag{B.9}$$

Confidence interval on the expected link weight. One of the benefits of Parisi et al.'s method is that one can calculate a confidence interval $[w^-, w^+]$ for each expected weight. The lower bound is given by

$$w^- = -\frac{\ln[e^{-1} + q^-]}{\lambda_{ij}^*} \ , \tag{B.10}$$

where q^- is a desired confidence level. The upper bound is given by

$$w^+ = -\frac{\ln[e^{-1} - q^+]}{\lambda_{ij}^*} \ . \tag{B.11}$$

We refer to Appendix E in Parisi et al. (2020) for the full derivation.

C Additional results

C.1 Weights

Ecuador. Figure C.1a shows the CCDF of the empirical weights of the full network (blue squares), the test network (green diamonds), the trimmed test networks (black dots) and that of the reconstructed networks (red triangles). All three weight distributions display heavy tails. Figure C.1b shows the power-law fit for one of the randomised networks. The distribution of the empirical weights has a power-law exponent of 1.1, while that of the reconstruction is higher and equal to 1.3. The cut-off point is also higher for the reconstructed weight distribution.

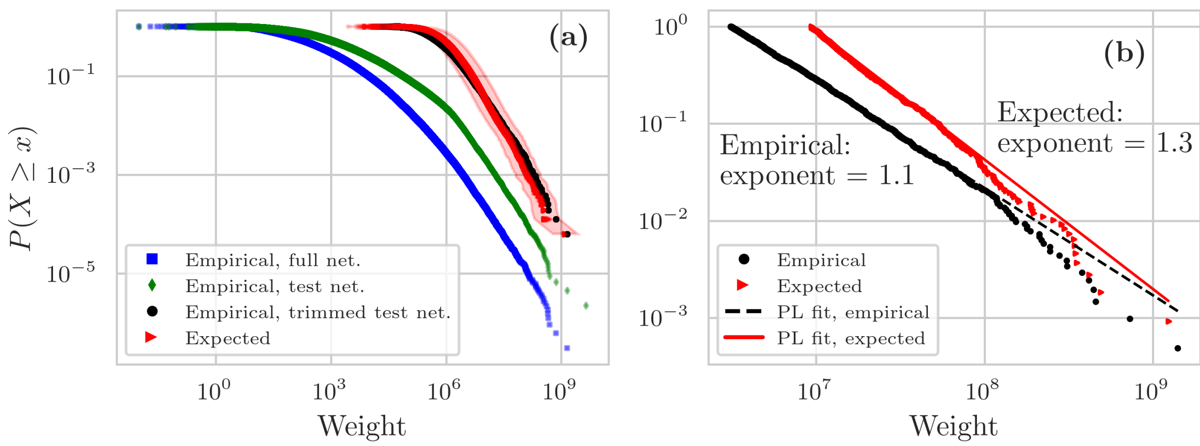

Figure C.1: **(a)** CCDF of the weights of the empirical full network (all firms in the Ecuadorian economy; blue squares), the empirical test network (green diamonds), the empirical trimmed test networks (black dots) and the reconstructed networks (red triangles); the shaded area indicates the 50% confidence interval. We show all 50 randomised test networks. **(b)** Power-law fit to the empirical (black dots) and expected (red triangles) weight distribution. The black dashed line shows the power-law fit to the empirical distribution, while the red solid line the power-law fit of the reconstructed distribution; we show results for one of the reconstructions.

The panels in the top row of Figure C.2 show the histograms of the relative prediction errors of Ecuador's weights for one of the randomised reconstructions. The top left panel shows the histogram of the non-positive error terms and the panel on the right the positive errors. The relative prediction error is defined as $\epsilon_i = (\boldsymbol{W}_{ij} - \boldsymbol{W}^*_{ij})/\boldsymbol{W}^*_{ij}$, where $\boldsymbol{W}^*_{ij}$ is the empirical value and $\boldsymbol{W}_{ij}$ is the reconstructed weight. The mean relative prediction error is 285.7%. Such a high mean is driven by a big outlier, which is 674,065%; the median is 40%.

The panels in the bottom row of Figure C.2 show the empirical weights on the x-axis and the relative prediction errors on the y-axis; we plot non-positive errors on the left and positive errors on the right. The reconstruction method consistently under-predicts weights with values above $\sim 10^8$.

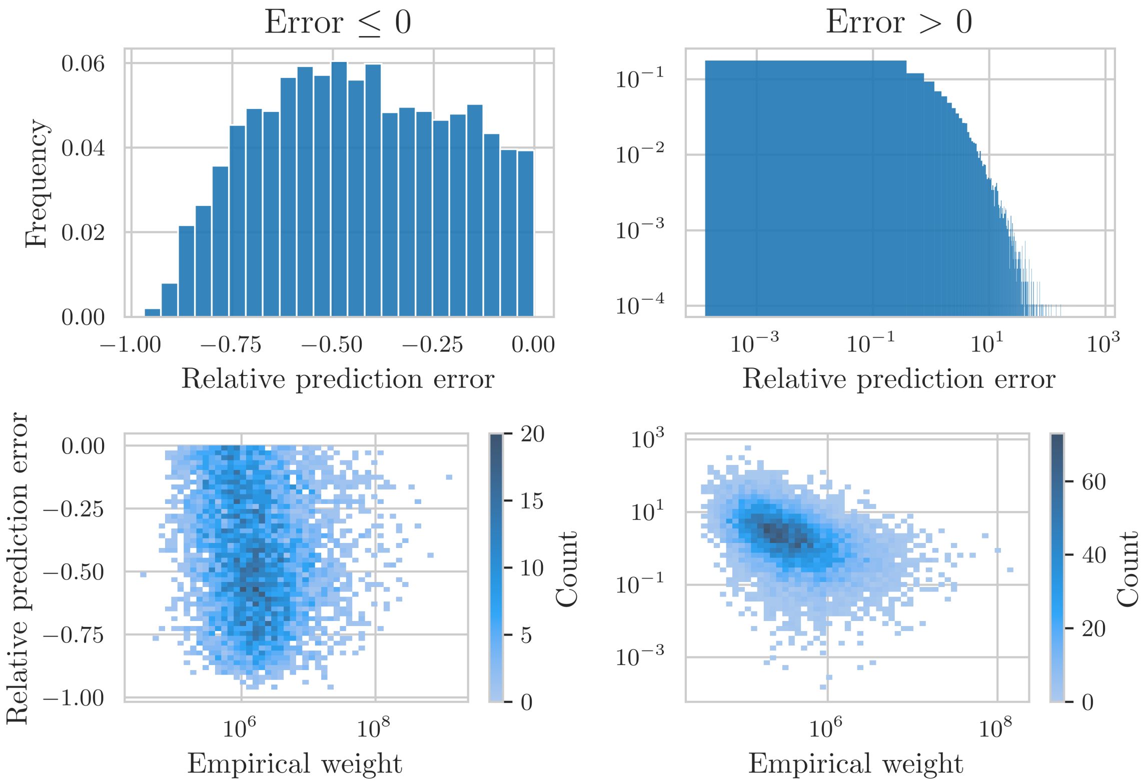

Figure C.2: **Top:** Histogram of the relative prediction errors of the weights for Ecuador. We plot separately the errors that are non-posite (left) and the errors that are positive (right). **Bottom:** Empirical weights (x-axis) against the relative prediction error (y-axis) for Ecuador. We plot the error terms that are non-positive on the left and those that are positive on the right. We divide both axes into 50 log-spaced bins and then count the number of data points falling in each square.

FactSet. Figure C.3 shows the CCDf of the expected weight distribution (blue squares) and its power-law fit (solid blue line) for FactSet. The estimated exponent equals 1.

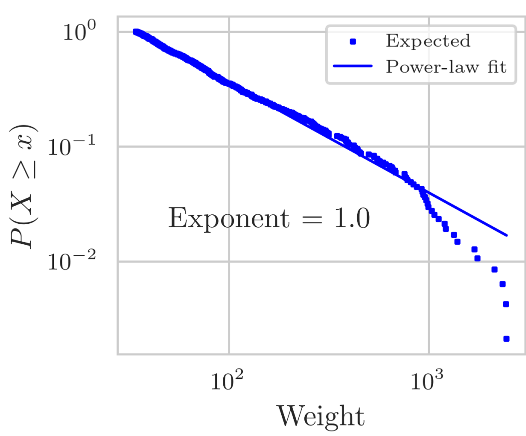

Figure C.3: CCDF of the expected weights (blue squares) and its power-law fit (solid blue line) for FactSet.

C.2 Technical and allocation coefficients

Ecuador. Figure C.4 shows the distribution of the empirical technical (left) and allocation coefficients (right) of the full network (blue squares), the test network (green diamonds), the trimmed test networks (black dots) and reconstructed networks (red triangles). The three distributions are right-skewed but do not display heavy tails.

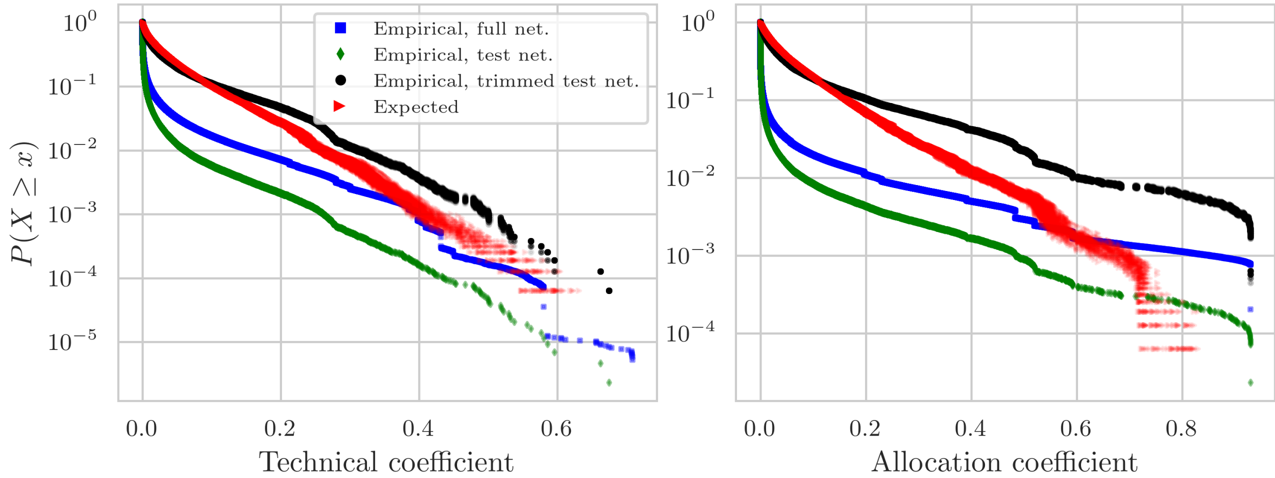

Figure C.4: CCDF of the technical (left) and allocation coefficients (right) of the empirical full network (blue squares), the empirical test network (green diamonds), the empirical trimmed test networks (black dots) and of the reconstructed networks (red triangles). We show all 50 randomised networks.

C.3 Multipliers

C.3.1 Influence vector

Ecuador. Figure C.5a shows the distribution of the empirical influence vector of the full network (blue squares), our test networks comprising 5,440 firms (black dots) and the reconstructed networks (red triangles). We compute the empirical influence vector for the firms in the (trimmed) test network using the network comprising all firms and then plot the distribution of only the 5,440 firms included in the (trimmed) test network. Therefore, the firms in the trimmed test network and the test network have the same influence. Figure C.5b shows the power-law fit for one of the reconstructions. The empirical distribution has a power-law exponent equal to 1.3, while the expected has a higher exponent, equal to 2. The CCDF of the expected influence vector also has a higher cut-off.

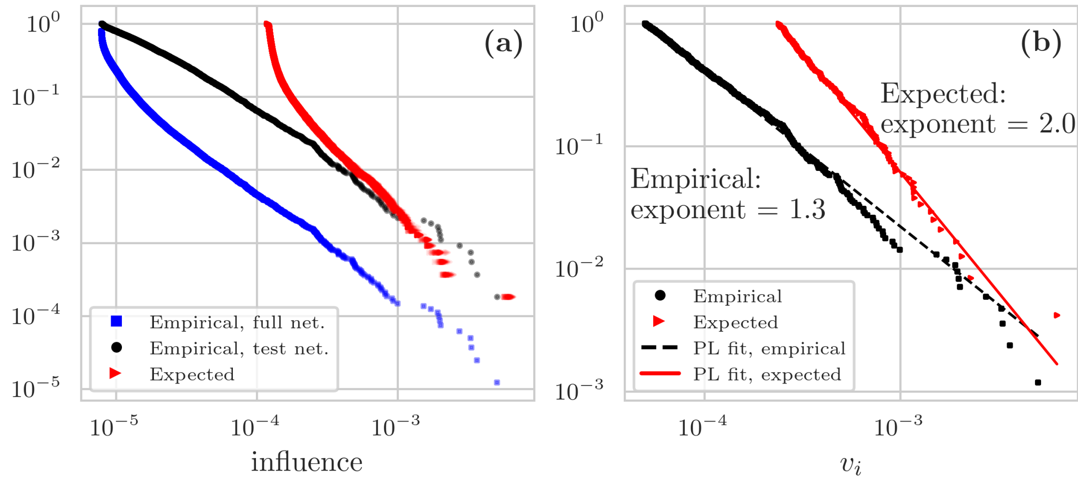

Figure C.5: **(a)** CCDF of the influence vector of the empirical full network (all firms in the Ecuadorian economy; blue squares), the empirical (trimmed) test network (black dots) and the reconstructed networks (red triangles). We show all 50 randomised test networks. **(b)** Power-law fit to the empirical (black dots) and expected (red triangles) influence vector distribution. The black dashed line shows the power-law fit to the empirical distribution, while the red solid line the power-law fit of the reconstructed distribution; we show results for one of the reconstructions.

FactSet. Figure C.6 shows the distribution of the expected influence vector (blue squares) and its power-law fit (solid blue line) for FactSet. The power-law exponent equals 1.9.

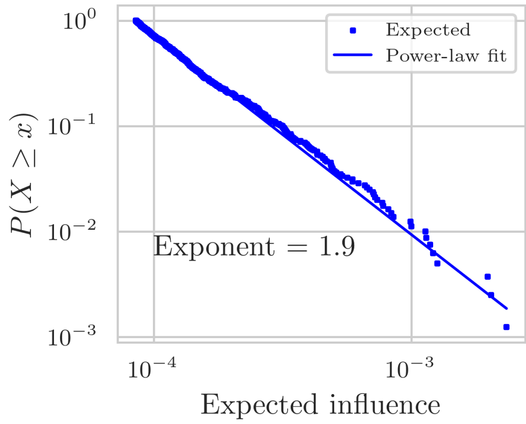

Figure C.6: CCDF of the expected influence vector (blue squares) and its power-law fit (solid blue line) for FactSet.

C.3.2 Prediction errors

Figure C.7a shows the histogram of the relative prediction errors for the output multipliers, while Figure C.8a for the influence vector, both are for Ecuador and for one of the 50 randomised networks. For the output multipliers, the mean relative prediction error is -0.7% and the median is -0.4%. For the influence vector, the mean relative prediction error is 735.1% and the median is 679.4%.

Figure C.7b shows the empirical output multipliers against the relative prediction errors for Ecuador, while Figure C.8b and C.8c for the influence vector. Figure C.8b shows observations for which the error term is non-positive, while Figure C.8c for observations for which the error term is positive.

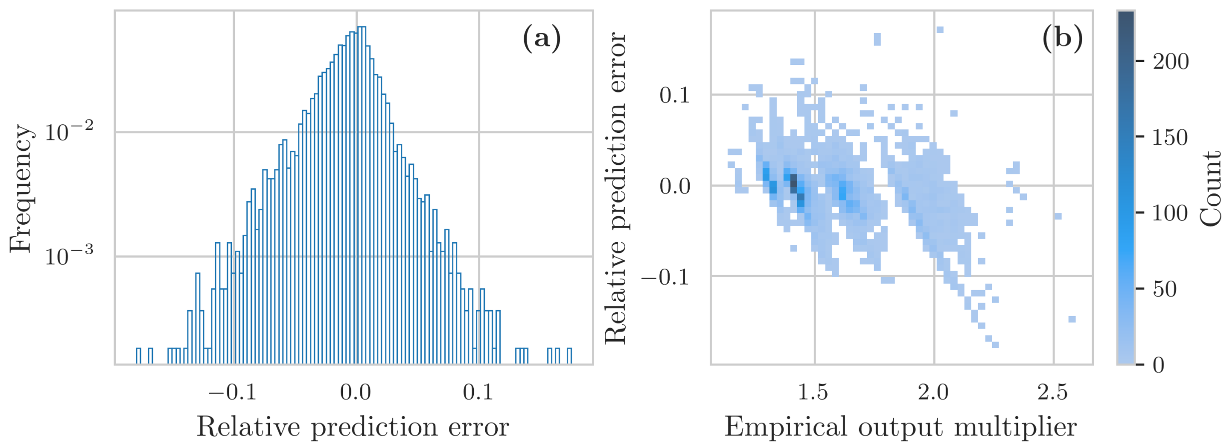

Figure C.7: **(a)** Histogram of the relative prediction errors of the output multipliers for Ecuador. **(b)** 2D histogram of the empirical output multipliers (x-axis) and the prediction errors (y-axis). We use 50 log-spaced bins for both axes and count the member of points falling into each square.

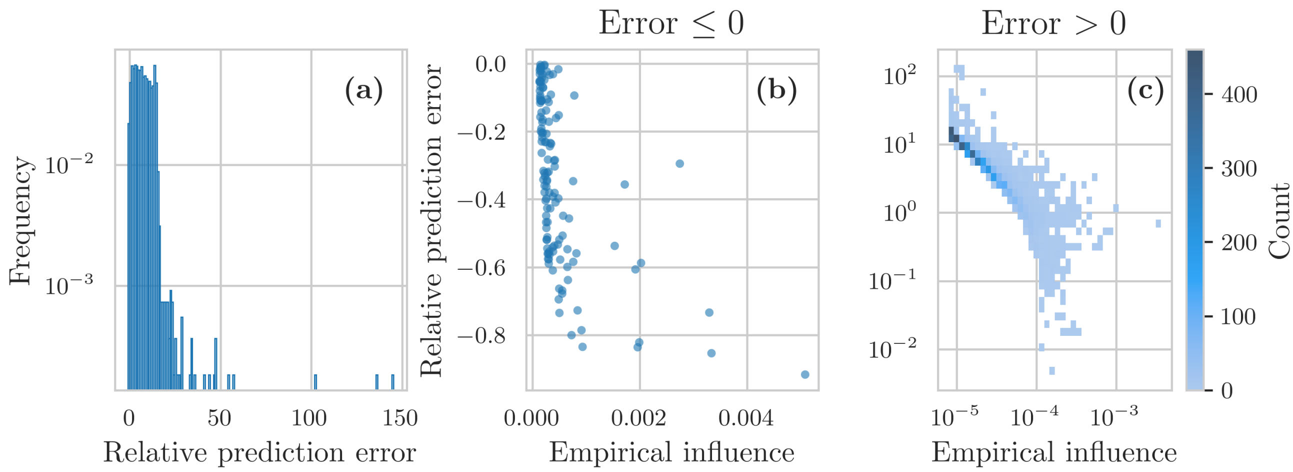

Figure C.8: **(a)** Histogram of the relative prediction errors of the influence vector for Ecuador. **(b)**, **(c)** 2D histogram of the empirical influence vector (x-axis) and the relative prediction errors (y-axis) for errors that are non-positive and positive, respectively. We use 50 log-spaced bins for both axes and count the number of points falling into each square.

C.3.3 Empirical and expected multipliers by sector

Output multipliers. Figure C.9 shows the empirical output multipliers on the x-axis against the expected output multipliers on the y-axis for Ecuador. Points are coloured based on the sector firms are in. Data points tend to organise in clusters, which form because the calculation of the technical coefficients requires knowing the value-added of each firm. Since we did not have access to firms' value-added, we used sector-level I-O tables to infer firms' value-added. The value-added of firm i is $y_i = \nu_s \cdot s_i^{\text{in}}$, where $\nu_s = y_s / s_s^{\text{in}}$ is the ratio of value-added and total intermediate expenses of sector s. We show ν_s in Figure C.9b, also colour-coded by sector.

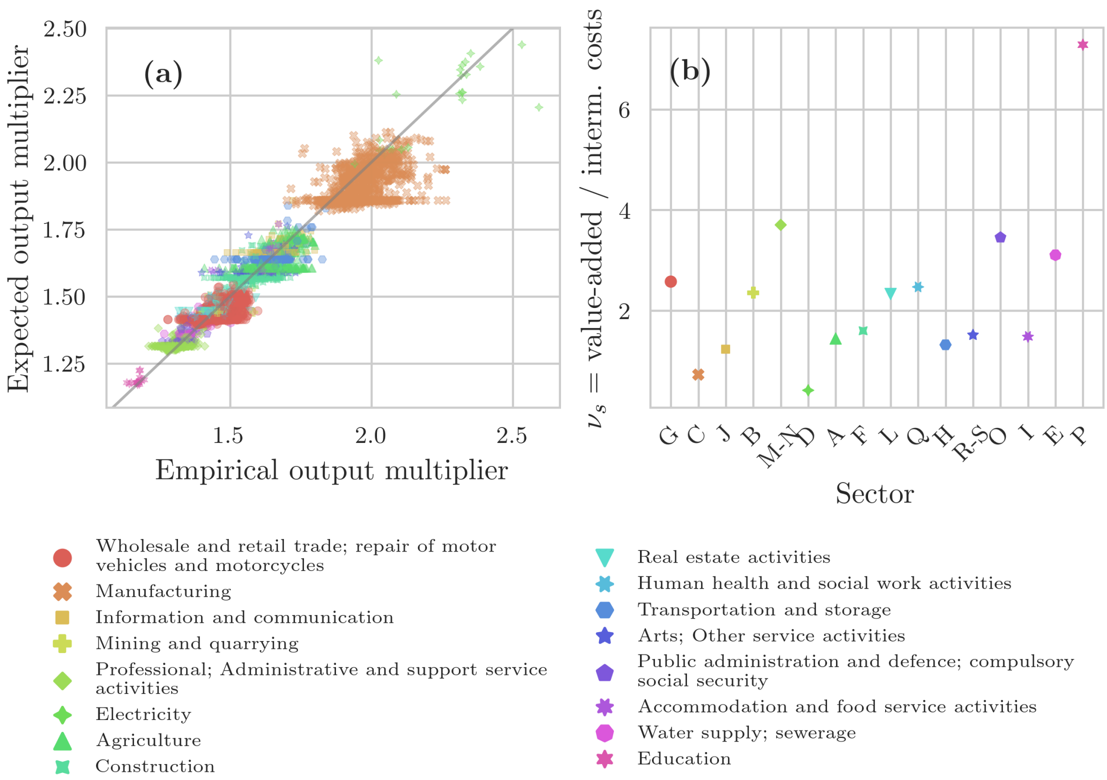

Figure C.9: **(a)** Empirical (x-axis) and expected (y-axis) output multipliers for Ecuador. Perfect prediction is achieved when all points lie on the 45-degree line (solid grey line). **(b)** $\nu_s = y_s/s_s^{\text{in}}$, i.e., the ratio of value-added and total intermediate expenses by sector in Ecuador's I-O table. Different colours and shapes correspond to the sectors in the I-O table in which firms are.

Influence vector. Figure C.10 shows the empirical influence vector on the x-axis against the expected influence vector on the y-axis for Ecuador. Points are coloured based on the sectors in the I-O table in which firms are.

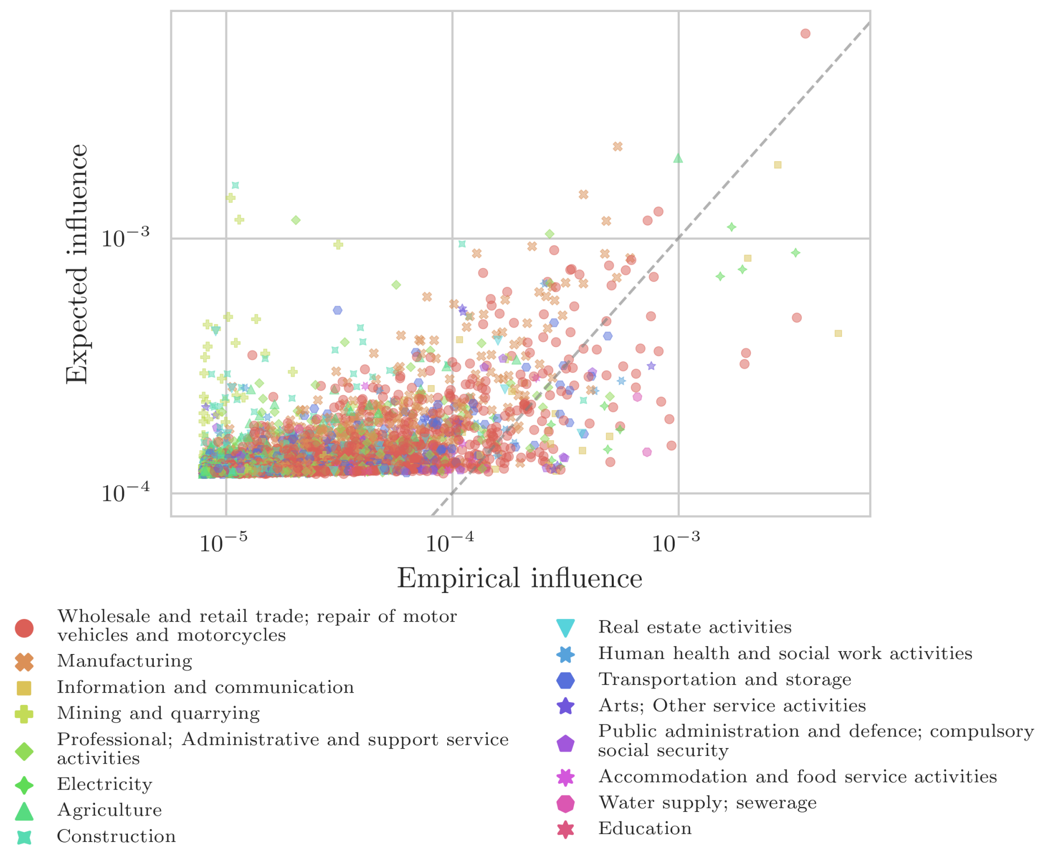

Figure C.10: Empirical (x-axis) and expected (y-axis) influence vector for Ecuador. Perfect prediction is achieved when all points lie on the 45-degree line (dashed grey line). The colours and shapes correspond to different sectors.

C.3.4 The effect of the proxy node on the centrality measures

Figure C.11a and Figure C.11d show the empirical multipliers against the reconstructed calculated including for all firms and the proxy node, while Figure C.11b and Figure C.11e exclude the proxy node from the calculations. Figure C.11c and Figure C.11f compare the expected multipliers that include the proxy node to those that exclude the proxy node from the calculations.

The influence vector is not that much affected by the inclusion or exclusions of the proxy node (Figure C.11d and Figure C.11e). Including the proxy node increases the influence of all nodes compared to not including it (Figure C.11f). While the error metrics are only slightly lower when the proxy node is excluded, the cosine similarity is slightly higher (0.56 vs 0.55) and the power-law exponent is slightly lower (2.0 vs 2.1). The mean and median are off for both, and we can only recover the variance if we include the proxy node, otherwise it is smaller. The proxy node is always the one with the highest influence.

On the contrary, the output multipliers are highly affected by the inclusion or exclusion of the proxy node (Figure C.11a and Figure C.11b). When the proxy node is excluded the output multipliers are severely underestimated (Figure C.11b). The proxy node is never the one with the highest centrality. When the proxy node is included, we can recover the mean, median and standard deviation, while when it is excluded, these are all underestimated (especially the mean and median).

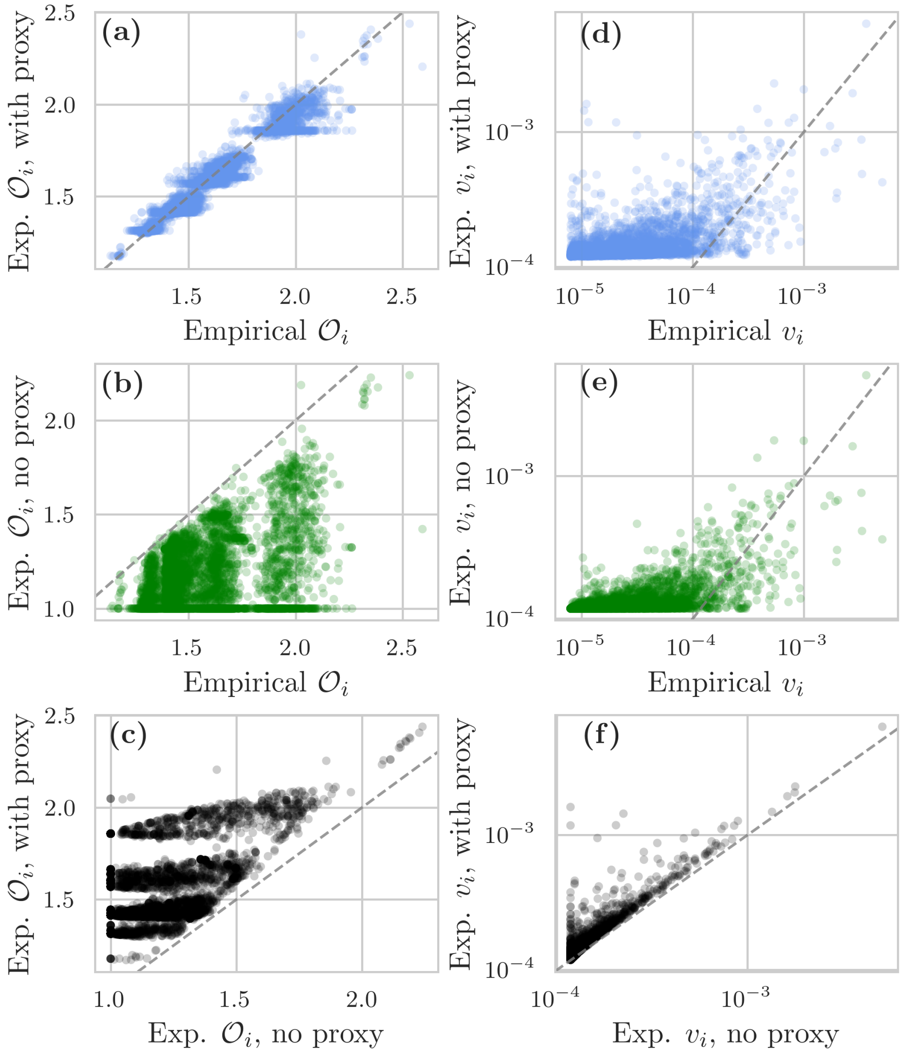

Figure C.11: **(a)**, **(d)** Empirical centrality measure (x-axis) and the expected calculated including the proxy node (y-axis) for the output multipliers ($\mathcal{O}_i$) and the influence vector (v_i), respectively. **(b)**, **(e)** Empirical centrality measure (x-axis) and the expected calculated excluding the proxy node (y-axis) for the output multipliers ($\mathcal{O}_i$) and the influence vector (v_i), respectively. **(c)**, **(f)** Expected centrality measure calculated excluding the proxy node (x-axis) and the expected centrality measure calculated including the proxy node (y-axis) for the output multipliers ($\mathcal{O}_i$) and the influence vector (v_i), respectively. The dashed grey line marks the identity line.

Before going into the details of why we observe these results, it is useful to define these two centrality measures again. The output multiplier is defined as

$$\mathcal{O} \equiv (\boldsymbol{I} - \boldsymbol{T}^{\top})^{-1}\mathbf{1} \ ,$$

or in scalar form

$$\mathcal{O}_j = \sum_i \frac{W_{ij}}{q_j}\mathcal{O}_i + 1 \ .$$

The output multipliers measure the centrality of a node by assigning a centrality score for each incoming edge – of course, these edges are weighted and normalised by the node's total costs, for the node for which we are calculating the multiplier. The influence vector is defined as

$$\boldsymbol{v} \equiv \frac{\alpha}{N}[\boldsymbol{I} - (1-\alpha)\boldsymbol{\Omega}]^{-1}\mathbf{1} \ ,$$

or in scalar form

$$v_i = \sum_j (1-\alpha)\frac{W_{ij}}{s_j^{in}}v_j + \frac{\alpha}{N} \ .$$

The influence vector assigns a centrality score to each node that the focal node points to. Thus, nodes that point to more nodes are more central – again, the edges are weighted and normalised by the total cost (in-strength) of the nodes to which the focal node points. In both centrality measures, a node may be highly influential because it has many links (incoming or outgoing) with nodes of low influence, or because it has few links (incoming or outgoing) with highly influential nodes. The output multiplier is based on incoming edges, while the influence vector is based on outgoing edges.

Since the only thing that changes in the calculations of multipliers is the inclusion or exclusion of the proxy node, let us focus on the proxy node's role:

$$\mathcal{O}_j \sim 1 + \frac{W_{pj}}{q_j}\mathcal{O}_p + \dots \ , \text{and}$$

$$v_i \sim \frac{\alpha}{N} + (1-\alpha)\frac{W_{ip}}{s_p^{in}}v_p + \dots \ ,$$

where p indexes the proxy node. Each link of the proxy node matters more in the output multipliers because (empirically) $W_{pj}/q_j > (1-\alpha)W_{ip}/s_p^{in}$ for all but the link the proxy node has with itself and with one another node.

C.3.5 Reconstructed and benchmark influence vector

Figure C.12 shows the "uniform" influence, which we use to benchmark aggregate volatility, against the reconstructed influence vector for the 50 randomised networks. It can be seen that the reconstructed influences tend to be bigger than those of the benchmark. However, the error metrics of the two influence vectors have very similar values, except for the cosine similarity, which is lower for the uniform influence vector (Table C.1).

Type	RMSE	MAE	MedAE	Cosine similarity
Influence vector	2×10^{-4}	1×10^{-4}	1×10^{-4}	0.560
	(9×10^{-7})	(1×10^{-7})	(5×10^{-8})	(3×10^{-3})
Uniform influence vector	2×10^{-4}	1×10^{-4}	1×10^{-4}	0.478
	(4×10^{-7})	(1×10^{-7})	(4×10^{-8})	(4×10^{-3})

Table C.1: Statistical indicators for the reconstructed and uniform influence vector for Ecuador. RMSE denotes the root mean squared error, MAE the mean absolute error and MedAE the median absolute error. For each metric, we show its mean value across the 50 randomised reconstructions. Below the mean value, the standard deviation in parenthesis. We excluded the proxy node from the calculations.

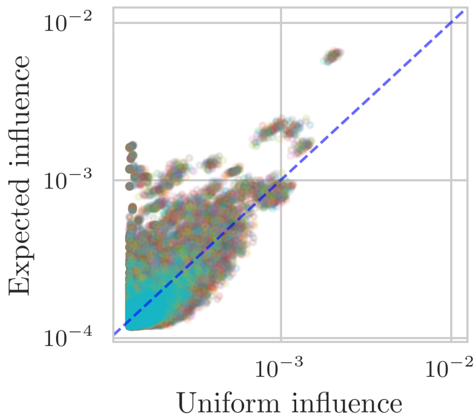

Figure C.12: "Uniform" influence on the x-axis and expected influence on the y-axis for the 50 randomised networks. The "uniform" influence is calculated by assuming that each firm buys inputs from its suppliers in the same proportion; this is the influence vector we use in our benchmark for aggregate fluctuations. The dashed blue line marks the identity line.

C.4 Different numbers of unknown links

C.4.1 Number of customers and suppliers

Figure C.13 shows the number of suppliers (in-degree) against the number of customers (out-degree) for different networks. The first panel on the top left-most corner shows the in- and out-degrees for the full network, the second panel for the test network and the consecutive panels for test networks with different numbers of unknown links. The last panel, at the bottom right-most corner, shows results for the trimmed test network that we discuss throughout the paper, which has the same average degree as FactSet. It can be noticed that in the full network firms tend to have more customers than suppliers, something also observed by Bacilieri et al. (2023). In creating our test network, firms lose more customers than suppliers. The maximum out-degree is 27,030 in the full network and decreases almost 7 folds in the test network (96% of missing links), reaching a value of 4,081, while the maximum in-degree is 2,543 in the full network and lowers to 781 in the test network (96% of missing links), which is a 3-fold decrease.

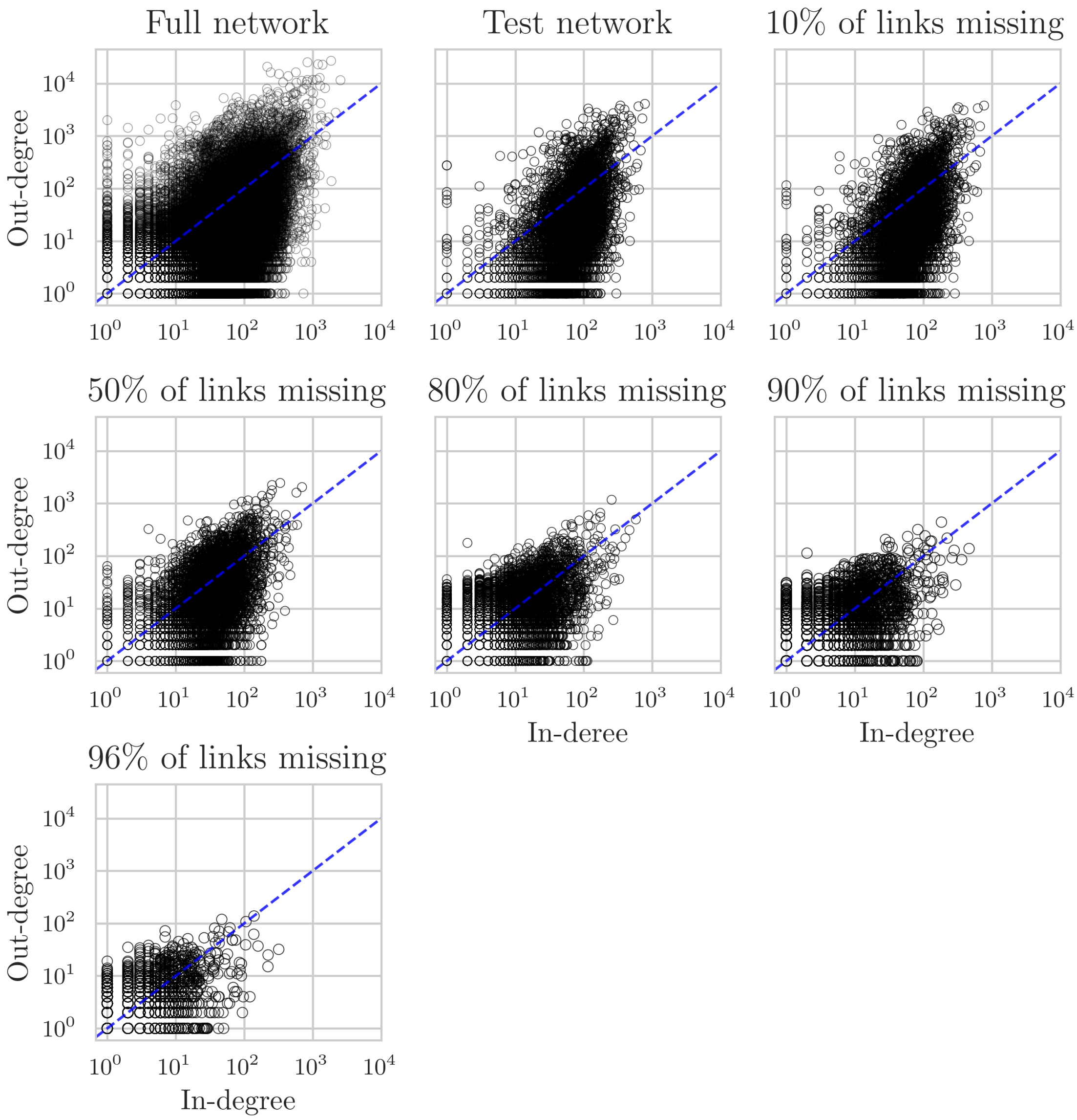

Figure C.13: Number of suppliers (in-degree) on x-axis and number of customers (out-degree) on y-axis for the full network, the test network and for test networks with different numbers of unknown links. The blue dashed line marks the identity line.

C.4.2 Comparison of empirical and expected quantities

Microscale quantities. Figure C.14 shows the empirical values against the expected values for the allocation coefficients (top panels), the technical coefficients (2^{nd} row), and the weights (bottom panels) for different numbers of unknown links. The allocation coefficients display a transition when the number of unknown links rises from 90% to 96%, something that the other two quantities do not experience. It also highlights the sampling process used to remove links: as the number of unknown links increases, more weights of higher values are removed. However, the shape of the joint density does not change, it is simply rescaled.

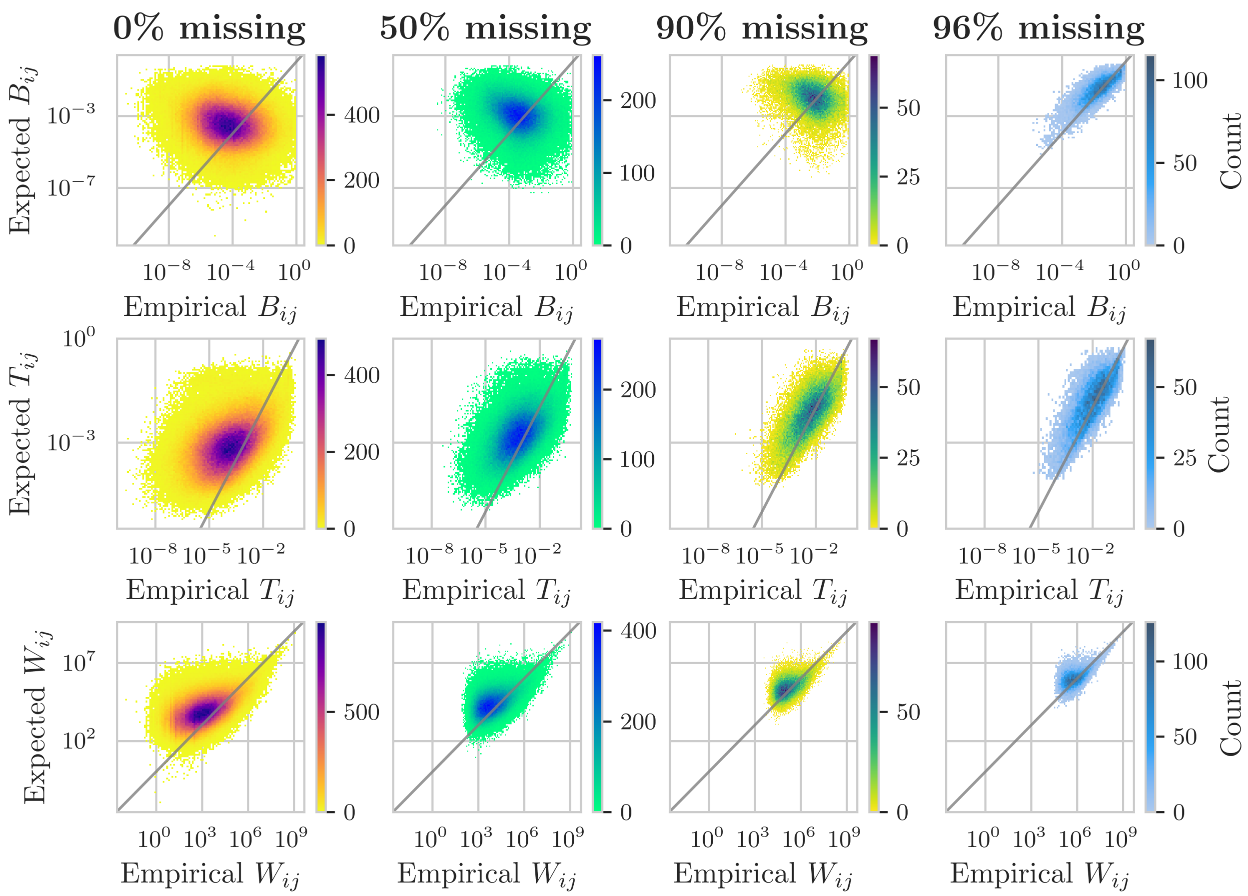

Figure C.14: 2D histograms for the empirical values on the x-axis and the expected values on the y-axis for different numbers of unknown links (0%, 50%, 90% and 96%) for the allocation coefficients (B_{ij}), the technical coefficients (T_{ij}) and the weights (W_{ij}). We bin each axis into 100 log-spaced bins and count the number of data points that fall in each square.

Higher-order quantities. Figure C.15 shows the empirical values against the expected values for the output multipliers (left column) and the influence vector (right column) for different numbers of unknown links (0%, 20%, 50%, 90% and 96%). Figure C.16 shows the empirical CCDF (black dots) for the output multipliers (left) and the influence vector (right). It also shows the CCDF of the reconstructed multipliers for different numbers of unknown links and for the 50 randomised networks.

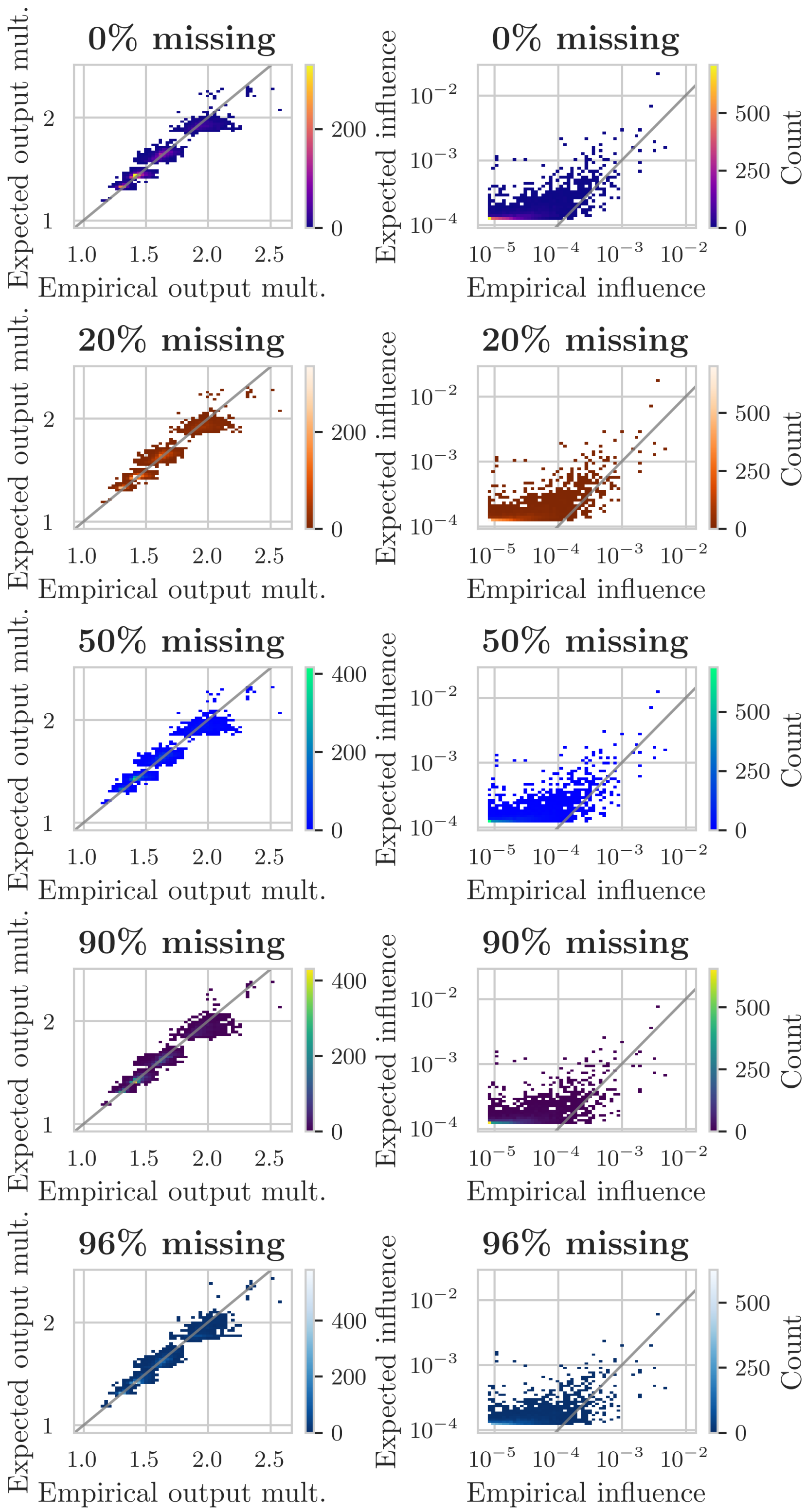

Figure C.15: 2D histograms for the empirical values on the x-axis and the expected values on the y-axis for different numbers of unknown links (0%, 20%, 50%, 90% and 96%) for the output multipliers (left column) and the influence vector (right column). We bin each axis into 50 log-spaced bins and count the number of data points that fall in each square.

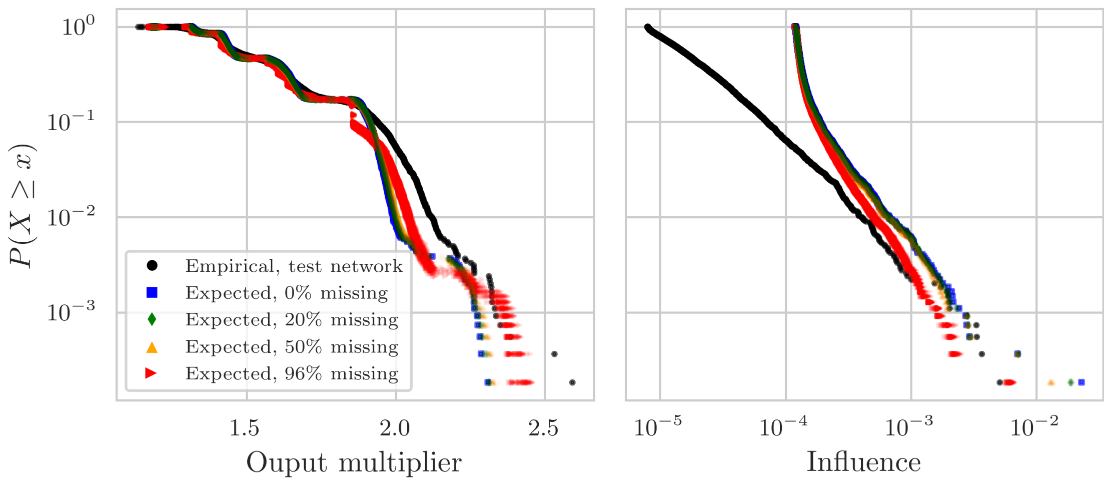

Figure C.16: CCDF of the output multipliers (left) and influence vector (right) for different numbers of unknown links (0% blue squares, 20% green diamonds, 50% yellow triangles and 96% red triangles) and the empirical (black dots).