Know what?
New lenses on productive knowledge shed light on long run development, structural change, job switching and the transition to the green economy



Penny Mealy

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2018

School of Geography and the Environment
Institute for New Economic Thinking at the Oxford Martin School
Smith School of Enterprise and the Environment
St Edmund Hall, Oxford University

This page has been intentionally left blank

...when taken as items in a process of cumulative change or as items in the scheme of life, these productive goods are facts of human knowledge, skill, and predilection; that is to say, they are, substantially, prevalent habits of thought, and it is as such that they enter into the process of industrial development.

Thorstein Veblen, 1898

This page has been intentionally left blank

Abstract

Tracing back to Adam Smith, scholars have emphasised the importance of knowledge and its profound implications for prosperity. However, measuring and analysing something as intangible and multifaceted as the knowledge collectively held and used within society has, not surprisingly, proved challenging.

This thesis advances quantitative empirical approaches for analysing productive knowledge within socio-economic systems and applies these methodologies to shed new light on (i) long run regional development and structural change (ii) labour market segregation and job switching, and (iii) the transition to the green economy.

This thesis shows that across a range of different contexts, from US regional development in the 19th century to modern-day labour markets, *what places and people know matters*. Measuring the type of knowledge concentrated in US states sheds new light on structural change, the decline of rust-belt states, and the reversal of patterns in income convergence. Similarly, measuring the type of knowledge held by individual workers in different occupations provides new insights on the division of labour, job transitions and future career prospects. Moreover, measuring the type of knowledge embedded within countries turns out to be particularly informative of a nation's likelihood of success in the green economy – and its potential for green diversification in the future.

This thesis also reveals *distinct segregation in societal knowledge across places and people*. Such divisions could matter more than we realise. For example, the segregation of productive knowledge across US states could play a role in undermining the nation's capacity for inclusive growth. The segregation of skills and productive capabilities across the US labour force could exacerbate the distributional consequences of labour market shocks. And moreover, the segregation across nations' capacities to both participate and benefit from the transition to the green economy could slow or even create barriers to progressing one of the most important socio-economic transitions of our time. While this thesis has advanced new ways to identify these divisions in knowledge, further work is required to better understand their implications – and what can be done.

Finally, this thesis suggests that *future improvements in what places and people know will need to become less incremental and more transformational*, if we are to successfully address the 'twin challenges' of poverty alleviation and climate change in the 21st century. Learning how to both navigate and drive the process of transformational change in productive capabilities could be the type of knowledge that matters more than anything else.

Contents

1 Introduction **11**

1.1 Theoretical framework: What do we mean by productive knowledge? 14

1.2 Methodological framework: Two lenses for examining productive knowledge 16

1.3 Overview of thesis and key findings 21

2 Literature Review **27**

2.1 The economic roots of productive knowledge and prosperity 27

2.2 The study of productive knowledge evolves - outside mainstream economics 29

2.3 Knowledge spillovers and agglomeration externalities 31

2.4 The relatedness of productive knowledge - and the Product Space . 32

2.5 Getting personal about productive knowledge 35

2.6 Taking productive knowledge to task 36

2.7 Productive knowledge over the long run 38

2.8 The complexity of productive knowledge 39

2.9 Productive knowledge and the green economy 40

2.10 The need to better understand transformational change in productive knowledge 41

3 A New Interpretation of the Economic Complexity Index **43**

3.1 Introduction 44

3.2 The Economic Complexity Index 46

3.3 Results 48

3.4 Discussion 63
3.5 Methods 65
3.6 Appendix 68

4 Determining the differences that matter: New insights on development, dynamics and divergence in US states over 1850-2010 72

4.1 Introduction 73
4.2 Understanding the ECI and PCI metrics as dimension reduction tools 77
4.3 Application to US States: 1850-2010 81
4.4 Explaining differences in economic prosperity 83
4.5 What separates rich and poor states? 91
4.6 New insights into shifting economic landscapes 96
4.7 Conclusion 103
4.8 Appendix 106

5 What you do at work matters: New lenses on labour 123

5.1 Introduction 124
5.2 Results 126
5.3 Discussion 143
5.4 Methods 144
5.5 Appendix 150

6 Economic Complexity and the Green Economy 159

6.1 Introduction 160
6.2 Green trade data and empirical patterns 165
6.3 Economic complexity and the green economy 174

6.4 The green product space and future green industrial transitions . . 186
6.5 Policy implications 198
6.6 Conclusion 201
6.7 Appendix 203

7 Transformational change: Parallels for addressing climate and development goals 219

7.1 Introduction 219
7.2 Transformational change in climate and development contexts . . . 222
7.3 Empirical patterns of transformational change in climate and development contexts 227
7.4 Methodological tools for modeling and understanding transformational change in climate and development contexts 236
7.5 Conclusion 248

8 Conclusion 250

8.1 Through the ECI-PCI looking glass 251
8.2 "Same-same but different" (and important) applications of knowledge similarity measures 254
8.3 Broader contributions to the literature and the study of productive knowledge 257

References 261

Acknowledgements

The process of advancing knowledge is a collective endeavour - and this thesis is certainly no exception. It would not have been possible without the insights and encouragement of a wonderful network of insanely talented and supportive people.

I would like to deeply thank my supervisors, Doyne Farmer and Cameron Hepburn. You both consistently demonstrate what it is to be a first rate scientist, a first rate economist, and most importantly, a first rate person. Learning from you has been a tremendous privilege, and I have enormous respect and admiration for your willingness to go into bat for ideas that matter.

I would also like to thank my fabulous co-authors:

To Ricardo Hausmann, you are the giant whose shoulders this thesis stands upon. It's been an honour - and an adventure (!) - to build on your work, and I have benefited enormously from your passion, insight and determination.

To Alex Teytelboym, thank you for your unfailing support, and the countless discussions, debates and disagreements - I learned and laughed so much, and loved every minute.

To Maria del Rio-Chanona, your formidable intellect and hilarious wit made research such a joy - I am immensely grateful for your insights, encouragement and, of course, your Mexican tequila.

A special thank you must also go to Simon Angus. Oliver Wendell Holmes Jr once said that *"one's mind, once stretched by a new idea, never regains its original dimensions"*. You have been instrumental in stretching both my mind and my life far beyond its original dimensions, and I continue to be incredibly appreciative of your ongoing support.

To the fabulous group of researchers at INET, thank you for your incredibly helpful feedback, encouragement and, in particular - your friendship. It has been such a privilege to work alongside you, learn from you, drink beer with you, and share all the inevitable ups and downs of the DPhil journey.

Also sincere thanks to the wonderful and welcoming research group at the Centre for International Development at the Harvard Kennedy School. It was an honour to share the work in this thesis with you, and I am particularly appreciative of your kind feedback and comments.

Finally, to Paul - there are no words. You have made all the difference in the world. And to my infinitely supportive parents and family, you can stop asking - *I'm finally finished.* But I wouldn't have got there without you.

This page has been intentionally left blank

1 Introduction

What lies at the heart of economic prosperity? Classical economists such as Adam Smith and Friedrich Hayek have long pointed to the nature of knowledge, and the way it is collectively held and used within society (Smith, 1776, Hayek, 1945). However, as societal knowledge is difficult to empirically measure, economic analysis has traditionally focused on aggregate production functions encompassing more tangible entities, such as capital and labour. In these frameworks, knowledge is commonly treated fairly simplistically as an imputed residual – a 'measure of our ignorance' (Solow, 1956, 1957, Mankiw et al., 1992, Abramovitz, 1993).

While such approaches have a central place in economic analysis and underpin much of classical economic growth and development theory, they unfortunately obscure the complex web of differentiated, interdependent productive capabilities that economists like Smith and Hayek believed to be so fundamental to productive socio-economic systems. The analytical treatment of knowledge (often as a single factor with an exogenous or endogenous improvement rate) also limits analysis into a number of important policy questions. For example, without recognising the type of skills and productive capabilities that characterise richer places relative to poorer places, policy makers seeking to advance living standards in poorer countries or regions may be less equipped to take meaningful steps towards alleviating poverty and closing income gaps. Similarly, without appreciating the tapestry of specialised knowledge underpinning the labour force, policy analysis of employment impacts may miss important labour market characteristics, such as segregated patterns in the division of capabilities across society. Such divides could generate severe inequality if not appropriately addressed. Moreover, without understanding the

characteristics of productive knowledge that will become increasingly important in a greener, more environmentally friendly economy, it is difficult for people, places and policy makers to make the necessary investments to secure a future that is both prosperous and sustainable.

In recent years, a number of innovative, empirical approaches have been advanced in the economic geography and economic complexity literature to better understand the nature of knowledge and the extent to which it is embedded within countries and regions. Rather than focusing on capital and labour aggregates, these approaches instead start from an alternative assumption: *what a place produces reveals a lot about what it knows.* By analysing detailed trade, industry and patent data, these approaches have made valuable progress towards better understanding the specificity (or type) and complexity (or sophistication) of production knowledge existing within different societies.

This thesis builds on these approaches, aiming to further advance a quantitative empirical framework for analysing productive knowledge within socio-economic systems. It firstly refines and extends the *economic complexity* methodology, which was originally introduced to infer information about countries' productive capabilities by making relative comparisons across their export baskets (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). This thesis addresses the lack of clarity about what the product-based Product Complexity Index (PCI) and country-based Economic Complexity Index (ECI) capture, and provides a precise mathematical interpretation of how these measures work. It then applies similar methods to shed new light on research and policy questions in four key areas.

First, this thesis looks at productive knowledge underpinning long run economic

development. Specifically, it examines the regional economic development of US states from 1850-2010. By examining US census data on state employment concentrations in detailed industries and occupation categories, it investigates the type of productive knowledge and capabilities that characterise richer and poorer regions over 160 years of development history. It also explores how the type of productive knowledge underpinning prosperous places has evolved over time and considers the possible implications that diverging regional knowledge specialisations may have for income convergence and inclusive growth.

Second, this thesis studies the web of differentiated productive capabilities characterising the US labour force. By measuring the similarity in specific tasks that people working in different occupations undertake, it explores the network structure of professional knowledge. Using this networks-based lens, this thesis investigates divisions of labour with respect to a number of characteristics, such as wages and education, gender and susceptibility to automation. It also examines the usefulness of occupational similarity measures for predicting job transitions.

Third, this thesis examines the productive knowledge and capabilities relevant to the green economy. By collating and analysing a new dataset of 'green' (or environmentally friendly) exports across countries, it seeks to shed light on two questions: (i) which countries currently have the productive capabilities to be successful in the green economy and (ii) how might countries re-orient their present set of capabilities to become more competitive in green products and industries.

Fourth, this thesis explores the importance of transformational change in the overlapping contexts of climate and development policy. Recognising that addressing climate and development goals in the 21st century requires *large leaps* in produc-

tive knowledge, rather than small steps, this thesis considers the relative merits of current frameworks and methodological tools for analysing the transformational change process. It also identifies possible synergies across climate and development fields and highlights key research and policy-related advantages from collaboration.

In the remainder of this introductory chapter, I initially outline the underlying conceptual framework that this thesis draws on for theorising about productive knowledge. I then discuss the methodological framework that this dissertation extends and builds upon. Finally, I provide an overview of this thesis' structure, objectives and key findings.

1.1 Theoretical framework: What do we mean by productive knowledge?

Although the conceptual underpinnings of knowledge are inherently elusive (Clark, 2018), its profound effects on society are easy to see. Heralded as the source of economic growth (Kuznets, 1973), the accumulation of knowledge has been held largely responsible for the remarkable advances in living standards experienced by many countries in the last two centuries (Solow, 1956, Mankiw et al., 1992, Romer, 1993, Jones, 2005). However, as noted by Mokyr (2002), the differences in prosperity between modern society and medieval peasants are not due to the fact that we are any smarter or more intelligent than our ancestors. There is very little to suggest that we are. Instead, "the central phenomenon of the modern age is that as an *aggregate* we know more" (Mokyr, 2002, p. 8).

This *collective* use and accumulation of productive knowledge is the primary con-

cern of this thesis. And as this phenomenon is both unperceivable and impenetrable in the aggregate production functions that form the mainstay of neoclassical economic analysis, this thesis seeks to advance a different lens. Like Galilleo's telescope, the hope is that with a little tool-based tinkering, and a little refocusing of the data resolution, it may be possible to gain new insights into the way society distributes and coordinates differentiated knowledge across individuals to serve productive purposes.

Numerous frameworks have been proposed to meaningfully organise the way we think about knowledge. In *Nichomachean Ethics*, Aristotle distinguishes between three types of knowledge: (i) *episteme*, which relates to scientific knowledge (or knowing 'what'), (ii) *techne*, which relates to human skills and crafts (or knowing 'how') and *phronesis*, which relates to ethics and practical wisdom (or knowing 'the right thing to do'). Economists have tended to be more preoccupied with the first two types of knowledge, consigning the study of phronesis to the realm of philosophy and theology. Although the work in this dissertation follows suit, it is worth mentioning that this narrowed perspective on knowledge may become increasingly less appropriate if technological progress continues to outrun progress on moral and institutional fronts. Indeed, prosperity—and the fate of humanity—may hang in the balance.

A different distinction relating to the nature of knowledge was highlighted in Michael Polanyi's classic phrase: '*we know more than we can tell'* (Polanyi, 1967, p 4). The limitations in codifying 'tacit' knowledge tends to make its diffusion much more challenging. Unlike 'explicit' knowledge which can be easily written down, tacit knowledge needs to be learned and appropriated 'by doing' (Arrow,

1962).

Building on these conceptual groundings, Hausmann (2016) recently proposed a related framework. He distinguished between *tools*, which represent knowledge embodied in technological artefacts, *recipes or blueprints*, which represent knowledge that can be codified, and *know-how*, which relates to the tacit knowledge that exists only in people's brains. Know-how, he argues, is often the binding constraint on the development process because it is more difficult to transfer and embed across the globe. Most of the analysis undertaken in this dissertation is broadly aligned with Hausmann (2016)'s conceptual framework, and will primarily aim to shed new empirical light on know-how in different settings.

1.2 Methodological framework: Two lenses for examining productive knowledge

Measuring something as intangible and multifaceted as the complex web of productive knowledge stored in people's brains has, not surprisingly, proved difficult. However, using detailed data on what places produce, scholars have recently advanced two important analytical lenses that have provided new, empirical insights into the productive knowledge underpinning different countries and regions. I discuss each of these in turn.

1.2.1 Knowledge complexity (or sophistication)

The *complexity* (or *sophistication*) of a place's productive knowledge has been posited to have important implications for its prosperity and future growth rates.

For example, in his theory of technological capabilities, Lall (1992, 2000) suggested that countries which developed more sophisticated technological know-how were more likely to experience rapid growth rates. Similar hypotheses have been proposed by Hausmann et al. (2007), Hidalgo and Hausmann (2009) and Hausmann et al. (2014). To test these conjectures, number of different approaches have been proposed to empirically measure the complexity (or sophistication) of countries' productive knowledge by studying detailed data on what they are able to export (Lall, 2000, Lall et al., 2006, Hidalgo and Hausmann, 2009, Tacchella et al., 2012, Hausmann et al., 2014).

This thesis is primarily focused on the approach approach advanced by Hidalgo and Hausmann (2009) and Hausmann et al. (2014). On the basis of the theory that countries become prosperous by developing complex capabilities that allow them to export a diverse range of products few other countries are able to export, Hidalgo and Hausmann (2009) introduced a recursive algorithm called the *Method of Reflections*. This approach involves measuring country *diversity* (the number of products a country can export competitively[1] and product *ubiquity* (the number of countries that can competitively export a product) and iteratively 'correcting' a country's diversity by the ubiquity of its products. While the exact solution to this recursive algorithm was later shown to be found by solving an eigenvalue equation (Caldarelli et al., 2012, Cristelli et al., 2013, Hausmann et al., 2014), the resulting product-based Product Complexity Index (PCI) and country-based Economic Complexity Index (ECI) proved to be particularly successful in explaining cross-country variation in per capita GDP and predicting future growth rates.

[1] where a country is considered to be competitive in an export if its revealed comparative advantage (as measured by the Balassa index (Balassa, 1965)) in that export is greater than 1.

1.2.2 Knowledge specificity and similarity

The *specificity* of a place's productive knowledge - and its *similarity* to other types of knowledge, has proven to have important implications for the directionality of development and structural change. Just as it is easier to learn how to manufacture trucks if you already know how to manufacture tractors, scholars have found that countries and regions are more likely to diversify into products or industries that require production knowledge similar to what they current possess.

The similarity of productive knowledge (often referred to as *relatedness*) underpinning different industries has been measured using a number of different approaches. Hidalgo et al. (2007) measured the relatedness between exported products by examining their probability of being co-exported within different countries. Boschma et al. (2013) applied a similar approach to investigate regional diversification in Spain. Alternative strategies have examined the flow of workers (or skill-relatedness) between industries (Neffke and Henning, 2013, O'Clery et al., 2016, Neffke et al., 2017) and firms (Guerrero and Axtell, 2013), and also looked into the strength input-output linkages across industries (Essletzbichler, 2015). Research has also explored the relatedness of productive knowledge underpinning different technologies by investigating patent citations (Leten et al., 2007, Rigby, 2015) and the co-classification of patents across different technology classes (Kogler et al., 2013, 2017).

A key novelty introduced by Hidalgo et al. (2007) was the use of network analysis to examine and visualise relatedness of knowledge underpinning products and industries. Specifically, Hidalgo et al. (2007) constructed the Product Space, which

is a network where exported products are represented as nodes linked to each other if they are more closely related. The Product Space proved to be a particularly useful representation for generating new insights into countries' productive structures and constrained development pathways. And indeed, in addition to building on the wider body of literature describing the path-depend nature of economic development (David, 1985, Arthur, 1989, Krugman, 1991a,b,c, Arthur, 1994, David, 1994, Aghion et al., 2014, 2016), these networks-based approaches have offered researchers and policy-makers new frameworks to both analyse and potentially catalyze industrial development in particular directions (Boschma and Gianelle, 2013, Thissen et al., 2013, Balland et al., 2018).

1.2.3 Addressing limitations and opportunities within the existing methodological framework

These empirical methodologies have made valuable progress towards better understanding the nature of productive knowledge and the process by which it is accumulated. However, as they represent a relatively nascent analytical framework, this thesis aims to tackle some of the present limitations and extend these methodologies in new, policy-relevant directions.

A key limitation this dissertation seeks to address relates to the ambiguity surrounding Hausmann et al. (2014)'s PCI and ECI metrics. Despite their recognised success in explaining patterns in prosperity, there has been some confusion about what the measures actually capture. Although the ECI was conceptually cast in terms of a 'generalised' or 'corrected' measure of a country's export diversity (Hidalgo and Hausmann, 2009, Hausmann et al., 2014), Kemp-Benedict (2014)

showed that the measure is actually mathematically unrelated (in fact, *orthogonal*) to country diversity. More recently, the ECI was described with reference to 'eigenvector centrality' (Morrison et al., 2017). However, as eigenvector centrality is directly proportional to diversity, Morrison et al. (2017)'s description is also inaccurate given Kemp-Benedict (2014)'s result.

In terms of new directions, this thesis seeks to make methodological contributions towards three key areas. First, it explores the potential of using *occupational* data to shed new light on the productive capabilities underpinning socio-economic systems. While existing efforts have largely focused on trade, industry and patent data, considerably less work has applied knowledge complexity and relatedness measures to occupational data. Occupational data has the advantage that it allows analysis of productive knowledge at the *individual* level. In addition to permitting a more granular analysis of the distribution of differentiated capabilities across individual persons within society, studying occupational data also provides a more direct link to tangible and specific skill-oriented policy levers. Such analysis could potentially yield useful clarity in policy discussions relating to education, skills and migration.

Second, this thesis aims to investigate the evolution of regional productive knowledge over 160 years of US developmental history – a length of time which does not appear to have been explored with these analytical approaches. This is largely due to the lack of consistent, detailed data stretching long time-spans. This thesis takes advantage of the Integrated Public Use Micro-series (IPUMS) database (Ruggles et al., 2017), which has only relatively recently made it possible to study regional employment in detailed industry and occupational categories over the 1850-2010

period.

Third, this thesis seeks to adjust the focus of these two existing analytical lenses to provide useful insights into a particular subset of productive knowledge that will become increasingly important over the coming decades – the knowledge and capabilities relevant to the green economy. While there have been some preliminary attempts in the literature to move in this direction (Huberty and Zachmann, 2011, Hamwey et al., 2013), this thesis develops new analytical metrics and creates a comprehensive dataset of green traded products to shed more extensive light on this area.

1.3 Overview of thesis and key findings

This dissertation has two key overarching objectives. First, it aims to further advance quantitative empirical approaches for analysing productive knowledge within socio-economic systems. Second, it seeks to demonstrate the value of these methodologies for shedding new light on important policy questions relating to (i) long run economic development, (ii) labour market segregation and job switching, and (iii) the transition to the green economy.

To address these objectives, this thesis undertakes a literature review, which provides an overview of the historical roots and some of the most relevant branches of the study of productive knowledge. It then proceeds with four empirical papers and a handbook chapter, which each aim to make methodological and/or policy-relevant contributions to the understanding of productive knowledge, and its applications, in different settings. Finally, it concludes with a discussion of the

key contributions, broader implications for research and policy, and suggestions for future work.

The first paper (Chapter 3), titled *A New Interpretation of the Economic Complexity Index*, is co-authored with Professor J. Doyne Farmer and Dr Alex Teytelboym and has been submitted to *Science Advances*. This paper makes a methodological contribution to the study of the complexity of productive knowledge by providing a new, mathematically precise interpretation of what the Hausmann et al. (2014) ECI and PCI measures capture. Rather than being linked to notions of 'generalised diversity' (Hidalgo and Hausmann, 2009) or eigenvector centrality (Morrison et al., 2017), this paper shows that these measures can be more appropriately understood as an application of a widely used spectral clustering algorithm, which is commonly used to partition a similarity graph (or network) into two components. When applied to country-export data, the ECI sorts countries into two clusters on the basis of the similarity in their export portfolios, where countries with positive ECI are assigned to one cluster and countries with negative ECI are placed in the other cluster.

However, this paper also shows that, more generally, the ECI and PCI metrics serve as particularly helpful dimension reduction tools. They collapse the high-dimensional space of country-export specialisations into two single-dimension vectors that best describe the similarities in the *types* of products that countries export. Analogous to the Dewey Decimal system for placing books with similar topic close together on library shelves, we show that the ECI provides an ordering that places countries with similar exports close to each other in the ranking, and those with dissimilar exports far apart. The corresponding PCI helps identify the

type of exports that countries at either end of the ECI spectrum have in common. That is, it indicates what types of exports distinguish countries with high ECI scores from countries with low ECI scores.

Importantly, this result allows us to re-interpret the empirical findings in Hidalgo and Hausmann (2009) and Hausmann et al. (2014). In accordance with the intuition of the Method of Reflections, these studies originally suggested that the ECI's strong positive correlation with per capita GDP and its ability to predict future growth rates underscored the importance of accumulating a diverse set of productive capabilities. However, this paper makes the difference between country diversity and the ECI explicit: country diversity captures *how many* products countries can export competitively, while the ECI reflects *what type* of products countries are specialised in. This distinction casts a number of existing results in the economic complexity literature in a new light, and also has some important implications for the broader literature on economic development.

The second paper (Chapter 4) further explores this new interpretation of the ECI and PCI metrics in a different setting – the regional development of US states over the 1850-2010 period. This paper, titled *Determining the differences that matter: New insights on development, dynamics and divergence in US states over 1850-2010*, is co-authored with Professor J. Doyne Farmer and Professor Ricardo Hausmann and is currently being prepared for submission to the *Journal of Economic Geography*. By applying these metrics to US census data on state employment concentrations in detailed industries and occupation categories, we show that over 160 years of developmental history, the ECI and PCI metrics consistently provides useful insights into the types of productive capabilities that dif-

ferentiate richer and poorer regions. Moreover, in congruence with the results for country-export data (Hausmann et al., 2014), these measures are also significantly predictive of future growth.

However, the application of the ECI and PCI metrics to such a long period of time yields a discovery not yet documented in existing work: the ECI and PCI metrics appear to be uniquely positioned to capture useful insights about the dynamics of productive knowledge in evolving economic landscapes. While the algorithm for calculating the PCI and ECI metrics *only* operates on the pattern of employment concentrations across US states (taking in no information about the types of industries or occupations, income or wages), we find that over the 1970-1980 period, the PCI of manufacturing-related activities declines substantially while the PCI of high-skill professional activities shows a marked rise. This paper discusses these dynamics and also considers the potential implications for the reversal of US regional income convergence from 1980 onwards and the economic decline of rust-belt states.

The third paper (Chapter 5) titled *What you do at work matters: New lenses on labour* is co-authored with R. Maria del Rio-Chanona and Professor J. Doyne Farmer and has been submitted to *Nature Human Behaviour*. This paper advances a novel approach to study the web of differentiated productive capabilities characterising the US labour force. Taking the analysis of productive knowledge 'to task', we measure the similarity (or relatedness) in specific work activities that people working in different occupations undertake, and show that this metric is more predictive of job-to-job transitions than existing benchmark measures.

On the basis of this measure, the paper then explores the network structure of

the division of labour. This network lens depicts strong patterns of segregation, particularly in relation to gender, skills and susceptibility to automation. We discuss these findings in light of the potential implications for social mobility as well as distributional concerns relating to the future of work.

The fourth paper (Chapter 6) then seeks to construct different analytical lenses to investigate the productive knowledge relevant to the green economy. This paper, titled *Economic Complexity and the Green Economy* is co-authored with Dr Alex Teytelboym, and has been submitted to the *Journal of the Association of Environmental and Resource Economists.* By collating and analysing a new dataset of green (environmentally friendly) exports and constructing new 'green complexity' measures, it provides novel insights into countries' current and future competitiveness in the green economy.

In particular, we develop a 'Green Complexity Index' (GCI), which ranks countries in accordance with the number and complexity of green products they can export competitively. Controlling for countries' per capita GDP and ECI, we find countries with higher GCI tend to have significantly higher environmental patenting rates, lower carbon emissions and more stringent environmental policies. We also construct a Green Complexity Potential (GCP) measure, which calculates each country's average relatedness to complex green products that it is not yet competitive in. We show that this measure is significantly predictive of countries' future competitiveness in green products. Additionally, in finding a strong positive correlation between GCI and GCP scores (suggesting significant path-dependence in the accumulation of green productive capabilities), we discuss the potential implications for green industrial policies.

Looking to the future, **the fifth paper** (Chapter 7) in this thesis suggests that the incremental accumulation of productive knowledge may no longer be able to guarantee prosperous and sustainable developmental outcomes. Instead, to address what Stern (2016) has described as 'the twin defining challenges of our century' – poverty alleviation and climate change – this paper draws attention to a somewhat under-appreciated, but profoundly important commonality in these dual objectives: both require societies to navigate and manage system-wide transformative change. Titled *Transformational Change: Parallels for addressing climate and development goals*, this paper is co-authored with Professor Cameron Hepburn and has been submitted as a chapter to the *Handbook on the Economics of Climate Change*.

Although significant efforts are underway to understand transformative change processes in respective climate and development fields, this paper highlights key advantages from integrating efforts across domains – particularly in relation to the understanding of technological spill-overs, diffusion and the evolution in industrial structures. It also reviews and discusses the relative merits of methodological approaches for analysing transformational change that have been used in both development and climate contexts, calling for cooperation to progress these important policy-making tools. Finally, it suggests that by combining and concentrating efforts in the climate and development fields towards these important goals, we might also make valuable progress towards advancing a much-needed generalised theory of transformational change in productive knowledge underpinning society.

2 Literature Review

2.1 The economic roots of productive knowledge and prosperity

The importance of productive knowledge for prosperity and economic development has deep roots in the economics literature. For example, in relating wealth to the division of labour, Adam Smith drew attention to the specificity of productive capabilities and highlighted how organising the productive efforts of society into different arrangements of tasks and expertise could profoundly influence the level of achievable output (Smith, 1776). Similarly, Friedrich Hayek stressed the importance of the "dispersed bits of incomplete and frequently contradictory knowledge which all separate individuals possess" and emphasised the role of markets and institutions in coordinating such disparate knowledge to serve productive purposes within society (Hayek, 1945, p. 519).

Thorstein Veblen's historic call for an evolutionary theory of economic development was underpinned by the desire to understand the 'sequence of change in the method of doing things'. He argued that, "when taken as items in a process of cumulative change or as items in the scheme of life, these productive goods are facts of human knowledge, skill and predilection; that is to say, they are, substantially, prevalent habits of thought, and it is as such that they enter into the process of industrial development" (Veblen, 1898, p. 387). Joseph Schumpeter shared Veblen's evolutionary view of the economy, but saw the entrepreneur as the key source by which new knowledge and new ways of doing things "keeps the capitalist engine in motion" and "incessantly revolutionises the economic structure from within"

(Schumpeter, 1994, p. 83).

Unfortunately, the dynamic, complex web of differentiated and interdependent productive capabilities conceptualised by these great minds did not easily lend itself to the dominate framework of analysis similar to most mainstream economists. Following the introduction of the Solow growth model, the most common treatment of knowledge still tends to be an exogenous (Solow, 1956, 1957) or endogenous (Romer, 1986, 1990, Lucas Jr, 1988) parameter in an aggregate production function. While these approaches are analytically tractable and provide useful insights into the patterns of development and long run growth, they provide no analytical lens on to the key features and dynamics that the aforementioned classical economists were so concerned with. Indeed, as Nordhaus and Tobin (1972, p. 2) so eloquently express:

> *"The steady equilibrium growth of modern neoclassical theory is, it must be acknowledged, a routine process of replication. It is a dull story compared to the convulsive structural, technological and social changes described by historically oriented scholars of development... The theory conceals, either in aggregation or in the abstract generality of the multi-sector models, all the drama of the events – the rise and fall of products, technologies, and industries and the accompanying transformation of the spatial and occupational distributions of the population".*

2.2 The study of productive knowledge evolves - outside mainstream economics

Seeking to address these issues, Nelson and Winter (1982) birthed a new strand of literature – a little ways outside the mainstream economics tradition. They introduced a theory of economic development that placed differentiated productive capabilities at centre stage. Moreover, in explicitly conveying the dynamics of economic change from an evolutionary perspective, their approach provided an important counterpoint to the steady equilibrium growth paths modelled under the neoclassical framework. Nelson and Winter (1982) described firms as heterogeneous organisations possessing different capabilities – referred to as *routines*. Defined fairly flexibly, routines related to skills or any 'repetitive pattern of activity' undertaken by an organisation, and were conceptualised as the analogous equivalent to genes in biological evolutionary theory. In Nelson and Winter (1982)'s theory, successful firms, which improved their routines and found better ways of doing things increased their profit and consequently grew and expanded at the expense of poorer performing firms.

However, Nelson and Winter (1982)'s firms behaved differently to the firms conceptualised in neoclassical economics. Rather than making make profit maximising decisions over exogenously given choice sets, Nelson and Winter (1982) described firms that learn new routines in a boundedly rational manner (Simon, 1972) and 'satisfice' somewhat imperfectly on the basis of their existing routines and knowledge. Further, drawing inspiration from Polanyi (1967), Nelson and Winter (1982) also recognised that a large proportion of organisational knowledge is tacit and can only be learned 'by doing' (Arrow, 1962). Such limits on knowledge codifiability

and the knowledge search process both constrain the development of firms' capacities, and consequently lead to path-dependence in the knowledge accumulation process (Kogut and Zander, 1992, Gertler, 2005).

While Nelson and Winter's (1982) theory spawned a new field of evolutionary economics (Anderson et al., 1988, Dosi et al., 1988, Arthur et al., 1997, Metcalfe, 1998, Dopfer and Potts, 2004, Dopfer et al., 2004, Dopfer, 2005, Dopfer et al., 2008, Metcalfe et al., 2018), their ideas were closely related to thinking taking place in other fields, such as the literature on strategic management and organisational studies. For example, there are clear overlaps with the resource-based view of the firm, which emphasises the role of strategic firm-specific capabilities for achieving sustainable competitive advantage (Penrose, 1959, Wernerfelt, 1984, Barney, 1991). Nelson and Winter (1982) ideas also provided a theoretical foundation for the 'dynamic capabilities' framework (Teece and Pisano, 1994, Teece et al., 1997), and the related concept of 'absorptive capacity' (Cohen and Levinthal, 1990), which focuses on a firm's ability to reconfigure itself and adapt to changing contexts and competitive landscapes.

The other important field focusing related notions of knowledge and productive capabilities was of course, economic geography. While the strategic management literature was primarily focused on the firm, economic geographers sought to understand the heterogeneous distribution of productive knowledge and economic activities across space. To answer questions on why manufacturing is concentrated in a few places (Krugman, 1991c), or why people cluster in cities despite the high rents (Glaeser et al., 1992), scholars had begun looking to theories of knowledge spillovers (which were also gaining interest in the economic growth lit-

erature (Romer, 1986, Lucas Jr, 1988, Grossman and Helpman, 1993, Jaffe et al., 1993)) and agglomeration externalities.

2.3 Knowledge spillovers and agglomeration externalities

One of the earliest theories of agglomeration was put forth by Alfred Marshall. Marshall (1890) suggested that firms within a given industry may benefit from locating in close proximity due to (i) lower transport costs for industry inputs and outputs, (ii) the sharing of a skilled labour pool and (iii) intra-industry knowledge-spillovers. In Glaeser et al. (1992)'s seminal work, the connection between Marshall (1890)'s early ideas and related formalisations by Arrow (1962) and Romer (1986) became known as the *Marshall-Arrow-Romer (MAR)* externality. The MAR externality suggests that the growth of a city should favour specialisation *within* a given industry. However, Jane Jacobs put forth an alternative view. She suggested that the most important knowledge spillovers occur *across* industries, arguing that "the greater and sheer number of and variety of division of labour, the greater an economy's inherent capacity for adding still more kinds of goods and services" (Jacobs, 1969, p. 59).

Following Glaeser et al. (1992), who aimed to put competing theories of agglomeration to the test in US cities, a swathe of empirical literature has sought to shed further light on whether a diverse or specialised productive base is more conducive to economic growth. Interestingly, the jury is still out: results are decidedly mixed (see Beaudry and Schiffauerova (2009) and de Groot et al. (2016) for extensive reviews). While differences in methodologies, geographical locations and time pe-

riods in history may account for some of the conflicting findings (Combes et al., 2004), studies have also demonstrated that agglomeration externalities differ by industry (Diodato et al., 2018), by plant (Neffke et al., 2011b), and over the industry life-cycle (Neffke et al., 2011).

2.4 The relatedness of productive knowledge - and the Product Space

One approach put forward to resolve what became known as the 'MAR vs Jacobs controversy' was the concept of related variety (Frenken et al., 2007, Castaldi et al., 2015, Content and Frenken, 2016). Postulating that growth and innovation is essentially a recombinant process (Weitzman, 1998) but some types of knowledge are easier to recombine than others, Frenken et al. (2007) drew on the hierarchical Standard Industry Classification (SIC) scheme to propose measures of *related* and *unrelated variety*. They found support for their hypothesis: Dutch regions with higher related variety (measured as the average entropy of employment across five-digit industry categories within each two-digit class) tended to have higher employment growth. They also found some (weaker) evidence that unrelated variety (measured as the entropy in employment across two-digit SIC industry classes) was significantly negatively related to unemployment growth, suggesting that the presence of unrelated sectors in a region could provide a buffer against unemployment shocks.

In the same year, a similar concept of *proximity* was proposed to understand how countries diversify their productive capabilities over time. Rather than deriving a measure of knowledge relatedness from a given classification scheme, Hidalgo et al.

(2007) measured the proximity between two exported products by considering how often they co-occur in countries' export baskets. The underlying intuition was that if two products are commonly co-exported, they are more likely to require similar productive capabilities. On the basis of this proximity measure, Hidalgo et al. (2007) constructed the *Product Space* - and introduced network analysis into the study of productive knowledge.

The Product Space is a network in which traded products are represented as nodes which are linked to each other if they are more likely to require similar production capabilities (i.e. have higher proximity to each other). In addition to providing a novel way to visualise countries' productive structures, the Product Space revealed some important new insights about economic development and the process of economic diversification. First, the Product Space helped visually highlight how not all industrial specialisations (and underpinning productive capabilities) are equal in their opportunities for knowledge spill-overs and future diversification. In particular, Hidalgo et al. (2007) showed that the Product Space had a distinct core-periphery network structure. Some products (such as metals, machinery and chemicals) occupied the densely connected core of the network, and were suggested to consequently have higher potential for knowledge spill-overs and diversification than other products (such as agriculture and mining), which were located in the more sparsely connected periphery.

Second, when examining different countries' export portfolios, Hidalgo et al. (2007) found that while developed countries' exports tended to be located in densely connected areas of the product space, many less developed countries were presently specialized in sparsely connected products located in the periphery. This lack of

connectedness, they argued, could translate into a limited potential for knowledge spill-overs and diversification – and explain some of the difficulties poorer countries face in attempting to undertake growth-enhancing structural transformation and converge to income levels of advanced countries.

Third, Hidalgo et al. (2007) demonstrated that the position of countries' exports in the Product Space could help identify specific industrial development opportunities that countries could more easily transition towards in the future. By developing a *proximity density* measure, which is based on the average proximity of a country's current exports to a given undeveloped product, they showed that countries were significantly more likely to diversify into products that were more proximate to their existing export portfolio.

The tendency for places to diversify into economic activities that are proximate (or related) to their current activities was also found to operate at the regional level - and in some cases, more strongly so (Boschma et al., 2013). Scholars have also investigated alternative approaches for measuring knowledge relatedness between industries. These methodologies include examining the co-occurrence of products belonging to different industries in manufacturing plants (Neffke et al., 2011a), inter-industry labour flows (Neffke and Henning, 2013, Neffke et al., 2017), and the strength of input-output linkages (Essletzbichler, 2015). Other studies have explored the relatedness of production knowledge underpinning different technologies by investigating patent citations (Leten et al., 2007, Rigby, 2015) and the co-classification of patents across different technology classes (Kogler et al., 2013, 2017).

2.5 Getting personal about productive knowledge

Some studies have also sought to understand the relatedness of productive knowledge through the lens of occupational data. Unlike trade, industry and patent data, occupation data allows for analysis of the distribution of knowledge at the *person* level. This provides some important advantages. First, to the extent that the primary agents responsible for the creation, recombination and transmission of knowledge are ultimately *individuals,* occupational data could provide a more direct lens onto the knowledge generation process. Second, studying what people do, rather than what firms or industries produce, could permit better empirical insights into the division of labour that Adam Smith originally theorised about. Third, scholars have recently highlighted the tendency for regions and cities to become more specialised in functions (with firms locating management and administrative headquarters in larger cities, and plants or production activities in other areas) (Duranton and Puga, 2005, Rossi-Hansberg et al., 2009, Martin et al., 2018) or even tasks (Acemoglu and Autor, 2011). As such, studying occupational employment concentrations may better capture the spatial separation of activities along the industrial value chain and could be more informative about the type of knowledge and capabilities distributed across different places.

One strand of research that has made progress in this direction is the work of Muneepeerakul et al. (2013). By studying the pattern of occupational employment concentrations in US metropolitan statistical areas (MSAs) from 2005-2011, they developed a co-location based 'Occupation Space' network, where occupations are linked to each other if they have a higher probability of being concentrated in the same MSA. Aligned with the findings for trade, industry and patent data,

Muneepeerakul et al. (2013) found that MSAs were more likely to develop new occupational concentrations that were related to the MSA's existing occupational concentrations. Controlling for MSA size, they also noted that productive MSAs were more likely to specialise in more unique (rare) occupations. In subsequent work, the same Occupation Space network was used to provide insights into the development of MSAs towards 'creative' (Shutters et al., 2016a) and 'green' urban economies (Shutters et al., 2016b).

However, it is questionable whether the co-location of occupational concentrations across MSAs provide a reasonable measure of knowledge-relatedness. Indeed, Shutters et al. (2016a) showed that as cities became more concentrated in 'creative' occupations (such as computer scientists, architects and designers), they also become more concentrated in non-creative occupations (such as cooks, subway operators and barbers). These occupations clearly involve different types of knowledge, and their co-location in cities is more likely to be a function of the demand for differentiated, complementary tasks in urban environments. As this thesis explores in Chapter 5, a more appropriate measure of occupation-based knowledge-relatedness could be based on similarities in tasks or activities that individuals actually perform in their jobs.

2.6 Taking productive knowledge to task

And indeed, following the seminal work of Autor et al. (2003)), there has been a growing body of literature devoted to the study of tasks in the labour market. The 'task framework' introduced by Autor et al. (2003) uses task intensity scores

from the Dictionary of Occupational Titles (DOT) (and its online successor – the online Occupational Information Network (O*NET) database) to measure the extent to which occupations are concentrated in routine, manual and abstract (or cognitive) tasks. This approach has been used to investigate a wide range of labour market questions relating to the impact of automation and technological change (Autor et al., 2003, Spitz-Oener, 2006, Autor, 2015, Berger and Frey, 2016), offshoring (Antràs et al., 2006, Grossman and Rossi-Hansberg, 2008, Blinder et al., 2009, Blinder and Krueger, 2013)), immigration (Peri and Sparber, 2009, Basso et al., 2017), structural employment shifts (Autor et al., 2006, Goos and Manning, 2007) and the evolution of wages (Acemoglu and Autor, 2011, Autor et al., 2008). In Chapter 5, this thesis contributes to this literature by applying some of the methodological techniques advanced in the relatedness literature to the occupation and task data – and provides a new, networks-based view onto tasks and the division of knowledge underpinning the labour-force.

While investigating tasks (what people do) provides one perspective on what workers know, an alternative approach has instead looked at workers studied at school or university. In particular, by examining the content (e.g. accounting, radiology or fashion design) and level (primary school through to post-graduate) of education for individual workers in Sweden, Neffke (2017) quantified the extent to which certain types of education are (i) substitutable (i.e. give access to the same occupations), and (ii) complementary (i.e. co-occur in economic establishments). He showed that the specificity of what workers know—and the extent to which it can be combined with complementary types of knowledge—matters. Swedish workers who possessed complementary skills to their co-workers were much better paid

than workers who were easily substitutable. Moreover, in demonstrating that these complementarity measures are able to explain a large proportion of the variance in well-known wage premiums, such as the returns to schooling, and working in larger plants or urban environments, Neffke (2017) underscored the importance of cultivating environments that allow for the effective combination and coordination of differentiated, interdependent types of productive knowledge.

2.7 Productive knowledge over the long run

Surprisingly few studies have attempted to investigate the long-term evolution of knowledge and productive capabilities – in part due to data constraints. However, one notable effort is the work of Michaels et al. (2013). This study drew on occupational data to investigate the extent to which US metro-areas had become more concentrated in 'interactive' tasks (involving thought, communication and inter-social activity) relative to non-metro areas over the 1880-2000 period. Using a novel approach for measuring occupational tasks by examining verb frequency in occupational descriptions given in the DOT, they show that the tendency for metro-areas to become more specialised in interactive occupations corresponds to the improvement and diffusion of communication and transport technologies. Michaels et al. (2013) argue that by increasing the demand for 'interactive' skills these types of technological improvements played a key role in shaping the task content of employment in urban areas. Chapter 4 of this thesis undertakes a different approach to examine the long-term evolution of productive capabilities. It analyses census data on occupation (and industry) employment concentrations in US states over the 1850-2010 period using a methodology originally introduced

to infer information about the *complexity* of places' productive knowledge.

2.8 The complexity of productive knowledge

The product-based Product Complexity Index (PCI) country-based Economic Complexity Index (ECI) were initially developed by Hidalgo and Hausmann (2009) to indirectly measure the relative complexity (or sophistication) of countries' productive capabilities from trade data. Their efforts followed a linage of attempts in the development economics literature to quantitatively measure the productive capabilities that were important for economic growth (Lall, 2000, Lall et al., 2006, Hausmann et al., 2007, Sutton and Trefler, 2016). However, unlike previous efforts, which incorporated GDP per capita into their country capability metrics (and consequently suffer endogeneity issues when used to subsequently explain economic growth (Lall et al., 2006, Hausmann et al., 2007)), the ECI and PCI metrics were calculated purely on the basis of country export portfolios.

The ECI and PCI metrics turned out to be remarkably successful in explaining cross-country patterns in economic prosperity. In addition to being strongly correlated with country per capita GDP, the ECI was shown to be able to explain more variance in future growth rates than explanatory variables traditionally employed in growth equations, such has governance and institutional quality, education and competitiveness (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). However, despite their notable predictive power, the use of the metrics in further studies has remained relatively limited – in part, due to confusion about what the metrics actually capture and how they work. As such, in Chapter 3, the first contribu-

tion this thesis seeks to make to the study of productive knowledge is to address the methodological issues associated with the ECI and PCI metrics and provide a mathematically precise interpretation of what these measures capture.

2.9 Productive knowledge and the green economy

In addition to contributing to the literature on productive knowledge underpinning economic growth, development and the labour market, this thesis also seeks to progress an analytical framework for studying a fundamentally important set of capabilities – the productive knowledge relevant to the green economy.

While there are burgeoning efforts to qualitatively and quantitatively analyse 'green' growth (Hallegatte et al., 2012, Jacobs, 2013, Smulders et al., 2014, Bowen and Hepburn, 2014, Jakob and Edenhofer, 2014, Hepburn et al., 2018), low-carbon transitions (Foxon, 2011, Knopf et al., 2010, Fankhauser, 2013, Geels et al., 2016) and sustainable development (UNEP, 2011, Sachs, 2012, World Bank, 2012, Griggs et al., 2013), the possibility of bringing to bear some of the conceptual and analytical techniques discussed in this literature review has remained a relatively unexplored angle. A key reason for this dearth relates to the difficulty in defining and measuring productive knowledge and capabilities that are relevant to the green economy (Bucher et al., 2014). Although global efforts to define what it means to be 'green' have been under-way since 2001 when the World Trade Organisation instigated a mandate to reduce tariffs on environmental good and services (WTO, 2001), we still, at the time of writing, lack a universally accepted definition and measurement strategy.

Nevertheless, some recent efforts have aimed to forge progress in this direction. Drawing on sectoral patenting rates, exports and industry output, Fankhauser et al. (2013) examined the 'green competitiveness' of 8 countries in 110 manufacturing sectors. Inspired by the methodological approaches underpinning the Product Space (Hidalgo et al., 2007), Hamwey et al. (2013) provided an initial 'illustrative' attempt to identify 11 green products in the Product Space, but did not undertake further analysis. Huberty and Zachmann (2011) looked at the position of 6 green products in the Product Space (relating to electric meters, solar cells, wind turbines and nuclear reactors) to analyse the effectiveness of green industrial policies in the context of European countries. In Chapter 6, this thesis extends these efforts by developing new 'green complexity' measures based on the economic complexity analytical framework, and applies these metrics to a much more extensive dataset of countries and green products.

2.10 The need to better understand transformational change in productive knowledge

Looking to the future, Chapter 7 argues that in order to progress prosperous and sustainable developmental outcomes in the 21st century, research into the knowledge generation, recombination and diffusion process requires a step change.

Many scholars are now stressing that without radical, non-marginal change in the development and diffusion of low-carbon technologies, limiting global warming to 2 degrees will be extremely difficult (Reid et al., 2010, Burch et al., 2014, Fankhauser and Stern, 2016). At the same time, developing countries are encountering a significant paradigm shift of their own. With automation making lower labour

costs in developing countries a less relevant comparative advantage (Rodrik, 2016), and with a rapidly diminishing global carbon budget making emissions-intensive industrial activities less feasible from a climate perspective (Stern, 2008, IPCC, 2007), many developing countries are currently facing the challenging prospect of scouting out new ways for accumulating growth-enhancing capabilities.

Climate and development fields are running up against the same key issue – the need to better understand the process of transformational change in productive knowledge. As such, the final contribution this dissertation seeks to make is to highlight key commonalities and opportunities for collaboration across these interrelated research domains. Aiming to focus research efforts towards the most promising pathways forward, Chapter 7 also reviews the comparative strengths and weaknesses of existing tools for analysing transformational change processes and identifies current analytical challenges that need addressing. As is often the case with the advancement of knowledge, the cooperation and coalescence of two distinct fields can generate the most important steps forward. While greater integration of the climate and development disciplines is clearly necessary to overcome the 'twin challenges' of poverty alleviation and climate change, their coming together could also catalyse progress towards a generalised theory of transformational change.

3 A New Interpretation of the Economic Complexity Index

Abstract:[1] Analysis of properties of the global trade network has generated new insights into the patterns of economic development across countries. The Economic Complexity Index (ECI), in particular, has been successful at explaining cross-country differences in GDP/capita and economic growth. The ECI aims to infer information about countries' productive capabilities by making relative comparisons across countries' export baskets. However, there has been some confusion about how the ECI works: previous studies compared the ECI to the number of exports that a country has revealed comparative advantage in ('diversity') and to eigenvector centrality. We show that the ECI is, in fact, equivalent to a spectral clustering algorithm, which partitions a similarity graph into two parts. When applied to country-export data, the ECI represents a ranking of countries that places countries with similar exports close together in the ordering. More generally, the ECI is a dimension reduction tool, which gives the optimal one-dimensional ordering that minimizes the distance between nodes in a similarity graph. We discuss this new interpretation of the ECI with reference to the economic development literature. Finally, we illustrate stark differences between the ECI and diversity with two empirical examples based on regional data.

[1]We would like to thank Simon Angus, R. Maria Del Rio-Chanona, Neave O'Clery, Devavrat Shah and Muhammad Yildrim for very helpful discussions and Ricardo Hausmann, Eric Kemp-Benedict, Luciano Pietronero and Bassel Tarbush for valuable comments on the original draft of this paper.

3.1 Introduction

Structural properties of the global trade network can explain differences in economic development across countries (Hidalgo and Hausmann, 2009, Tacchella et al., 2012, Caldarelli et al., 2012, Hausmann et al., 2014, Saracco et al., 2015, Straka et al., 2017). One network measure, known as the Economic Complexity Index (ECI), aims to infer information about countries' productive capabilities by making relative comparisons across their export baskets (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). The ECI has been successful in explaining cross-country differences in GDP/capita and in predicting economic growth. However, there has been some confusion about what the ECI is and why it explains variation in economic development. First, the ECI has been cast in terms of a 'corrected' diversity measure (where diversity is the number of exports a country has a revealed comparative advantage in; see definition in Eq. (1) (Hausmann et al., 2014)). Second, the ECI has been compared to an eigenvector centrality measure (Morrison et al., 2017). This paper stresses that neither of these descriptions is accurate. The ECI is actually orthogonal to diversity (Kemp-Benedict, 2014) (i.e. the dot product between the ECI and diversity vectors is 0), which means that the ECI captures a feature of the global trade network that is distinct from diversity. Moreover, we highlight that eigenvector centrality is much more closely related to diversity than to the ECI.

We show that the ECI is equivalent to a classic spectral clustering algorithm (Shi and Malik, 2000), which partitions a similarity graph into two parts. Hence, when applied to country-export data, the ECI represents a ranking of countries that places countries with similar exports close together in the ordering and countries

with dissimilar exports far apart. More generally, the ECI can be seen as a dimension reduction tool, which gives the optimal one-dimensional ordering that minimizes the distance between nodes in a similarity graph.

Our results reveal several interesting insights for economic development and motivate potential new research avenues. First, in distinguishing the ECI from diversity, we highlight that both variables play important but different roles in the development process. Put simply, diversity captures *how many* products countries are competitive in. Country diversification patterns have been found to exhibit an inverted-U shape: countries diversify early in their developmental phase, and begin specializing at higher levels of per capita income (Imbs and Wacziarg, 2003). In contrast, the ECI captures *what type* of products countries are competitive in. By grouping together countries with similar exports and separating countries with dissimilar exports, the ECI sheds light on the type of production capabilities that separate high- and low-income countries and provides empirical validation of the long-standing theory of technological capabilities in development economics (Lall, 1992, 2000, Sutton and Trefler, 2016).

Second, by making the link between the ECI and spectral clustering precise, we open the door for further applications of these dimension reduction tools in economic development. While partitioning the country-export similarity graph into two clusters has proven to reveal a great deal about countries' productive capabilities at different developmental stages, future work could readily exploit the family of clustering approaches to glean further insights from data on economic networks.

Finally, we show that our new interpretation of the ECI helps explain its potential for applications in contexts other than international trade data. We briefly present

two empirical examples from forthcoming work that illustrate the difference between the ECI and diversity in regional settings. In these two settings, the ECI can explain differences in economic outcomes that cannot be captured by diversity.

3.2 The Economic Complexity Index

The ECI (and its related Product Complexity Index (PCI) measure for exported products) was originally defined using an algorithm that operates on a binary country-product matrix M with elements M_{cp}, indexed by country c and product p (Hidalgo and Hausmann, 2009). $M_{cp} = 1$ if country c has a *revealed comparative advantage* (RCA) > 1 in product p, where RCA is calculated using the Balassa index (Balassa, 1965), given by

$$RCA_{cp} = \frac{x_{cp} / \sum_p x_{cp}}{\sum_c x_{cp} / \sum_c \sum_p x_{cp}}, \tag{1}$$

where x_{cp} is country c's exports of product p. $M_{cp} = 0$ otherwise. In the remainder of this paper, we say that a country c has revealed comparative advantage or is *competitive* in product p if $M_{cp} = 1$.

Summing across the rows and columns of M gives a country's *diversity* (denoted $k_c^{(0)}$) and product *ubiquity* (denoted $k_p^{(0)}$), defined as

$$k_c^{(0)} = \sum_p M_{cp} \tag{2}$$

and

$$k_p^{(0)} = \sum_c M_{cp}. \tag{3}$$

The ECI and PCI were originally defined through an iterative, self-referential *Method of Reflections* algorithm which first calculates diversity and ubiquity and then recursively uses the information in one to 'correct' the other (Hidalgo and Hausmann, 2009) (see Methods).

However, it can be shown (Caldarelli et al., 2012, Cristelli et al., 2013) that the Method of Reflections is equivalent to finding the eigenvalues of a matrix $\widetilde{M}$, whose rows and columns correspond to countries and whose entries are given by

$$\widetilde{M}_{cc'} \equiv \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} = \frac{1}{k_c^{(0)}} \sum_p \frac{M_{cp} M_{c'p}}{k_p^{(0)}}. \tag{4}$$

Equivalently, we can write $\widetilde{M}$ in matrix notation

$$\widetilde{M} = D^{-1} M U^{-1} M', \tag{5}$$

where D is the diagonal matrix formed from the vector of country diversity values and U is the diagonal matrix formed from the vector of product ubiquity values. Since $\widetilde{M}$ is a row-stochastic matrix (which means its rows sum to 1), the leading eigenvalue is 1, and the associated leading eigenvector is constant.

The ECI is defined as the eigenvector associated with the *second-largest* eigenvalue of $\widetilde{M}$. The precise equivalence between the Method of Reflections and the eigenvector approach is discussed in the Methods section. In the Appendix (section 3.6), we present the analogous derivation of the PCI.

3.2.1 Interpreting the $\widetilde{M}$ matrix

Given the central role $\widetilde{M}$ plays in calculating the ECI, we briefly discuss what it represents. When applied to country trade data one can think of $\widetilde{M}$ as a *weighted* similarity matrix, reflecting how similar two countries' export baskets are.

Further, from Eq. (5), we can see that

$$\widetilde{M} = D^{-1}S, \tag{6}$$

where $S = MU^{-1}M'$ is a symmetric similarity matrix in which each element $S_{cc'}$ represents the products that country c has in common with country c', weighted by the inverse of each product's ubiquity. This formulation also makes it clear that the entries of the row-stochastic $\widetilde{M}$ matrix can also be interpreted as conditional probabilities in a Markov transition matrix (Hidalgo and Hausmann, 2009, Kemp-Benedict, 2014).

3.3 Results

In this paper, we denote the ECI vector by $\widetilde{y}^{[2]}$ and the ECI of country c is denoted $\widetilde{y}_c^{[2]}$. We also denote the diversity vector by d where $d_c = k_c^{(0)}$ is the diversity of country c. Additionally, we note that the ECI is commonly standardised by subtracting the mean and dividing by the standard deviation to allow for comparisons across years (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). However, for clarity, we use the unstandardised ECI vector throughout this paper.

3.3.1 The ECI and Diversity

The ECI measure is conceptually cast in terms of measuring country diversity and product ubiquity, and then iteratively 'correcting' a country's diversity by the ubiquity of its products. This follows from the hypothesis that prosperous countries have capabilities that allow them to competitively export a diverse range of products which few other countries are competitive in (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). However, *the ECI is orthogonal to diversity* (Kemp-Benedict, 2014) i.e. the dot product of the diversity and the ECI vectors is zero,

$$d \cdot \widetilde{y}^{[2]} = 0. \tag{7}$$

That said, when applied to the export data, the ECI and diversity are nevertheless positively correlated, as shown in Fig. 1 (Pearson $\rho = 0.64$, p-value $= 1.1 \times 10^{-15}$). Recall that unless the mean of one of the variables is zero, a zero dot product does not imply zero correlation. Neither diversity nor the (unstandardised) ECI have zero means in the data.

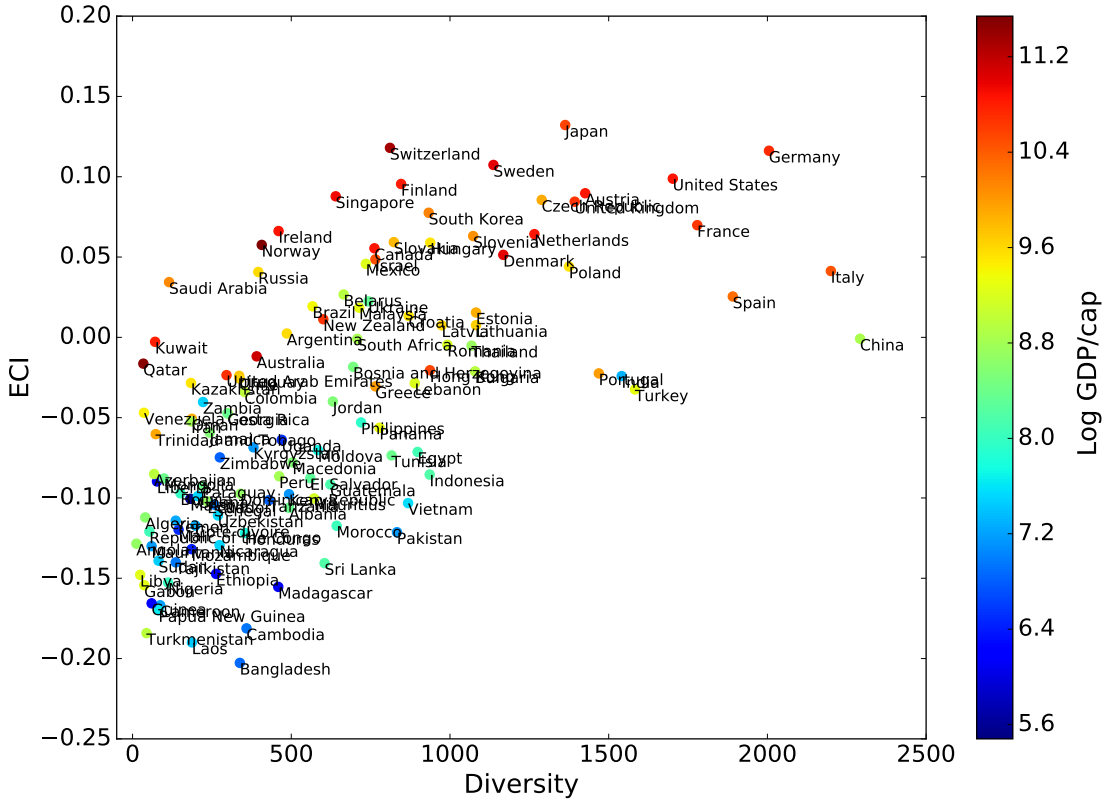


Figure 1: Country Diversity vs. the ECI compared to GDP per capita. Analysis is based on HS6 COMTRADE data for the year 2013 and GDP per capita data from the World Bank.

Despite being correlated, the orthogonality of the ECI and diversity measures suggests that they are capturing different aspects of the similarity of countries based on their export baskets. Intuitively, diversity captures *how many* exports a country is competitive in whereas the ECI captures *what type* of exports a country is competitive in.

We note that an alternative measure based on global trade data called *Fitness* Tacchella et al. (2012) is much more strongly correlated with diversity (Morrison et al., 2017). Spearman rank correlations for Fitness and diversity tend to be be-

tween 0.94 and 0.97 (Mariani et al., 2015). However, making comparisons between these the ECI and Fitness measures is beyond the scope of this paper.

3.3.2 The ECI and eigenvector centrality

The ECI has recently been described as 'standard eigenvalue centrality algorithm' (Morrison et al., 2017). Eigenvector centrality is defined as the eigenvector corresponding to the largest eigenvalue of an adjacency matrix and this definition is standard for symmetric matrices that represent undirected networks. In the case of directed networks, the natural definition is to take the left eigenvector corresponding to the leading eigenvalue of the adjacency matrix (Newman, 2010, p. 178).[2]

Therefore, the eigenvector centrality vector of $\widetilde{M}$ is the left (row) eigenvector x corresponding to the largest eigenvalue of the following eigenvalue equation

$$x\widetilde{M} = \lambda x. \tag{8}$$

Since $\widetilde{M}$ is row-stochastic, its largest eigenvalue is 1. Hence we are in interested in solutions to

$$x\widetilde{M} = x. \tag{9}$$

As the rows $\widetilde{M}$ have been normalized by diversity, it is easy to check that any vector proportional to d is a solution to Eq. (9). Since *eigenvector centrality of*

[2] In (Newman, 2010)'s exposition the adjacency matrix is transposed (i.e. $\widetilde{M}_{cc'}$ is an edge from c' to c), this is equivalent to finding the right eigenvector of the transpose of $\widetilde{M}$. The right leading eigenvector of $\widetilde{M}$ is a constant vector since $\widetilde{M}$ is row-stochastic.

$\widetilde{M}$ *is proportional to diversity*, it does not add anything to what we already know about $\widetilde{M}$.

From Eq. (6), we can also consider the symmetric similarity matrix $S = D\widetilde{M}$ which, when applied to country trade data, represents the similarity in countries' exports. Since S is a symmetric matrix, it represents an undirected network and is a more natural candidate for eigenvector centrality. In Fig. 2 we show that the *eigenvector centrality of S is highly correlated with diversity*. This is unsurprising. Diversity is the degree centrality of S and degree centrality is correlated with eigenvector centrality in many networks.

The fact that the ECI is defined as the eigenvector associated with the *second-largest* eigenvalue, whereas eigenvalue centrality is associated with the leading eigenvalue, makes it clear that *the ECI is not eigenvector centrality*. In fact, in the export data, eigenvector centrality is much more closely correlated to diversity than to the ECI (see Fig. 2).

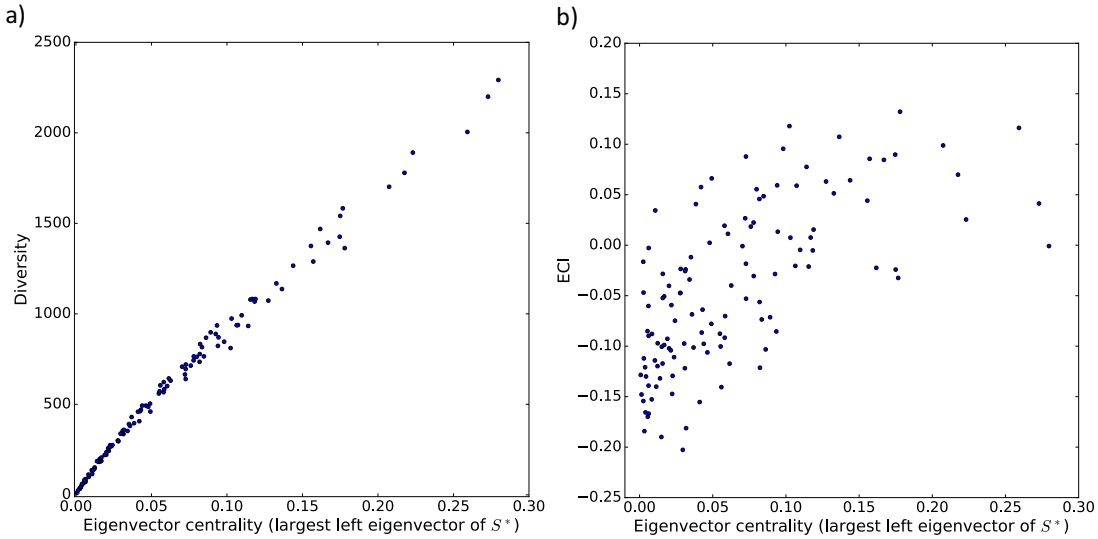


Figure 2: Panel (a) shows the eigenvector centrality of the export similarity matrix S vs country diversity (d) and Panel (b) shows the eigenvector centrality of S vs the ECI. Analysis is based on H6 COMTRADE data for the year 2013.

3.3.3 The ECI and Spectral Clustering

We now show that the ECI is equivalent to a spectral clustering method for the problem of partitioning an undirected weighted graph, represented by an adjacency matrix S, into two components (Shi and Malik, 2000). Spectral clustering is a widely used technique for community detection and dimensionality reduction and has a range of diverse applications including image recognition, web page ranking, information retrieval and RNA motif classification.

The goal of the spectral clustering approach is to minimize the sum of edge weights cutting across the graph partition, while making the size (number of nodes) of the two components relatively similar. As we discuss below, finding the exact solution to this problem is NP-hard. However, it is possible to obtain an approximate

solution by minimizing the *normalized cut (Ncut) criterion* (Shi and Malik, 2000). We demonstrate that the ECI is equivalent to this approximate solution.

3.3.3.1 The *Ncut* criterion

Consider an undirected graph $G = (V, E)$ with vertices V and edges E. We allow the graph G to be weighted, with non-negative weights so the adjacency matrix entries are $S_{ij} \geq 0$ where $S_{ij} = S_{ji}$. While the export matrix is one possible example, we can consider S to be any matrix with these properties. The degree of vertex i is defined as

$$d_i = \sum_{j \in V} S_{ij}, \tag{10}$$

and the size or "volume" of a set of vertices $A \subseteq V$ can be measured as

$$vol(A) = \sum_{i \in A} d_i. \tag{11}$$

One way to partition a graph into two disjoint sets is by solving the *cut* problem. The objective is to find a partition of V into complementary sets A and $\bar{A}$ that minimize the number of links between the two sets. The *cut* problem is to find the minimum of

$$cut(A, \bar{A}) = \sum_{i \in A, j \in \bar{A}} S_{ij}. \tag{12}$$

This objective function has the undesirable property that its solution often partitions a single node from the rest of the graph. To avoid this problem, the *normalized cut* (*Ncut*) *criterion* (Shi and Malik, 2000) penalizes solutions that are

not properly balanced. The objective is to partition the graph in such a way that each cluster contains a reasonable number of vertices. This can be achieved by minimizing the objective function

$$Ncut(A, \overline{A}) = (\frac{1}{vol(A)} + \frac{1}{vol(\overline{A})}) \sum_{i \in A, j \in \bar{A}} S_{ij}. \tag{13}$$

Let D be the diagonal degree matrix with $D_{ii} = d_i$ and $D_{i \neq j} = 0$. Then finding the minimum value of $Ncut$ is equivalent to solving the optimization problem

$$\min_{A} Ncut(A, \bar{A}) = \min_{y} \frac{y^T(D - S)y}{y^T D y}, \tag{14}$$

subject to $y_i \in \{1, -vol(A)/vol(\bar{A})\}$ and $y^T D\mathbf{1} = 0$.

Due to the fact that y_i is restricted to one of two possible values, this is not a simple linear algebra problem, and finding the true minimum of the $Ncut$ criterion has been shown to be NP-hard (Shi and Malik, 2000). However, by letting y_i take on any real value, an approximate solution can be obtained by finding the eigenvector $y^{[2]}$ corresponding to the second-smallest eigenvalue of the generalized eigenvalue equation

$$(D - S)y = \lambda D y. \tag{15}$$

Recall that $L_S = D - S$ is called the *Laplacian* matrix of S. By making the substitution

$$y = D^{-1/2} z, \tag{16}$$

this can be rewritten as a standard eigenvalue equation

$$D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}z = \overline{L}_S z = \lambda z, \tag{17}$$

where $\overline{L}_S = D^{-\frac{1}{2}}(D-S)D^{-\frac{1}{2}}$ is the *normalized Laplacian* of S. Because the normalized Laplacian is a stochastic matrix its smallest eigenvalue is zero. The eigenvector $z^{[2]}$ associated with the second-smallest eigenvalue of $\overline{L}_S$ is called the *normalized Fiedler vector*, and is a solution to the standard eigenvalue equation in Eq. (17). Transforming back to y using Eq. (16) to solve the original problem gives the solution

$$y^{[2]} = D^{-1/2}z^{[2]}. \tag{18}$$

The solution $y^{[2]}$ provides a useful approximate solution that minimizes the normalized cut criterion and is equal to a simple transformation of the normalized Fiedler vector.

3.3.3.2 The relationship between the ECI and the $Ncut$ criterion

Recall that $\widetilde{M}$ is the matrix whose eigenvector corresponding to the second-largest eigenvalue is the ECI. To see the relationship between spectral clustering and the ECI, note that the similarity matrix $S = D\widetilde{M}$ characterising country export similarity is in the same form used to minimise the normalised cut criterion. Multiplying both sides of Eq. (17) by $D^{-\frac{1}{2}}$ and re-arranging terms gives

$$D^{-1}SD^{-\frac{1}{2}}z = (1-\lambda)D^{-\frac{1}{2}}z. \tag{19}$$

Substituting $\widetilde{M} = D^{-1}S$ gives

$$\widetilde{M}D^{-\frac{1}{2}}z = (1-\lambda)D^{-\frac{1}{2}}z. \tag{20}$$

The eigenvalue equation for $\widetilde{M}$ is

$$\widetilde{M}\widetilde{y} = \widetilde{\lambda}\widetilde{y}. \tag{21}$$

Now, comparing Eqs. (20) and (21), we can see that the eigenvalues and eigenvectors of $\widetilde{M}$ are related to those of $\overline{L}_S$ by

$$\widetilde{\lambda} = 1-\lambda, \text{ and} \tag{22}$$

$$\widetilde{y} = D^{-\frac{1}{2}}z. \tag{23}$$

Thus the second-smallest eigenvalue of $\overline{L}_S$ corresponds to the second-largest eigenvalue of $\widetilde{M}$, and comparison to Eq. (18) makes it clear that the *ECI is equivalent to the spectral clustering solution of the normalized cut criterion*, i.e.

$$\widetilde{y}^{[2]} = y^{[2]} = D^{-\frac{1}{2}}z^{[2]}, \tag{24}$$

where $\widetilde{y}^{[2]}$ represents the second largest eigenvector of $\widetilde{M}$.

That is, the ECI ($\widetilde{y}^{[2]}$) is equivalent to the approximate solution ($y^{[2]}$) that minimises the normalized cut criterion. Moreover, the ECI is related to the normalized Fiedler vector by a simple transformation. In the Appendix (section 3.6), we also show how this interpretation can be applied to the PCI, and describe the precise relationship between the ECI and PCI.

3.3.3.3 The ECI partitions the country-export similarity graph

When interpreted as a clustering algorithm the ECI sorts countries into two clusters. It does this by assigning each country a real number on an interval with both positive and negative values, such that countries with similar ECI have similar exports. Countries with positive ECI are in one cluster and countries with negative ECI are in the other cluster, and the absolute value of a country's ECI measures the distance of any given country to the boundary between the clusters.

A visual representation can be seen in Fig. 3. Drawing on country trade data for 2013, we show the graph based on the similarity matrix S on the right hand side, where countries are represented as nodes and weighted links are given by $S_{cc'}$ (for visualisation purposes, we are only showing links with a weight larger than a given threshold. In this case we plot all links with a weight > 3). Here countries are coloured by their ECI, with darker shades of green representing higher complexity and darker shades of pink representing lower complexity. On the left hand side of the plot we show ECI values for each country plotted in ascending order. Countries with an ECI value greater than zero are in the green cluster and countries with a negative ECI are in the pink cluster. The ECI value (above or below zero) provides an indication of the distance from the cut.

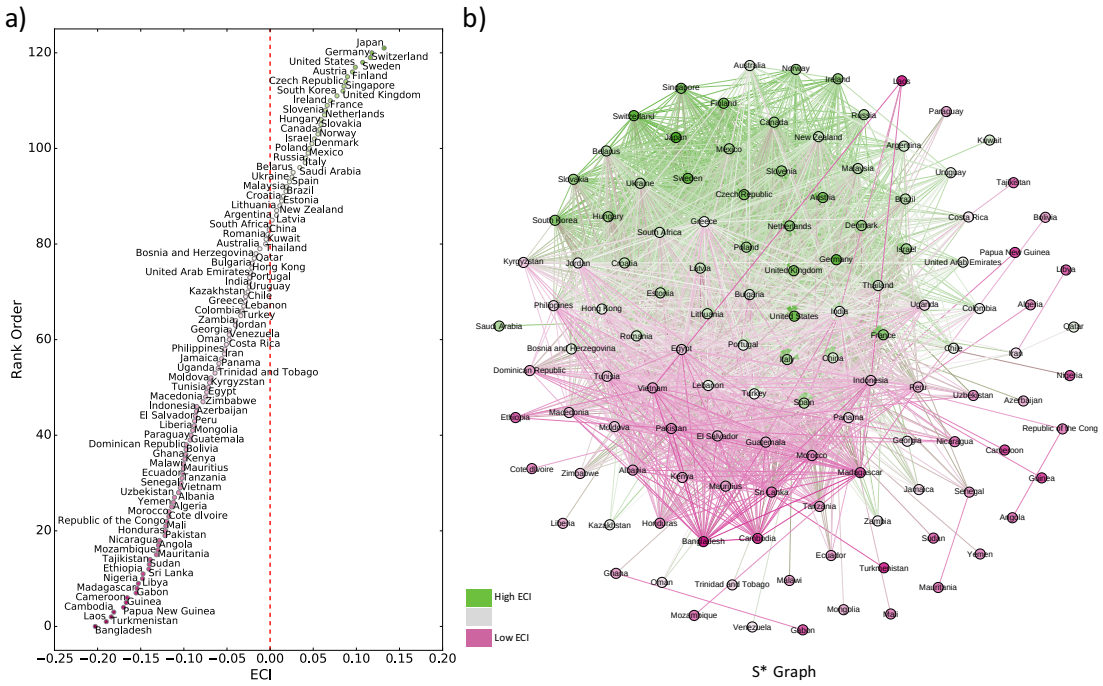


Figure 3: Visual representation of how the ECI vector partitions the graph based on similarity matrix S. The graph in Panel (b) shows country nodes coloured by their ECI value, with shades of green representing more positive ECI and shades of pink representing more negative ECI. The ECI rank ordering of countries in Panel (a) shows how countries having positive ECI fall on one side of the partition and countries having negative ECI fall on the other side of the partition.

3.3.4 The ECI as a dimension reduction tool

For the economic applications presented here, the ECI does not yield a clear separation into disjoint clusters. It nonetheless provides a useful rank ordering that places countries with similar exports near each other. To illustrate why this ordering is useful, suppose we hypothesize that there exists a relationship between the exports of a country and some quantity of interest, such as GDP per capita. To reduce the dimensionality of the problem it would be useful to find a rank

ordering that places countries with similar exports close to each other. However, there are potentially $C!$ ways to order a set of C countries. For $C \approx 100$ this number is intractably large, and searching for all possible rank orderings would be impossible.

It turns out that the ECI is *the* unique way of assigning a real number to each country in order to minimize the sum of the squared distances between countries, where the distances are weighted according to the similarity matrix S (Shi and Malik, 2000). While $y^{[2]}$ only approximates the normalized cut criterion, it exactly minimizes

$$\frac{\sum_{ij} (y_i - y_j)^2 S_{ij}}{\sum_i y_i^2 d_i}, \tag{25}$$

subject to the constraint

$$\sum_i y_i d_i = 0. \tag{26}$$

Note how the orthogonality between the ECI and diversity (Eq. 7) is hard-wired into in this minimization problem as a constraint (Eq. 26). In contrast to its application to graph partitioning, the ECI is an *exact* solution to this problem. The ECI thus provides a reduction of the high-dimensional space of countries and their exports onto a single dimension that proves to be very useful for problems such as understanding the relationship between exports and per capita GDP. In the Appendix (section 3.6) we discuss this result further and contrast it to more commonly used dimension reduction approaches such as principal components analysis and ordinary least squares.

3.3.5 The ECI can be more useful than diversity in regional settings

While the ECI has traditionally been applied to country trade data, we also provide a preview to forthcoming work which applies the ECI to data on regional employment in industries and occupations. In the context of trade data, the orthogonality between diversity and the ECI is often masked because these variables are empirically positively correlated (see Fig. 1). However, in the regional examples we present here, regional diversity and the ECI are no longer positively correlated.

Fig. 4 shows the relationship between diversity and the ECI measures calculated on the basis of UK regional employment concentrations in industries and US employment concentrations in occupations. In Panel a), the ECI of UK local authorities is negatively correlated with its industry diversity. In Panel d), there is no correlation between the state ECI and occupational diversity.

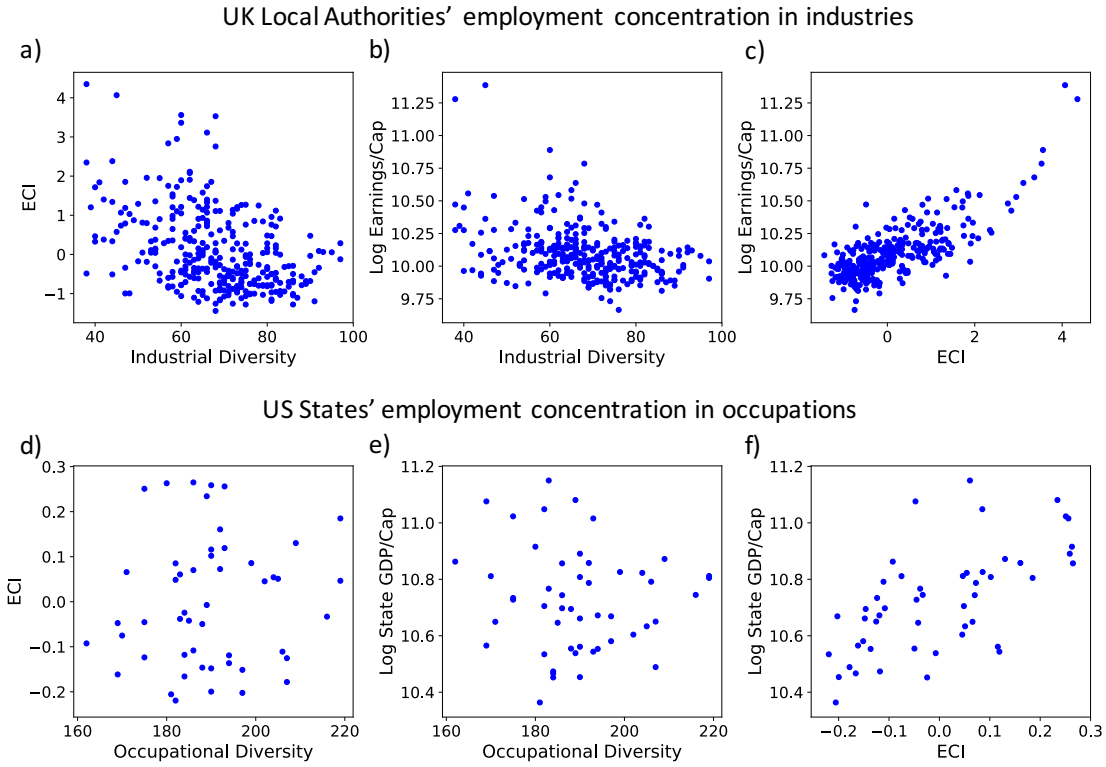


Figure 4: **Panels (a), (b), and (c)** show relationships between UK local authority industrial diversity, industrial-based ECI and log earnings per capita. Industrial employment data is sourced from the Business Register and Employment Survey for the year 2017 and 2011 work place earnings data is from the UK Office of National Statistics. **Panels (d), (e) and (f)** show relationships between US state occupational diversity, occupational-based ECI and log State GDP/cap. Occupational employment data is based on IPUMS data for the year 2010. State GDP per capita (2010) data is from the Bureau of Economic Analysis

Interestingly, in these two examples, we find that the ECI continues to be significantly positively correlated with regional per capita earnings and income (for UK local authorities, Pearson ρ between the ECI and log earnings per capita $= 0.76$, p-value $= 3.9 \times 10^{-61}$, and for US States, Pearson ρ between the ECI and GDP per capita $= 0.63$, p-value $= 7.46 \times 10^{-7}$). To the extent that a region's employment concentrations in industries or occupations capture information about its produc-

tive capabilities, our results suggest that the ECI is also able to shed light on the *type* of capabilities that separate rich and poor regions. Diversity, on the other hand, fails to meaningfully distinguish between high- and low-income regions in these two examples.

3.4 Discussion

In this paper we have shown that, rather than being linked to export diversity or eigenvector centrality, the ECI is equivalent to the spectral clustering solution of a normalized cut criterion (Shi and Malik, 2000). Our results have some important implications for the application of these measures to the development context. In particular, by making the difference between the ECI and diversity explicit, we can distinguish between the roles these measures play in the development process.

The relationship between diversification and development is well established in the economics literature. The general finding is that countries tend to follow a U-shaped pattern, where they first diversify and then begin specialising relatively late in the development process (Imbs and Wacziarg, 2003). This pattern aligns with other empirical studies that have described a positive association between export diversification and economic growth, which tends to be stronger for less developed countries (Al-Marhubi, 2000, Herzer and Nowak-Lehnmann D, 2006, Hesse, 2009).

In contrast to diversity, the ECI provides additional information relevant for economic development. As we illustrate in Fig. 3, the ECI provides a rank ordering of countries in terms of how similar their exports are to each other. The fact that this

ordering is useful in explaining variation in per capita GDP and predicting growth suggests that different types of exports (and by extension, productive capabilities) are associated with different growth and development outcomes. There is, indeed, empirical evidence that specializing in some products can lead to higher economic growth than specializing in others (Hausmann et al., 2007). Manufacturing products in particular show strong unconditional convergence in labour productivity (Rodrik, 2012). Within developing countries, technologically sophisticated products are more strongly associated with export and income growth (Lall, 2000).

We have also shown that the ECI can be thought of as a dimension reduction tool. It is the unique way to assign distances to countries such that the sum of their squared distances from each other is minimized, where the distance is measured using a weighted similarity matrix S. The ECI thus takes the complicated high dimensional space of countries and their exports and reduces it to a linear ordering, analogous to the Dewey-Decimal System for classifying books. The proven empirical success of applying this particular clustering algorithm to the country-export similarity matrix opens the door for future work to examine potential applications to the wider family of clustering and dimension reduction tools to learn further insights from the global trade network—or other data, such as country input-output tables.

Finally, we have shown two empirical examples to illustrate how in some settings, the ECI can be more useful than diversity in explaining variation in regional earnings and income per capita. Moreover, our new interpretation of the ECI suggests that, at least for these particular cases, the *type* rather than the *number* of industries and occupations concentrated in a region matters more for its economic

prosperity.

3.5 Methods

3.5.1 The Method of Reflections and its equivalence to the $\widetilde{M}$ eigenvalue problem

Following original Method of Reflections, we can take the measures of country diversity and product ubiquity (Eqs. 9 and 10) as a starting point and then recursively calculate the average values associated with country and product nodes' neighbors from the previous iteration step as shown in Eqs. (11) and (12), yielding

$$k_c^{(N)} = \frac{1}{k_c^{(0)}} \sum_p M_{cp} k_p^{(N-1)}, \tag{27}$$

and

$$k_p^{(N)} = \frac{1}{k_p^{(0)}} \sum_c M_{cp} k_c^{(N-1)}. \tag{28}$$

As $N \to \infty$ these variables converge to constant vectors, i.e. $k_c^{(\infty)} = k$ independent of c. Originally, values associated with a moderately large value of N ($N = 18$) were considered (Hidalgo and Hausmann, 2009). These measures produced useful deviations from the constant vector, and were shown to have a strong positive correlation between $k_c^{(N)}$ and log GDP per capita (Hidalgo and Hausmann, 2009).

The Method of Reflections was later reframed as an eigenvalue problem (Caldarelli et al., 2012, Cristelli et al., 2013, Hausmann et al., 2014), where it is shown that

inserting Eq. (28) into Eq. (27) and rewriting gives

$$\begin{aligned} k_c^{(N)} &= \frac{1}{k_c^{(0)}} \sum_p M_{cp} \frac{1}{k_p^{(0)}} \sum_{c'} M_{c'p} k_{c'}^{(N-2)} \\ &= \sum_{c'} k_{c'}^{(N-2)} \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} \\ &= \sum_{c'} \widetilde{M}_{cc'} k_{c'}^{(N-2)}, \end{aligned} \tag{29}$$

where

$$\widetilde{M}_{cc'} \equiv \sum_p \frac{M_{cp} M_{c'p}}{k_c^{(0)} k_p^{(0)}} = \frac{1}{k_c^{(0)}} \sum_p \frac{M_{cp} M_{c'p}}{k_p^{(0)}}. \tag{30}$$

The *Economic Complexity Index* (ECI) can then be defined as the eigenvector $\widetilde{y}^{[2]}$ associated with the second-largest eigenvalue of $\widetilde{M}$. Since $\widetilde{M}$ is row-stochastic, the leading eigenvalue is one and the leading eigenvector is constant. The iterative method gives essentially equivalent results due to the fact that the eigenvector for the second-largest eigenvalue corresponds to the direction in which the system converges most slowly onto the leading eigenvector.

Note that with the earlier recursive definition, while diversity and ubiquity are given as initial conditions, they become irrelevant in the limit as $N \to \infty$. Eqs. (27) and (28) define a linear dynamical system with a stable fixed point attractor, in which the solution is independent of the initial condition. Diversity and ubiquity are relevant only because they are incorporated into the definition of the dynamical system itself.

3.5.2 Calculating the ECI for UK and US regional employment data

3.5.2.1 UK Local Authorities and Industries

Using data from the UK Business Register and Employment Survey (BRES), we construct a binary *region-industry matrix* W on the basis of a region r's *Location Quotient* (LQ) in industry i

$$LQ_{ri} = \frac{e_{ri} / \sum_p e_{ri}}{\sum_r e_{ri} / \sum_r \sum_i e_{ri}}, \tag{31}$$

where e_{ri} is the number of people employed in industry i in region r and $W_{ri} = 1$ if $LQ_{ri} > 1$ and $LQ_{ri} = 0$ otherwise. Note that Eq. (31) is analogous to Eq. (1). We then constructed $\widetilde{W}$ matrix from W is the same way as $\widetilde{M}$ is constructed from M (Eq. 5). Finally, we calculate the *industry-based* ECI for UK Local Authorities by finding the eigenvector associated with the second-largest eigenvalue of $\widetilde{W}$.

3.5.2.2 US States and Occupations

We apply the same methodology to calculate the *occupation-based* ECI for US states. (We also find consistent results using data on US states and industries.) Drawing on census data for the US, which is available from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al., 2017), we construct a *state-occupation matrix* using state's location quotient in *occupation* i. We then compute the *occupation-based* ECI for US states analogously to the industry-based ECI for UK Local Authorities.

3.6 Appendix

3.6.1 The *Ncut* criterion also applies to the PCI

The spectral clustering interpretation can also be applied to the Product Complexity Index (PCI). Given the symmetry of the two measures, the PCI is the spectral clustering solution of the normalized cut criterion applied to the product-by-product similarity matrix $S = U\widehat{M}$ where

$$\widehat{M} = U^{-1}M'D^{-1}M, \tag{32}$$

and M is the M_{cp} matrix, D is the diagonal of the country diversity vector and U is the diagonal of the product ubiquity vector.

From the spectral clustering perspective, the PCI vector approximately solves the problem of partitioning the product similarity graph into two clusters of products that are similar in terms of the countries that export them.

We now show the precise relationship between the ECI and PCI.

Proposition 1. *A country's ECI is equal to the average PCI of products that the country is has revealed comparative advantage in.*

Proof. Recall that

$$\widetilde{M} = D^{-1}MU^{-1}M'. \tag{33}$$

The ECI is one of the solutions $\widetilde{y}$ to the following eigensystem:

$$\widetilde{M}\widetilde{y} = \widetilde{\lambda}\widetilde{y}. \tag{34}$$

To calculate the PCI for all products, we are interested in the second eigenvector of the matrix $\widehat{M}$, which is given by

$$\widehat{M} = U^{-1}M'D^{-1}M. \tag{35}$$

Hence, PCI is one of the solutions $\widehat{y}$ to the following eigensystem:

$$\widehat{M}\widehat{y} = \widehat{\lambda}\widehat{y}. \tag{36}$$

To prove the proposition, take Eq. (34), the eigensystem for the ECI, and substitute in Eq. (33):

$$D^{-1}MU^{-1}M'\widetilde{y} = \widetilde{\lambda}\widetilde{y}, \tag{37}$$

$$M^{-1}DD^{-1}MU^{-1}M'\widetilde{y} = \widetilde{\lambda}M^{-1}D\widetilde{y}, \tag{38}$$

$$U^{-1}M'\widetilde{y} = \widetilde{\lambda}M^{-1}D\widetilde{y}, \tag{39}$$

which is equivalent to the eigensystem for PCI,

$$U^{-1}M'D^{-1}M\widehat{y} = \widehat{\lambda}\widehat{y} \tag{40}$$

for

$$\widetilde{y} = D^{-1}M\widehat{y}. \tag{41}$$

as required. □

Therefore, the ECI can be immediately obtained from the PCI by using M. Moreover, note that all the eigenvalues of $\widehat{M}$ and $\widetilde{M}$ are the same.

3.6.2 Understanding the ECI as a dimension reduction tool

The ECI can also be useful as a dimension reduction tool. Appropriating another result from Shi and Malik (2000), we can see that while the eigenvector $y^{[2]}$ only approximates the normalized cut criterion, it exactly minimizes

$$\frac{\sum_{ij}(y_i - y_j)^2 S_{ij}}{\sum_i y_i^2 d_i}, \tag{42}$$

subject to the constraint

$$\sum_i y_i d_i = 0. \tag{43}$$

We have written this in coordinate notation for clarity. Given the similarity matrix S, the problem is to find real numbers y_i for each node i that minimize Eq. (42). To understand how this problem differs from that of finding the normalized cut, compare Eq. (42) here to Eq. (14) in the main paper. The denominators are the same and the constraints are the same, but the numerator in Eq. (14) of the main paper is replaced here by the sum of the squared distances between nodes weighted according to the similarity matrix S.

The numerator is suggestive of principal components analysis or ordinary least squares, but the problem being solved is different. Rather than being given a set of data points and finding a vector or vectors that minimize the distance from the data, here we are given a matrix S that characterizes the similarity of the nodes and we need to assign each node i a real number y_i. The solution $y^{[2]}$ has the special property that it minimizes the sum of the squared distance between

nodes, weighted by the similarity matrix S.[3] This provides a precise justification for the intuition of Hausmann et al. (2014, p. 24), who stated "[The ECI] is the eigenvector that captures the largest amount of variance in the system and is our measure of economic complexity." This also justifies Kemp-Benedict's (2014) interpretation of the ECI as a ranking countries according to how similar their exports are.

[3] The constraint of Eq. (43) is the orthogonality condition between diversity and the ECI, which guarantees that the interval spanned from lowest ECI to highest ECI contains zero, and more specifically that the sum of the diversity of the countries in one cluster equals the sum of the diversities of the countries in the other cluster.

4 Determining the differences that matter: New insights on development, dynamics and divergence in US states over 1850-2010

Abstract:[1] Understanding the differences between rich and poor places is complicated by the fact that places differ from each other in numerous ways. In this paper, we show how a dimension reduction algorithm can unveil hidden patterns in US census data and consistently yield useful insights into the type of economic activities that separate rich and poor states over 160 years of development history. Moreover, we find this approach has a unique ability to shed light on the dynamics of evolving landscapes and changes in relevance (or 'fitness') of particular types of activities, such as the shift from manufacturing to high skill services that occurred in the US over the last 40 years. We also highlight how this novel analytical lens can provide important new insights into the decline of the rust-belt and the reversal of US regional income convergence from 1980 onwards.

[1]We would like to thank Gordon Clark, Frank Neffke, Ljubica Nedelkoska, the CID Growth Lab Research Group and the INET Complexity Research Group for their valuable discussions and feedback on the work in this paper.

4.1 Introduction

How can we better understand the differences between rich and poor places? A key challenge for economists and economic geographers alike is that places differ to each other on many dimensions. Take Massachusetts and Mississippi: One state has a strong employment concentration in high tech and financial services, while the other is more concentrated in agricultural activities. But they also differ in terms of their natural resources, military facilities, manufacturing industries and concentrations of retail outlets and religious organisations. How can we know whether some of these differences matter more for economic performance than others? And what if we wanted to compare more than two places? How might we analyse the differences between Massachusetts, Mississippi and Michigan – or the entire set of US states?

One way to overcome the high dimensionality of making such comparisons is to ignore the *type* of industries or occupations a place is concentrated in, and just examine the extent to which a place is more or less diversified. And indeed, the topic of specialisation vs diversification has received significant attention within the economic geography literature, with mixed findings in terms of the relationship with economic performance (Attaran, 1986, Quigley, 1998, Duranton and Puga, 2000, Kemeny and Storper, 2015). More recently, scholars have suggested that when it comes to regional development, the 'what' might be more important than the 'scale' of specialisation (Kemeny and Storper, 2015, Martin et al., 2018).

In this paper, we demonstrate an alternative methodology for analysing differences in employment concentrations across places. We draw on a dimension reduction

algorithm that has been previously applied to country export data to calculate the country-based Economic Complexity Index (ECI) and product-based Product Complexity Index (PCI). This approach, which proved to be successful in explaining cross-country differences in per capita GDP and economic growth was originally advanced on the intuition that more prosperous places would be likely to export a large number of products that few other countries would be able to export (Hidalgo and Hausmann, 2009). However, the ECI was later shown to be mathematically unrelated to the number ('diversity') of products a country can export competitively (Kemp-Benedict, 2014). Instead, the value of the ECI and PCI metrics was suggested to lie in their ability to reduce the high-dimensional space of country-export specialisations into two single-dimension vectors that best describe the similarities in the type of products that countries export (Mealy et al., 2017).

As we discuss further in section 4.2, the ECI provides an ordering that places countries with similar exports close together in the ranking, and those with dissimilar exports far apart. Analogous to the Dewey Decimal system for placing books with similar topics close together in library shelves, the ECI turns out to be the optimal way to arrange countries in terms of the similarity in *what* they export. The corresponding PCI helps identify the type of exports that countries at either end of the ECI spectrum have in common. That is, it indicates *what* exports distinguish countries with high ECI scores from countries with low ECI scores.

Could these dimension reduction tools be useful in other settings? This paper investigates this new interpretation of the ECI and PCI metrics in the context of

regional economic development. We apply the Hausmann et al. (2014) algorithm to US census data on industrial and occupational employment concentrations across US states over 160 years of development history. For all periods for which we have state income data (1880-2010), we find that this approach consistently provides useful insights into the types of industries and occupations that underpin places of higher (and lower) economic prosperity. Moreover, in congruence with results for country-export data (Hausmann et al., 2014), we find these measures to be significantly predictive of future economic growth.

Applying this approach to census data over such a long time period also yields a further interesting discovery not yet documented in applications to country-export data: the ECI and PCI metrics appear to have a unique ability to capture the dynamics of evolving economic landscapes. Over the 1970-1980 period, the PCI of manufacturing activities declines substantially, while high skill services show a marked rise. It is important to stress that the dimension reduction algorithm takes in no information about types of industries or occupations, income or wages. It only operates on the pattern of employment concentrations across US states. However, this pattern turns out to be surprisingly informative of the changing relevance of particular *types* of specialisations in a given economic context.

We show that this structural change influence states' ECI trajectories in different ways. Although rust-belt states and eastern states had very similar ECI scores leading up to 1980 (suggesting they were specialised in similar types of activities), in 1980 the ECI of these two groups diverge – signalling very different future economic prospects. We find this new divergence between these groups of states is well explained by their differences in education, which were much higher in east

coast states than in rust-belt states. While educational differences did not strongly distinguish between states at either end of the ECI spectrum in previous periods, these differences have become particularly important for economic success in the new knowledge economy (Glaeser et al., 2014).

Finally, we suggest that the dimension reduction tools described in this paper could provide a new, complementary analytical angle onto the reversal in the pattern of state income convergence that occurred from 1980 onwards. While this 'great inversion' has been studied from numerous standpoints (Storper, 2018), we are not aware of literature that has explored income convergence through the lens of these dimensions reduction tools. Complementing the Krugman Index (or dissimilarity index) (Krugman, 1991d, 1993), which has previously been applied to investigate regional income convergence (Kim, 1998, Martin et al., 2018), we find the changes in the PCI and ECI over this period to be particularly informative of a new spatial divide separating US regions. Not only do we observe a bifurication in the PCI of manufacturing and high skill services over this period (Moretti, 2012), we also observe the error bars on these mean values becoming much further apart, suggesting that these activities have become much more strongly spatially segregated. While investigating the drivers of regional income convergence is beyond the scope of this paper, these preliminary findings align with existing theories linking slowing regional convergence to the spatial segregation of skills, which have become increasingly concentrated in cities (Glaeser and Saiz, 2003, Berry and Glaeser, 2005, Moretti, 2012).

The remainder of this paper is organised as follows. Section 4.2 reviews the ECI and PCI metrics and explains their use as dimension reduction tools. Section 4.3

describes the data and approach for applying the ECI and PCI metrics to state employment concentrations. Section 4.4, 4.5 and 4.6 detail our results, and section 4.7 concludes.

4.2 Understanding the ECI and PCI metrics as dimension reduction tools

The ECI and PCI metrics were originally introduced to infer information about cross-country differences in productive capabilities or 'know-how' from export data (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). Their calculation involves applying an algorithm to a binary country-product matrix M, with elements M_{cp} indexed by country c and product p. $M_{cp} = 1$ if country c has a *'revealed comparative advantage'* (RCA) > 1 in product p and 0 otherwise. RCA is calculated using the Balassa (1965) index

$$RCA_{cp} = \frac{x_{cp} / \sum_p x_{cp}}{\sum_c x_{cp} / \sum_c \sum_p x_{cp}}, \tag{1}$$

where x_{cp} is country c's exports of product p.

Note that this M matrix is a $c \times p$ dimensional object, which specifies each of the p export specialisations for every country c.

We can calculate how many products a country has RCA in (its *diversity*) by summing across the rows of the M matrix (denoted d_c). Similarly, we can count how many countries have RCA in a given product (its *ubiquity*) by summing aross the columns of the M matrix (denoted u_p). That is,

$$d_c = \sum_p M_{cp} \tag{2}$$

and

$$u_p = \sum_c M_{cp}. \tag{3}$$

While the diversity and ubiquity measures provide one way to compare the number of export specialisations across countries, they don't tell us anything about the *type* of exports countries are specialised in. However, the ECI and PCI metrics can be particularly helpful in this regard.

In Mealy et al. (2017), it was shown that the ECI vector gives the optimal one-dimensional ranking that places countries with similar types of exports close to each other in the ordering and countries with dis-similar types of exports far apart. Conceptually, the ECI is analogous to the Dewey-Decimal system for ordering books in a library. In order to make the process of finding a books easier, this system aims to place books on similar topics close together in library shelves and books on dis-similar topics further apart.

While this interpretation is explained in detail in Mealy et al. (2017), we provide a brief overview here. The ECI is defined as the eigenvector associated with the second largest eigenvalue of the matrix

$$\widetilde{M} = D^{-1}S, \tag{4}$$

where D is the diagonal matrix formed from the diversity vector, and S is a matrix

whose rows and columns correspond to countries and whose entries are given by

$$S_{cc'} = \sum_p \frac{M_{cp} M_{c'p}}{u_p}. \tag{5}$$

One can think of S as a symmetric similarity matrix, corresponding to how similar two countries' exports baskets are. Each element $S_{cc'}$ sums up the scarcity (one divided by ubiquity) of each product that countries c and c' both have RCA in.

Drawing on the work of Shi and Malik (2000), it can be shown that the ECI vector exactly minimises

$$\frac{\sum_{cc'} (y_c - y_{c'})^2 S_{cc'}}{\sum_c y_{c'}^2 d_c}, \tag{6}$$

subject to the constraint

$$\sum_c y_c d_c = 0, \tag{7}$$

where y_c and $y_{c'}$ represent the ECI scores of country c and c' respectively, and d_c denotes the diversity of country c. That is, the ECI vector is the exact solution to the problem of assigning real numbers y_c to each country c to minimise the sum of the squared the distances between countries, where distances are weighted by the similarity matrix S. The constraint $\sum_c y_c d_c = 0$ forces country ECI scores to take on positive and negative values, and to be reasonably balanced in their distribution above and below zero.[2]

From this perspective, we can see that countries with high ECI scores have similar types of exports to countries with other high ECI scores, and dis-similar types of exports to countries with low ECI scores. Moreover, this ECI ordering of countries turns out to be the optimal way to reduce the high-dimensional M matrix of

[2]It also forces the ECI vector to be exactly orthogonal to the country diversity vector.

country-export specialisations into a single dimension that best separates countries in terms of *what* they export.

Mealy et al. (2017) also showed that the ECI vector is the exact eigenvector that is used in Shi and Malik (2000)'s widely-used *Ncut* algorithm for bi-partitioning a similarity graph. That is, given a graph where countries are nodes linked to each other by their export similarities $S_{cc'}$, the ECI vector can be used to separate this graph into two balanced components that are internally similar and externally dis-similar. Here we would assign countries with positive ECI scores to one cluster and countries with negative ECI scores to the other cluster. The absolute value of a country's ECI score gives an indication of how likely it is to belong to its assigned cluster.

Since the PCI measure is symmetrically defined as the eigenvector associated with the second largest eigenvalue of the transpose of the $\widetilde{M}$ matrix, everything that applies to the ECI similarly applies to the PCI measure. That is, with respect to the countries that export them, products receiving high PCI scores are similar to products receiving high PCI scores and dis-similar to products receiving low PCI scores. Moreover, the PCI vector helps us identify the types of products that countries at one end of the ECI spectrum have *in common*. Products with high PCI scores are more likely to be exported by countries with high ECI scores, and less likely to be exported by countries with low ECI scores.

In Hidalgo and Hausmann (2009) and Hausmann et al. (2014), it was shown that countries with high ECI scores tend to be advanced countries that are more likely to export technologically sophisticated ('complex') products, while countries with low ECI scores tend to be less developed countries that tend to export less tech-

nologically sophisticated products. As such, the application of the ECI and PCI measures to country-trade data offers strong empirical support for developmental theories based on 'technological capabilities' (Lall, 1992, 2000, Lall et al., 2006, Sutton and Trefler, 2016), which contend that countries possessing more advanced technological capabilities are more likely to grow and prosper.

4.3 Application to US States: 1850-2010

Can the interpretation of ECI and PCI as dimension reduction tools give us new insights to regional development in the US? We explore this question by drawing on US census data over the period 1850-2010 from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al., 2017).

IPUMS provides decadal data on state employment in 3-digit industries and occupations. While regular changes in industry and occupation classification schemes have historically complicated longitudinal analysis, the IPUMS project has recoded census data into several sets of 'standardised' occupation classification schemes. We draw on the IND1950 and OCC1950 codes (based on the US Census 1950 classification scheme) to analyse the 1850-1950 period, the OCC1990 and IND1990 codes to analyse the 1950-1990 period and the OCC2010 and IND2010 codes to analyse the 2000-2010 period.[3] Unfortunately the 1890 period is missing due to a fire that destroyed these records. We also omit the year 1940 due to significant data inconsistencies (most likely due to the Second World War.)[4]

[3]While the IPUMS project makes the OCC1950 and IND1950 harmonised codes available for the entire 1850-2000 period, our analysis suggests that only backward projection of these harmonised codes (1850-1950) is likely to be appropriate. Using them to project into the future does not account for the introduction of new industry or occupation codes.

[4]The number of US states also changes over the study period. In 1850, the United States

To apply the Hausmann et al. (2014) algorithm to the IPUMS data, we first construct binary matrices W of state employment concentrations for each time period. These matrices are analogous to the binary M country-product matrix for export data. We calculate separate W matrices for state industrial employment concentrations (based on the harmonised IND codes) and state occupational employment concentrations (based on the harmonised OCC codes).

Each element W_{si} relates to a state's location quotient in industry or occupation i, given by

$$LQ_{si} = \frac{e_{si} / \sum_s e_{si}}{\sum_r e_{si} / \sum_s \sum_i e_{si}}, \tag{8}$$

where e_{si} is the number of people employed in industry or occupation i in state s. Analogous to the approach taken for country-export data, we let $W_{si} = 1$ if $LQ_{si} > 1$ and $LQ_{si} = 0$ otherwise.

We then define $\widetilde{W}$ as

$$\widetilde{W} = \widetilde{D}^{-1} \widetilde{S}, \tag{9}$$

where $\widetilde{D}^{-1}$ is the diagonal matrix formed from the diversity vector $d_s = \sum_i W_{si}$ of state employment concentrations (i.e the row sums of W) and $\widetilde{S}$ is a symmetric state similarity matrix whose rows and columns correspond to states, and whose entries are given by

$$S_{ss'} = \sum_i \frac{W_{si} W_{s'i}}{d_s}. \tag{10}$$

For each time period, we calculate a *State Industrial Complexity* (SIC) index by

comprised of 31 states and experienced subsequent additions until 1960. Our analysis includes data on all states available in each census year.

finding the eigenvector associated with the second-largest eigenvalue of $\widetilde{W}$ based on industrial employment concentrations. We calculate the corresponding *Industry Complexity* (IC) index by finding the second-largest eigenvalue of the transpose of the $\widetilde{W}$ matrix. We apply the same approach to the $\widetilde{W}$ matrix of state occupational employment concentrations to calculate the *State Occupational Complexity* (SOC) index and *Occupation Complexity* (OC) index.

The SIC (and SOC) indices for states are analogous to the ECI for countries based on export data. They give a one-dimensional ranking of states that places states with similar types of industrial (occupational) employment concentrations close to each other in the ordering and states with di-similar employment concentrations far apart. The IC (and OC) indices for industries (and occupations) are similarly analogous to the PCI for exported products.

4.4 Explaining differences in economic prosperity

In this section, we explore the relationship between the SIC and SOC indices and state per capita income over all periods for which we have income data. As the Gross Domestic Product (GDP) measure was only developed in the 1960s, we draw on estimates of state personal income per capita from Easterlin (1960), US Department of Commerce (1962) and Klein (2013) for the earlier 1880-1950 years. We unfortunately do not have income data for the 1850-1870 period. More information about these income sources and the differences between state personal income and state GDP can be found in the Appendix (section 4.8.1).

Not surprisingly, the SIC and SOC indices are highly correlated. Over the 1880-

2010 period, the Pearson correlation coefficient between these two measures lie between 0.74 and 0.93 (see Fig 7 in the Appendix section 4.8.2). This indicates that states that have similar industrial employment concentrations also tend to have similar occupational concentrations.

In Table 1, we examine the relationship between the indices and state per capita income. Model (1) and (2) regress state per capita income on the SOC index (with and without state fixed effects), Model (3) and (4) perform the same analysis with the SIC index, and Model (5) and (6) include both SIC and SOC variables in the regression. While the SIC and SOC indices are both strongly positively correlated with state per capita income over the 1880-2010 period, we find that the SOC index consistently has a slightly greater ability to explain variance in state per capita income than the SIC index.

Table 1: Regression analysis comparing the relationship between state per capita income and the SIC and SOC indices

	Dependent variable:					
	Log state income per capita (1880-2010)					
	(1)	(2)	(3)	(4)	(5)	(6)
State Occupational Complexity	0.230***	0.114***			0.241***	0.067***
	(0.010)	(0.013)			(0.019)	(0.025)
State Industrial Complexity			0.196***	0.079***	−0.012	0.042**
			(0.012)	(0.010)	(0.021)	(0.018)
Constant	5.136***	4.868***	5.136***	4.778***	5.136***	4.840***
	(0.062)	(0.075)	(0.066)	(0.074)	(0.062)	(0.077)
Year Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects?	No	Yes	No	Yes	No	Yes
Observations	593	593	593	593	593	593
Adjusted R^2	0.986	0.992	0.982	0.991	0.986	0.992
Note:						*p<0.1; **p<0.05; ***p<0.01

We suggest two possible reasons for the comparably stronger association between the SOC index and state per capita income. In the periods prior to 1910 the US census only asked respondents to report their occupation – no separate industry

question was posed. While it is usually straightforward to infer the appropriate industry form the occupational response, in some cases it is less clear. As such, industry coding for these earlier periods is less precise (Rondander, 1999). However, the stronger relationship between the SOC index and income per capita is not only limited to the period prior to 1910; we find this pattern to extend over the entire 1880-2010 period.

A second possible reason which could explain this phenomenon in the later years is the evolution from sectoral to functional specialisation. Rather than being home to an industry's entire value chain, cities and regions have experienced an increasing tendency to be specialised in functions (Duranton and Puga, 2005, Rossi-Hansberg et al., 2009) (or even tasks (Acemoglu and Autor, 2011)), with firms locating management and administrative headquarters in larger cities and plants or production activities in other areas. As such, the SOC index based on state occupational employment concentrations may better capture this spatial separation of activities along the industrial value chain. From this perspective, the index's closer alignment with per capita income could suggest that functional, rather than sectoral specialisations have stronger implications for a region's average level of economic prosperity.[5] For these reasons, we focus the most of the analysis in the remainder of this paper on the SOC index.

It is important to highlight that the high adjusted R^2 is primarily driven by the inclusion of the year fixed effects, which alone account for 97.3% of the variance in state per capita income. However, without year fixed effects, the SOC index alone still captures a fairly large proportion of the variance in state per capita

[5] Martin et al. (2018) similarly suggest that functional structure may be more important for productivity growth than sectoral specialisation.

income. As shown in the year-on-year correlations in Figure 1, the SOC index consistently accounts for at least 24% of the variance in state per capita income over the 1880-2010 period, and in some years explains up to 86%.

In Table 2, we also test whether the close association between the SOC index and state per capita income is robust to a number of controls. We find that the relationship remains strongly significant, even with the inclusion of variables relating to state population, occupational diversity (d_s) and each state's proportion of overseas and interstate workforce.

Table 2: Regression analysis testing the robustness of the relationship between SOC and state per capita income to a series of control variables

	Dependent variable:					
	Log state income per capita (1880-2010)					
	(1)	(2)	(3)	(4)	(5)	(6)
State Occupational Complexity	0.173***	0.126***	0.141***	0.052***	0.119***	0.075***
	(0.015)	(0.012)	(0.013)	(0.011)	(0.017)	(0.011)
Log State Pop	−0.108***	−0.215***			−0.063***	−0.162***
	(0.011)	(0.028)			(0.010)	(0.024)
State Occupational Diversity	0.004***	0.004***			0.002***	0.003***
	(0.001)	(0.0003)			(0.001)	(0.0003)
Proportion overseas workforce			1.428***	1.644***	1.204***	1.283***
			(0.163)	(0.170)	(0.153)	(0.159)
Proportion interstate workforce			0.388***	0.075	0.246***	0.104
			(0.065)	(0.100)	(0.061)	(0.095)
Constant	6.382***	7.716***	4.660***	4.536***	5.477***	6.747***
	(0.166)	(0.408)	(0.047)	(0.052)	(0.152)	(0.346)
Year Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects?	No	Yes	No	Yes	No	Yes
Observations	593	593	593	593	593	593
Adjusted R^2	0.989	0.994	0.990	0.994	0.991	0.995

Note: *p<0.1; **p<0.05; ***p<0.01

Population, overseas and interstate workforce data are also drawn from the IPUMS database (Ruggles et al., 2017). Proportion of overseas and interstate workforce statistics are calculated on the basis of the IPUMS 'Birthplace' variable, which indicates the US state or foreign country in which a worker was born.

Table 3 explores the relationship between a state's SOC index, and its future economic growth rate. In Model (1), we regress states' decadal compound annual growth rate against their SOC scores and control for their initial level of per capita income. We find that states with higher SOC tend to have higher future growth rates. In Models (2), (3) and (4), we show that this result is also robust to a series of control variables.

Table 3: SOC Growth Regression Analysis

	Dependent variable:			
	Compound Annualised Growth Rate per capita (by decade): 1880-2010			
	(1)	(2)	(3)	(4)
State Occupational Complexity	0.002**	0.002*	0.002***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Initial Log Income Per Capita	−0.023***	−0.026***	−0.024***	−0.026***
	(0.003)	(0.004)	(0.004)	(0.004)
Log State Pop		−0.001		−0.001
		(0.001)		(0.001)
State Occupational Diversity		0.0001		0.0001
		(0.00003)		(0.00003)
Proportion Overseas Workforce			−0.0005	−0.002
			(0.010)	(0.010)
Proportion Interstate Workforce			0.005	0.004
			(0.004)	(0.004)
Constant	0.130***	0.150***	0.134***	0.147***
	(0.017)	(0.022)	(0.020)	(0.023)
Year Fixed Effects?	Yes	Yes	Yes	Yes
Observations	542	542	542	542
Adjusted R^2	0.875	0.875	0.875	0.875
Note:				*p<0.1; **p<0.05; ***p<0.01

Let us pause briefly to digest what these regression results mean in light of our interpretation of the SOC index as a one-dimensional ordering that places states with similar types of occupational specialisations close together in the ranking and those with dis-similar types of occupational specialisations far apart. The strong and robust relationship between the SOC index and state per capita income portrayed in Table 2 suggests that over the 1880-2010 period, states with similar *types* of occupational concentrations tend to have similar levels of per capita income.

This finding is, in a sense, both trivial and profound. It is trivial if one consid-

ers it in the two-region setting. Imagine two states that had identical economic structures. It would probably be fairly surprising if they *didn't* have very similar levels of per capita income. Now imagine 51 states with economic structures having varying degrees of similarities. It is not immediately obvious that if you collapse this high dimensional space of state occupational specialisations into a single ordering reflecting relative state similarities, this sequence would correlate so strongly with the ordering of state per capita income. And yet, this result holds over the 130 year period for which we have state income data.

Returning to our two-region setting: suppose we did find two states that had identical economic structures, but *different* levels of per capita income. Most economists would expect some degree of convergence in incomes - the state with lower per capita income should be likely to grow relatively faster, and 'catch up' with the living standards enjoyed by the other state (see for example Kim (1998)). And indeed, this basic intuition is reflected in our growth regression analysis: states that have a higher SOC index than one would expect given their current per capita income (i.e. look more similar to wealthier states) do tend to grow faster over the subsequent decade.

How does a measure that captures *type* of occupational specialisations differ to measures that reflect the *number* or *spread* of state occupational specialisations? In Figure 1, we compare the SOC index to state occupational diversity (d_s), the inverse Herfindahl-Hirschman Index (HHI) and Shannon's Entropy in terms of their relationship with state per capita income over the 1880-2010 period. These three measures each capture related notions of the extent to which a state's employment distribution is diversified over different occupational categories. State

occupational diversity simply sums up the number of occupations for which a state has a $LQ > 1$. The inverse HHI sums up the square of each state's employment share s_i in each occupation i (and takes the inverse), and is given by

$$Inv\ HHI = \frac{1}{\sum_i s_i^2}. \tag{11}$$

Shannon's entropy treats state employment shares as a probability distribution and calculates the degree of uncertainty that a given worker in a state will be employed in a particular occupation

$$Entropy = -\sum_i s_i \log(s_i). \tag{12}$$

For each of these measures, higher (lower) values indicate greater (lesser) state employment diversification. As shown in Figure 1, states with higher diversification tended to have higher levels of per capita income up to around 1960. However, this association starts breaking down from 1970 onwards. In contrast, the SOC index continues to be strongly positively correlated with state per capita income over the entire 1880-2010 period. This suggests that over time (most likely due to reductions in transport and communication costs (Gaspar and Glaeser, 1998, Glaeser and Ponzetto, 2007, Michaels, 2008, Michaels et al., 2013, Leamer and Storper, 2014)) the *type* of occupations states are specialised in has become more important in distinguishing between richer and poorer places.

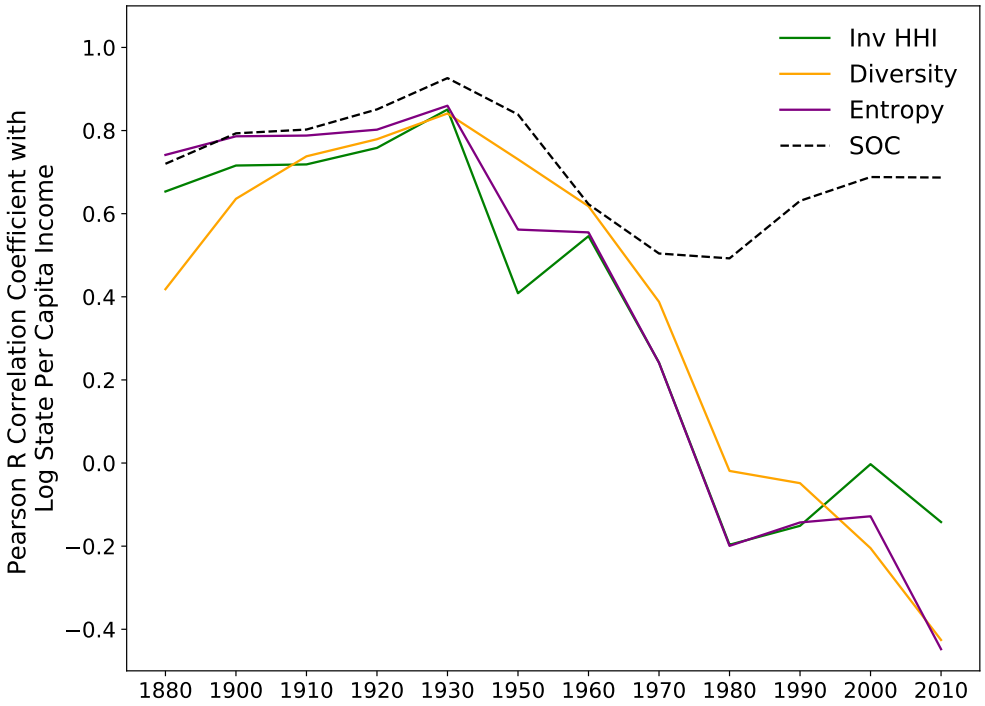


Figure 1: Comparison of the relationship between log state per capita income and Herfindahl-Hirschmann index (HHI), Occupational Diversity, Entropy and the SOC index over the 1880-2010 period.

4.5 What separates rich and poor states?

Relative to poorer places, what type of occupations are richer places concentrated in? Recall from Section 4.2 that in the context of country-export data, the PCI metric helps us identify the types of products that countries at one end of the ECI spectrum have in common. In this section, we draw on the Occupational Complexity (OC) index to identify the types of occupations that states with high (low) SOC scores have in common.

It is important to emphasise that the tendency for places with high (low) SOC

scores tend to be more (less) prosperous is *not in-built* into these metrics. The algorithm only operates on the binary pattern of state employment concentrations in the W matrix, and takes in no further information on income, wages, skill level, or occupational classification (such as agricultural, manufacturing or services). However, as we demonstrated in the previous section, the SOC index tends to consistently distinguish between states with higher and lower levels of income - and this strong and robust result reflects the differences in the types of occupations concentrated in richer and poorer places.

To test this even further, we calculate a measure that explicitly captures an occupation's concentration in prosperous places. Inspired by the *'PRODY'* income-content measure for exported products (Hausmann et al., 2007), we calculate an *Occupational Income Content* ($OccY$) measure, which calculates the average per capita income of states that have a $LQ > 1$ in a particular occupation. That is, for each occupation i, $OccY_i$ measures the average prosperity of states that are concentrated in it, and is given by

$$OccY_i = \frac{1}{\sum_s W_{si}} \sum_s W_{si} Y_s, \tag{13}$$

where W_{si} relates to elements in the binary W matrix of state occupational concentrations, and Y_s relates to the per capita income of state s.

Figure 2 shows the Pearson R correlation coefficient for the year-on-year correlations between each occupation's $OccY$ and Occupational Complexity index over the 1880-2010 period. This strong relationship is on par with the year-on-year correlations between SOC and state per capita income over the same period (see

Figure 1), and provides further empirical evidence that the occupational differences separating rich and poor regions are well captured by the OC measure.[6] In the Appendix (section 4.8.4), we show that this relationship is robust to a series of controls including occupational wage, educational requirement and total employment.

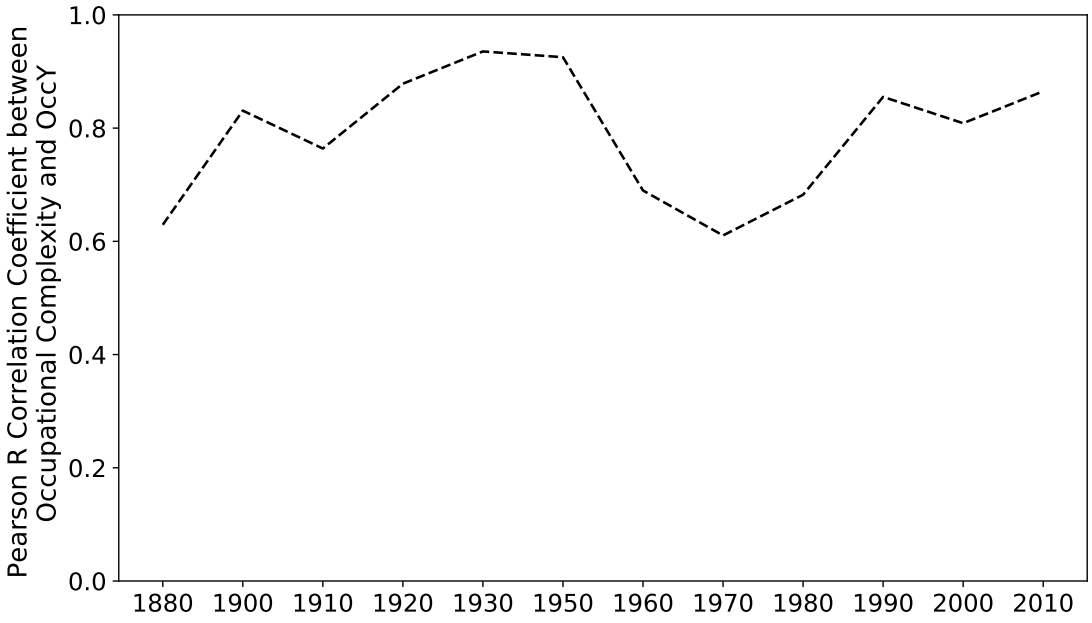


Figure 2: Pearson R correlation coefficient for the relationship between Occupational Income Content (OccY) and Occupational Complexity over the 1880-2010 period.

In Table 4, we show the highest and lowest ranked occupations in terms of their OC index for 2010. A cursory comparison of the top and bottom ranks suggests that, not surprisingly, prosperous places tend to be concentrated in high-skill services,

[6] A more subtle point is that, unlike the ordering given by $OccY$, which is directly related to per capita income, the OC index reflects the similarity gradient specifying the direction along which the similarity of occupations (in terms of the locations they are concentrated in) changes as much as it possibly can. To our knowledge, there is no theory that explains why this should be the case. As such, it opens the door for new developments in location theory to explain this robust empirical pattern.

while less prosperous places are underpinned by industrial activities. However, reflecting the surrounding ecosystem of activities that co-evolve with a region's tradable activities, a number of lower skilled occupations, such as baggage porters also occupy the top-ranks.[7]

While one might be tempted to refine the SOC and OC indices by activities that are typically non-traded from the W matrix, we show in the Appendix (section 4.8.6) that this reduces the ability of the SOC index to explain variance in state income per capita and predict future growth. The non-tradable supporting ecosystem of increasingly specialised activities turn out to be particularly important in distinguishing between rich and poor places. And indeed it is worth noting that even Adam Smith (1776) found cause to highlight the unique presence of particular activities that are *only* able to exist in prosperous places:

> *"There are some sorts of industry, even of the lowest kind, which can be carried out nowhere but in a great town. A porter for example, can find employment and subsistence in no other place."*

[7] This result is consistent with Shutters et al. (2016a), who examined the concentration of 'creative' and knowledge-intensive cities in US metropolitan statistical areas (MSAs). They found a tendency for 'creative' urban economies to specialise in both creative and non-creative occupations and suggested that this could be because particular non-creative occupations complement the tasks performed by creative occupations.

OC Rank	Occupation
1	Lawyers, and judges, magistrates, and other judicial workers
2	Actors, Producers, and Directors
3	Financial Analysts
4	Securities, Commodities, and Financial Services Sales Agents
5	Accountants and Auditors
6	Editors, News Analysts, Reporters, and Correspondents
7	Software Developers, Applications and Systems Software
8	Computer Scientists and Systems Analysts/Network systems Analysts/Web Developers
9	Personal Financial Advisors
10	Baggage Porters, Bellhops, and Concierges
11	Managers, nec (including Postmasters)
12	Taxi Drivers and Chauffeurs
13	Medical Scientists, and Life Scientists, All Other
14	Writers and Authors
15	Managers in Marketing, Advertising, and Public Relations
16	Architects, Except Naval
17	Other Business Operations and Management Specialists
18	Parking Lot Attendants
19	Management Analysts
20	Designers
…	…
434	Molders and Molding Machine Setters, Operators, and Tenders, Metal and Plastic
435	Chefs and Cooks
436	Industrial Production Managers
437	Paper Goods Machine Setters, Operators, and Tenders
438	Stock Clerks and Order Fillers
439	Industrial and Refractory Machinery Mechanics
440	Clergy
441	First-Line Supervisors of Production and Operating Workers
442	Industrial Truck and Tractor Operators
443	Woodworking Machine Setters, Operators, and Tenders, Except Sawing
444	Metal workers and plastic workers, nec
445	Laborers and Freight, Stock, and Material Movers, Hand
446	Other production workers including semiconductor processors and cooling and freezing equipment operators
447	Assemblers and Fabricators, nec
448	Welding, Soldering, and Brazing Workers
449	Millwrights
450	Driver/Sales Workers and Truck Drivers
451	Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic
452	Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic
453	Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders

Table 4: Pearson R correlation coefficient for the relationship between Occupational Income Content (OccY) and Occupational Complexity over the 1880-2010 period.

4.6 New insights into shifting economic landscapes

How has the type of occupations underpinning prosperous places has changed over time? And what implications might this have for long term regional development? In this section we investigate the evolution of SOC and OC indices over the 1850-2010 period.

Examining the trajectory of detailed occupational OC scores over such a long time horizon is complicated by changes occurring in the harmonised occupational classification scheme in 1950-1960, and 1990-2010. As detailed occupations across these classifications are not directly comparable, we look at the average OC values of detailed occupations falling into broader categories that we can compare across the entire period.[8]

In Figure 3, we show the average OC scores of occupations falling into three broad categories: professional, technical and finance occupations (including management, scientific, financial and other professional and technical activities), service occupations (including retail and administrative activities) and production (construction, extraction and manufacturing) occupations. Figure 3 reveals a fairly striking shift in OC scores from 1980 onwards. Professional, technical and finance occupations experience a distinct rise in their average OC scores, while production activities see a marked decline. The distance between these average OC scores also become much larger from 1980 onwards, as shown by the error bars which represent the 95% confidence interval around the mean values.

[8]More about this mapping of detailed occupations into broader categories can be found in the Appendix (section 4.8.3).

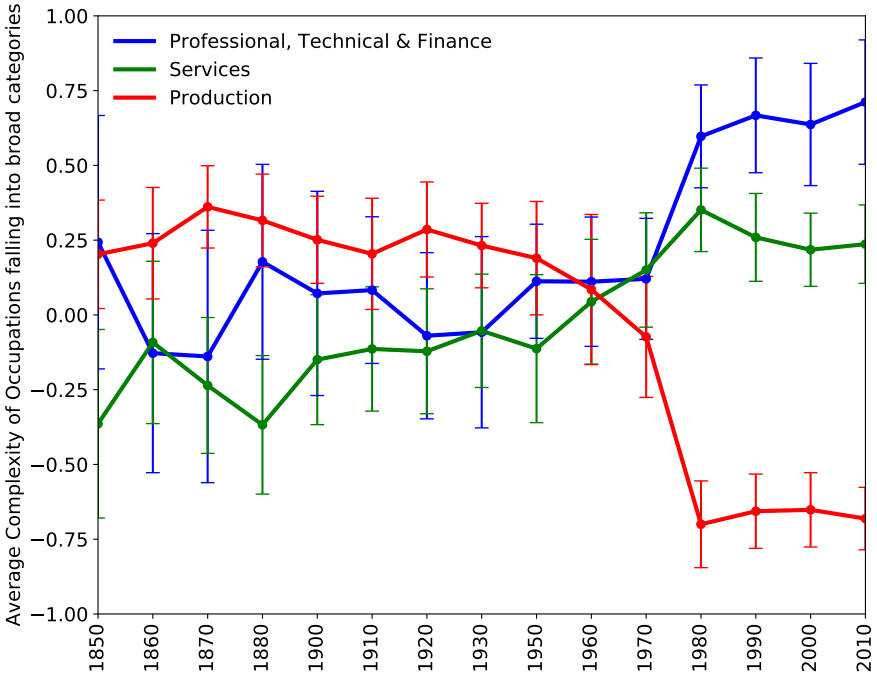


Figure 3: Average OC scores of Production, Services and Professional, Technical and Finance Occupations. Error bars relate to the 95% Confidence Interval.

It is important to note that this distinct change does not reflect changes in the occupation classification (these changes occurred in over 1950-1960, and 1990-2000). As we show in the Appendix, this structural shift in OC scores can also not be well accounted for by changes in occupational employment shares in these broad categories (as shown in Figure 13, this evolution was much smoother). Moreover, a very similar shift is evident when analysing metropolitan (rather than state) level data (see Figure 12), and when analysing industry complexity index scores (see Figure 11).

Before we digest the OC dynamics further, we first examine the corresponding evolution of state SOC scores over the 1850-2010 period. Figure 4 shows four key groups of states that experienced similar SOC development trajectories. While

California and a number of East-coast states maintained consistently high SOC scores over the entire period (Panel A), Southern agricultural-based states tended to have comparatively low SOC scores (Panel B). Panel C depicts many resource-rich states located in the US's North-West Mid-West region, which have shared similar SOC trajectories, but perhaps the most striking SOC dynamics occur in rust-belt states in Panel D. While rust-belt states had relatively high SOC scores (comparable to states in Panel A) for most of the earlier development phase, the 1970-1980 period sees their SOC scores decline dramatically.

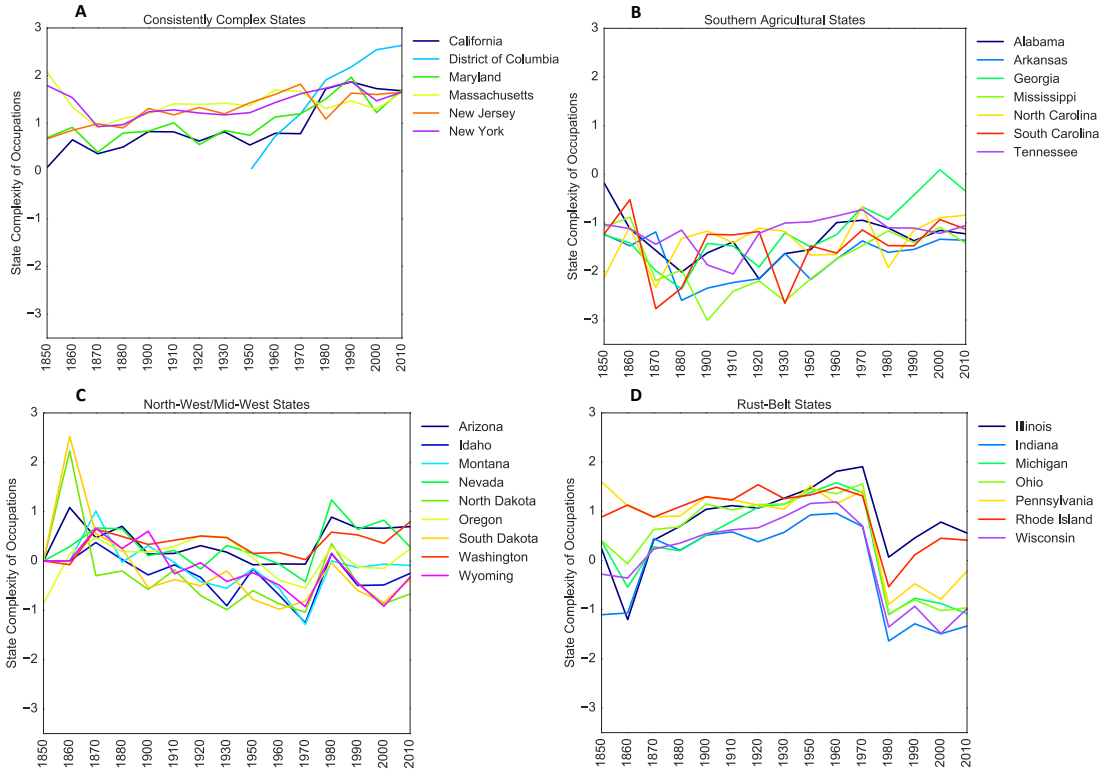


Figure 4: Evolution of State Occupational Complexity over time for a selection of US states with similar SOC trajectories.

Why did the OC and rust-belt SOC scores experience such a significant shift in

1970-1980? This particular decade is historically interesting on a number of fronts. The 1973 oil-crisis, which led to a quadrupling of oil prices, severely impacted US automotive businesses that were concentrated in rust-belt states (Bresnahan and Ramey, 1993, Lee and Ni, 2002). With higher fuel prices and severe gas shortages, domestic demand for 'gas gazzling' vehicles, which places like Detroit were famous for, shifted to more efficient models imported from Japan (Carrillo, 2004). However, for states endowed with coal, oil and gas reserves, the oil-crisis generated a short-lived boom, as higher energy prices stimulated mining and drilling activities around the country (Black et al., 2005, Allcott and Keniston, 2017).

At the same time, computers began proliferating the workforce - augmenting the productivity of higher skill occupations specialised in cognitive tasks and reducing demand for labour concentrated in routine and manual activities (Autor et al., 2003, Berger and Frey, 2016). However, skill-biased technological change was not the only force impacting US labour markets. As highlighted by (Autor, 2015, p10),

> *"Many forces distinguish the labour markets of these two epochs of 1940-1980 and 1980-2010. A partial list would include changes in the relative supply of college and non-college labour, rising trade penetration offshoring and globalization of production chains, declines in labour union penetration, the changing 'bite' of the minimum wage, and certain shifts in tax policy. Of course, many of these factors combine and interact as well such that attributing changes to a single cause would be foolish."*

Heeding Autor's words, we do not attempt to identify causal channels for the changes in SOC and OC scores in this paper. However, we do seek to reflect on the insights that these metrics give us about the changes in the US economic landscape over this period.

As reflected by their high SOC scores, both rust-belt states and East-coast (and

California) states were concentrated in relatively similar types of occupations in the period leading up to 1980. However in 1980, the SOC scores of these two groups of states diverge – signalling different future economic prospects for these states. The changes in OC scores in Figure 3 help identify the nature of these differences. East-coast states (and California) become much more concentrated in high-skill professional services, while rust-belt states remained largely specialised in manufacturing.

We stress again that the algorithm for calculating SOC and OC metrics is blind to the type (or classification) of occupations states are concentrated in – it only operates on the binary pattern of 1's and 0's in the W matrix and finds the direction in which states and occupations are maximally separated, in terms of the similarity in their employment concentrations. Yet this pattern of similarities appears particularly useful for identifying changes in relevance (or 'fitness') of particular activities for generating prosperity – without having any a priori theoretical intuition.

Why were the rust-belt states less able to transition to high-skill service activities like their East-coast counterparts? Figure 5 suggests that differences in education had a lot to do with it.[9] As shown in Panel A of Figure 5, which plots 1970 SOC scores against 1970 average state educational attainment levels, education does not well separate states at either ends of the SOC spectrum. However, a different picture emerges in 1980. Panel B shows that 1980 SOC scores are well explained by differences in 1970 average state educational attainment levels. And, indeed as we show in Figure 6, which shows the year-on-year Pearson correlation coefficient between the SOC index and average state educational attainment for

[9] Entrenched unionism and lack of competition in rust-belt states' labour and output markets may also have played a key role (Alder et al., 2014).

all years education data is available, the relationship between the SOC index and statement education attainment shows a marked shift in 1980. While it comes as no surprise that in the new knowledge economy, education clearly matters – we find that the SOC and OC indices have a unique ability to pin-point what has become an increasingly important divide (Berry and Glaeser, 2005).

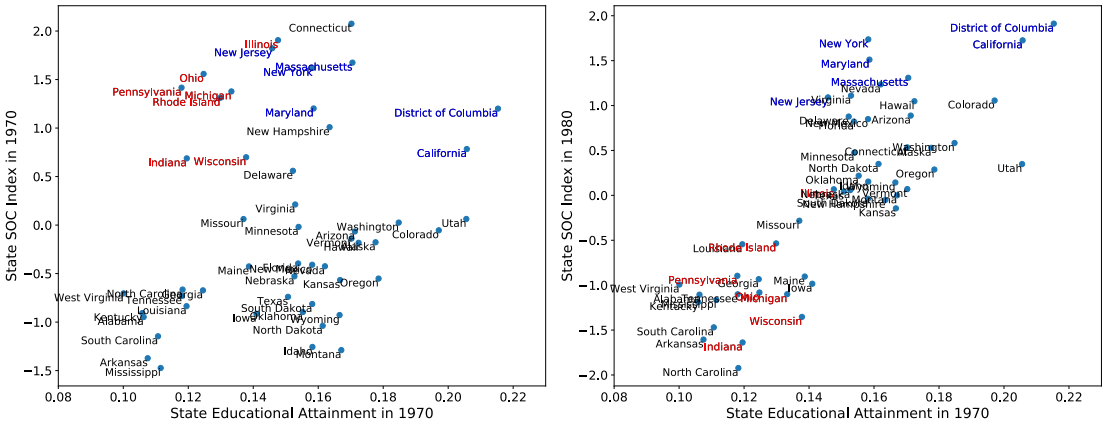


Figure 5: Panel A: Relationship between State SOC index in 1970 and State Educational Attainment in 1970. Panel B: Relationship between the SOC index in 1980 and State Educational Attainment in 1970. Rust-belt states are labelled in red, and East-coast states (and California) that consistently maintained high SOC scores are labelled in blue. State educational attainment are drawn from the IPUMS database (Ruggles et al., 2017) and calculated as the proportion of the state workforce having completed at least one year of college as given by the 'EDUC' variable.

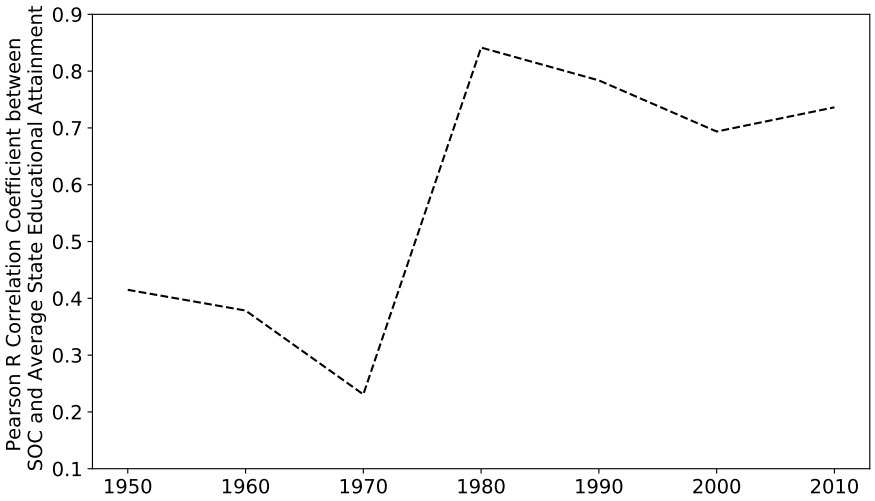


Figure 6: Pearson R correlation coefficient for the relationship between the SOC index and average state educational attainment over time.

One further important topic that these dimension tools could shed light on is regional income convergence. Although considerable economic disparities have long separated US states, throughout most of the 20th century poorer states were tending to grow faster relative to richer states (Barro and Sala-i-Martin, 1992). However, this process slowed substantially from 1980 onwards. A number of scholars have suggested that this phenomenon could be underpinned by the increasing returns to skills and innovation, which have become much more concentrated in cities (Glaeser and Saiz, 2003, Berry and Glaeser, 2005). As highlighted by Moretti (2012), p 106,

> *"It is almost as if, starting in the 1980s, the American economy bifurcated. On the one side, cities with little human capital and traditional economies started experiencing diminishing returns and stiff competition from abroad. On the other, cities rich in human capital and economies based on knowledge-intensive sectors started seeing increasing returns and took full advantage of globalised markets."*

Interestingly, this 'bifurification' is well captured by the OC and SOC dynamics

over this period. In Figure 3, not only do we observe a new polarisation of production and high skill services at either end of the PCI spectrum in 1980 (suggesting the activities became much more strongly spatially segregated), we also observe the error bars becoming much further apart.

Undertaking an in-depth study of the drivers of regional income convergence and divergence is beyond the scope of this paper. However, we believe that the SOC and OC indices could be useful additions to the current set of analytical tools available to investigate this phenomenon, complementing existing measures such as the Krugman Index of 'dissimilarity' (Krugman, 1991d, 1993), which has been used to examine regional convergence in the US (Kim, 1998) and the UK (Martin et al., 2018).[10] [11] Moreover, to the extent that convergence and divergence may depend on the type of economic activities that richer and poorer states are specialised in, the metrics we have described here could be particularly advantageous.

4.7 Conclusion

This paper has used a dimension reduction algorithm, which has previously been applied to country-export data (Hidalgo and Hausmann, 2009, Hausmann et al., 2014), to study differences in state employment concentrations over 1850-2010. This algorithm reduces the high-dimensional space of state specialisations into two single-dimension vectors that capture relative similarities in the *type* of economic

[10] We make comparisons to the Krugman Index over the 1850-2010 period in the Appendix (section 4.8.7).

[11] Berry and Glaeser (2005) also utilise a dis-similarity index and an isolation index to show that skilled workers have become more segregated across cities. While these results are consistent with the OC dynamics (which also hold at metro level, as shown in Figure 11), the OC measures have the advantage of not requiring information or a priori intuition on skilled workers.

activities states are concentrated in. In particular, the SOC index examined in this paper places states with similar occupations close together in the ranking and states with dissimilar occupations far apart. The corresponding OC index helps identify the types of occupations that distinguish states with high SOC scores from states with low SOC scores.

Despite operating only on the binary pattern of state location quotients in different occupations, we find that the resulting SOC and OC indices provide remarkably consistent insights into the types of activities distinguishing rich and poor places over the 130 year period for which state income data is available. The SOC index also outperforms measures of diversification, which become less useful in explaining differences in per capita income from 1930 onwards. It also proves to be a useful predictor of future economic growth.

While these findings align with intuition based on a two-region setting (two states with similar occupations should converge to similar levels of per capita incomes), we are not aware of any theories which explain why the linear ordering of states along the dimension captured by the SOC index should correlate so strongly with the ordering of state per capita incomes. New theories are needed to explain this robust empirical pattern.

We also find that in this particular setting, the SOC and OC indices demonstrated a unique ability to capture changing relevance of particular types of productive activities as the US economy underwent a fundamental structural shift in 1970-1980. While this finding suggests that these metrics could provide a new analytical lens to study structural change and regional convergence patterns more generally, future work is needed to verify this in other contexts undergoing similar patterns

of change. The UK regional economy could be a particularly interesting test case (Martin et al., 2018).

Finally, we believe the bifurification of SOC and OC scores in 1980 warrants further attention. The divergence in the SOC trajectories and economic prospects of rust-belt states remains pertinently relevant today. While more research is required to understand whether the *type* of differences separating regions could influence a nation's capacity for inclusive growth, we believe that these new metrics could help shed light on this important question.

4.8 Appendix

4.8.1 Historical state income data

Studying economic development over such a long stretch of history is complicated by the fact that state Gross Domestic Product (GDP) is only available from 1963 onwards. For previous years, we draw on estimates of state personal income per capita. Although GDP and personal income typically show very similar trends, the measures do have some key differences in their calculation. Personal income represents the value of economic activity received by people in a given economy. It includes wages, benefits, proprietor income, dividends, interest, rent and transfer payments (such as social security). GDP represents the value added in production by the labour and capital located in a state. It is calculated by summing up a state's gross output (sales or receipts and other operating income, commodity taxes and inventory change), minus its intermediate inputs (consumption of goods and services from other US industries or imports)(Landefeld and Marcus, 2006).

Unfortunately, state-level estimates of personal income for these earlier years are not available from a single source. Moreover, to our knowledge, no estimates are available for 1850-1880. We consequently draw on multiple sources to arrive at state-level estimates for the years 1880-1950. For the years 1880-1910 we draw on estimates available in Klein (2013). For 1920, we employ Easterlin (1960)'s estimates. For the period 1930-1950, we draw on a statistical abstract from the US Department of Commerce (1962).

4.8.2 Correlation between SIC and SOC indices

In Figure 7, we show the Pearson correlation coefficient for the relationship between the SIC and SOC indices for the period 1880-2010. Throughout this period, these measures are very strongly correlated, suggesting that states that have similar occupational structures also have similar industrial structures.

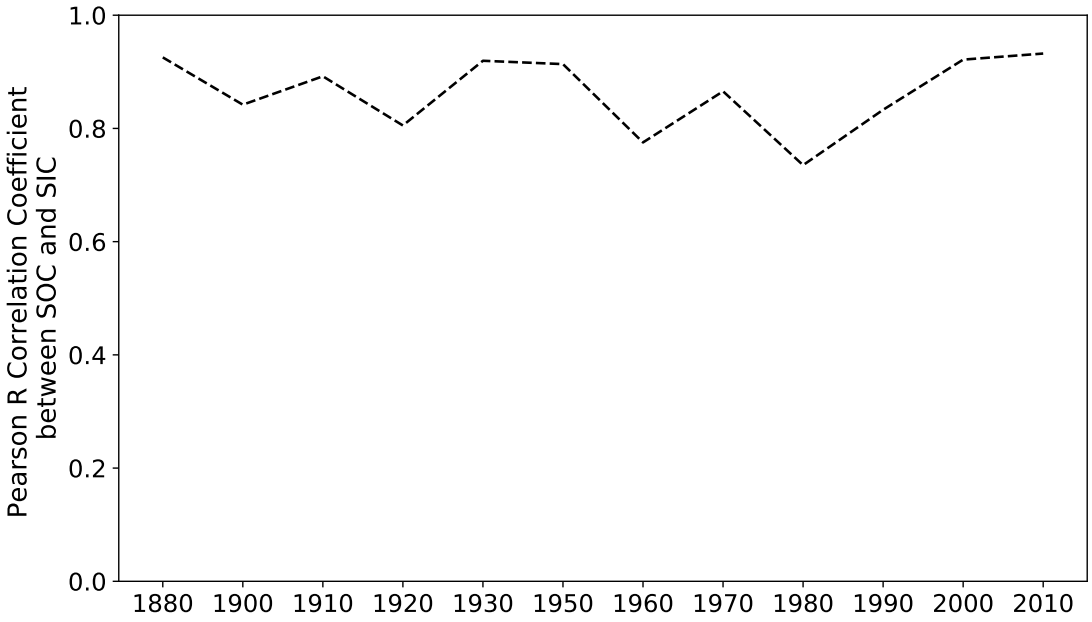


Figure 7: Pearson Correlation between State Industrial Complexity and State Occupational Complexity indices over 1880-2010.

4.8.3 Aggregation fo 3-digit IPUMS occupations and industries to broad categories

Analysing the OC and IC scores for detailed occupations and industries over such a long time horizon is complicated by changes occurring in the harmonised classification scheme in 1950-1960, and 1990-2000. As 3-digit occupation and industry

codes across classifications are not directly comparable over these periods, we construct broader classification groupings, and arrive at the results in Figure 3 and Figure 11 by calculating the average OC scores of 3-digit occupations falling into these broader groups over time.

For occupations, we primarily construct the broad 'Production', 'Services' and 'Professional, Technical and Finance' groupings for the OCC1950, OCC1990 and OCC2010 classifications by drawing on each 3-digit occupation's more aggregated 2-digit occupation, and then further aggregating the relevant 2-digit occupations into the relevant broad groupings. However, in some cases the 2-digit categories do not perfectly align with the broader grouping. For example in the OCC1950 classification, most occupation categories receiving the 2-digit category 'Professional, Technical' are assigned to the 'Professional, Technical and Finance' Group. However, although the 3-digit occupations 'Religious workers' and 'Music teachers' are assigned to the 'Professional, Technical' category, we assign these occupations to the Services category. We undertake a very similar procedure for grouping IND1950, IND1990 and IND2010 3-digit occupations into the broad industry categories shown in Figure 11. Crosswalk files specifying the treatment of each 3-digit occupation and industry code are available on request.

4.8.4 Additional analysis and robustness tests for the OC index

In this section, we examine the relationship between the occupational OC index and occupational wage, educational requirements and total employment over time. We also test whether the relationship between the OC index and the occupational $OccY$ measure is robust to a series of control variables.

The strong tendency for SOC and OC to correlate with income-related measures does lead one to question whether we could just be capturing wage-related effects associated with occupations. To test this, we draw on IPUMS occupational earnings variables and examine the relationship between an occupation's median wage and its complexity score. As we have three different occupational classifications relating to 1950, 1990 and 2010 occupational codes, we test the relationship for each classification year.[12] As shown in Figure 8, we find a fairly weak positive relationship between occupational earnings and OC scores, suggesting that our complexity measures are capturing more than general wage effects.

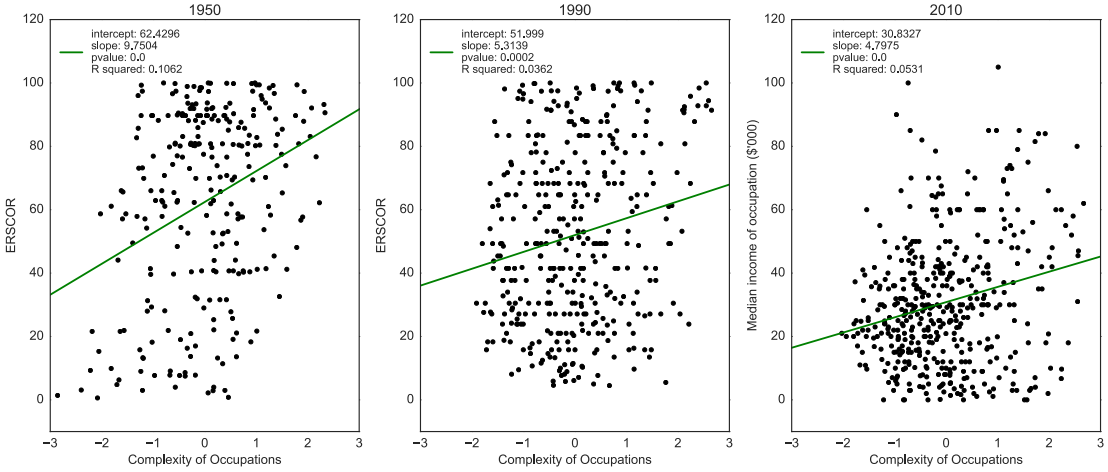


Figure 8: Complexity of occupations vs median earnings for each occupation.

Since our SOC and OC measures are based on employment concentrations in each state, it could also be the case that we are capturing effects related to the sheer number of people employed in different occupations. As such, we investigate the relationship between the total number of people employed in each occupation

[12]ERSCOR50 and ERSCOR90 are IPUMS variables relating to the standardised median income earned by an occupation in 1950 and 1990. ERSCOR is not available for 2010, so we calculate the occupation's standardised median income on the basis of 2010 census data.

and its complexity score in Figure 9. Again, we find limited evidence that OC scores are capturing effects associated with the quantity of people employed in each occupation category.

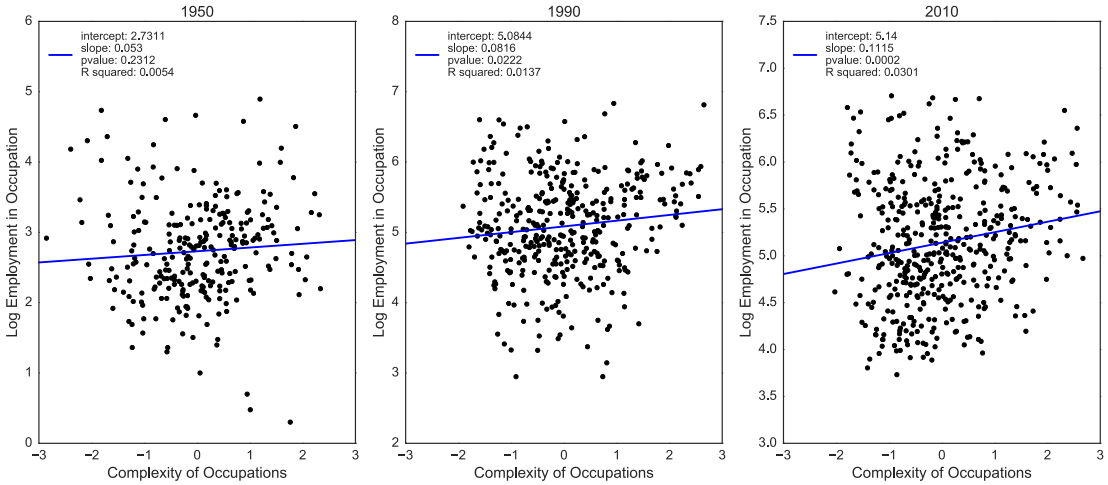


Figure 9: Complexity of occupations vs total employment in each occupation.

Finally, in Figure 10 we consider the relationship between the OC index and educational requirements associated with each occupation. This time, we draw on IPUMS education attainment variables associated with each occupation for 1950, 1990 and 2010. [13]

[13]EDSCOR50 and EDSCOR90 are IPUMS variables relating to the percentage of people in each occupation category that have completed at least one year of college in 1950 and 1990. EDSCOR is not available for 2010, so we use 2010 census data to calculate the percentage of people in each occupational category that have completed at least one year of college

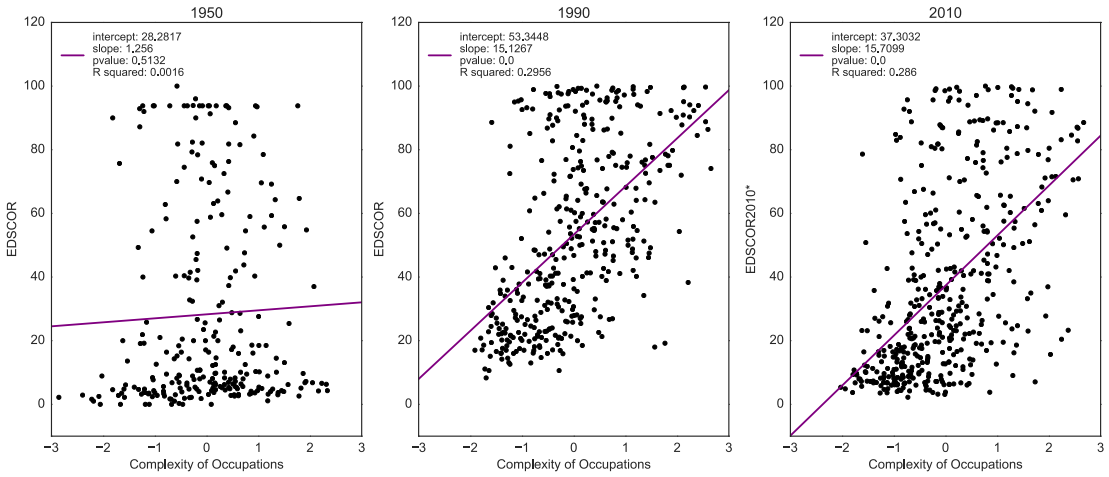


Figure 10: Complexity of occupations vs each occupation's educational requirement.

Interestingly in 1950 (the first year that educational attainment data is available), we find no correlation between an occupation's OC score and its associated educational requirements. However, in more recent years we find that OC is more strongly positively correlated with occupational education requirements, suggesting that over time, OC and skill content have become more closely aligned.

As a final robustness test, we examine how the relationship between OC and OccY changes when we control for each occupation's earnings, employment and educational requirements. As the occupation classification changes in 1950, 1990 and 2010, we cannot undertake a panel regression, but instead estimate a regression model for each classification year:

$$OccY_i = \beta_1 OC_i + \beta_2 wage_i + \beta_3 edu_i + \beta_4 lemp_i + u_i \tag{14}$$

where $wage_i$ relates to the standardised median income earned by a given occupa-

tion, edu_i relates to the percentage people in each occupation category that have completed more than 1 year of college, $lemp_i$ represents the log of the number of people employed in each occupation and u_i represents the error term.

We present the results of these regressions in Table 5. Model (1) and (2) relate to the 1950 classification scheme and compare the change in the OC variable's explanatory power when we introduce earnings, employment and education controls. Model (3) and (4) similarly make comparisons for the 1990 classification while Model (5) and (6) relate to the 2010 classification.

Table 5: Regression analysis of the relationship between the Occupational Complexity Index and Occupational Income Content ($OccY$)

	Dependent variable: OccY (Occupational Income Content)					
	1950		1990		2010	
	(1)	(2)	(3)	(4)	(5)	(6)
Occupational Complexity Index	0.080***	0.080***	0.077***	0.079***	0.083***	0.081***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
Occupational Median Income		−0.00004		−0.00001		0.0000
		(0.0001)		(0.0001)		(0.000)
Occupational Educational Requirement		0.0003***		−0.0001		0.0001
		(0.0001)		(0.0001)		(0.0001)
Log Total Employment		0.004		0.004		0.003
		(0.004)		(0.003)		(0.004)
Constant	7.287***	7.270***	10.036***	10.027***	10.778***	10.760***
	(0.002)	(0.012)	(0.002)	(0.019)	(0.002)	(0.019)
Observations	268	268	383	383	453	453
Adjusted R^2	0.855	0.863	0.740	0.742	0.755	0.754
Note:						*p<0.1; **p<0.05; ***p<0.01

In all cases, the OC index accounts for the overriding majority of explained variance in occupational income content.

4.8.5 Robustness of structural change pattern

In this section, we further scrutinize the observed shift in the OC scores over the 1970-1980 period.

First we test whether the distinct change in the OC scores are also present in the industrial complexity indices based on state industrial employment concentrations. Figure 11 shows the average IC scores for 3-digit industries falling into broad categories over time. Here we find a very similar pattern - with the average IC of manufacturing industries falling significantly and the average IC of information and communication industries experiencing a marked rise. The average IC scores of Professional, scientific, technical and financial industries also increase over this period, though not as sharply as information and communication industries.

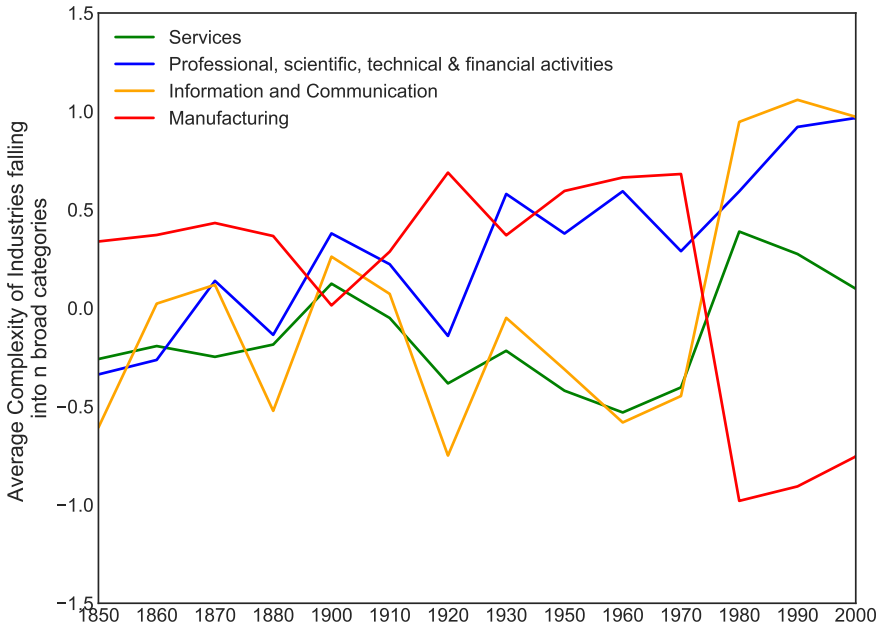


Figure 11: Average Industrial Complexity index scores of broad industry categories over 1850-2010.

In Figure 12, we examine whether this change also occurs if we calculate the OC metrics for metro-level regions rather than state level. While metro-level data was only available from 1970 onwards at the time of writing, we again find a similar change occurring over the 1970-1980 period, suggesting that the dynamics are not an artefact of using state-level data.

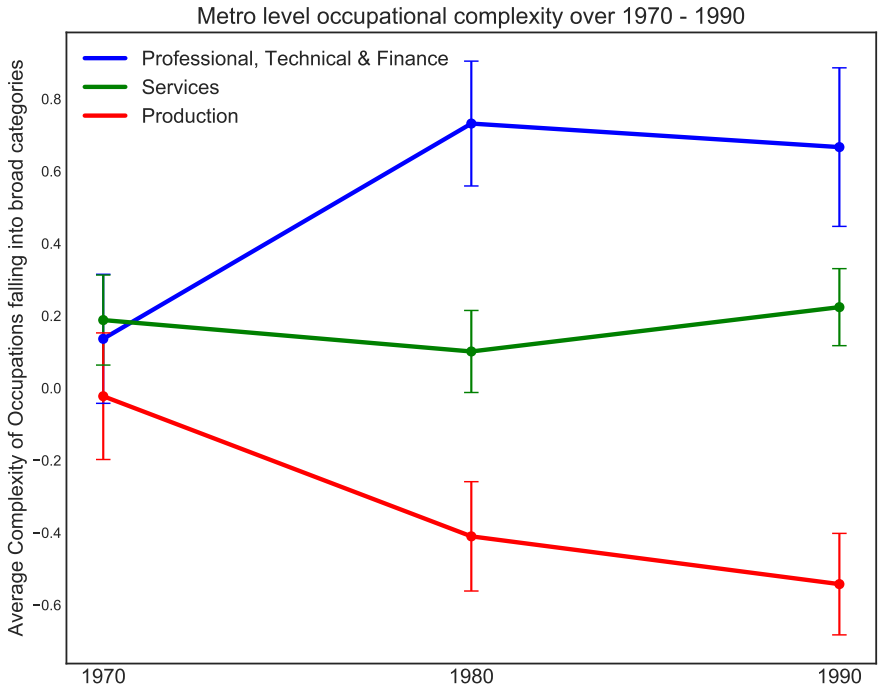


Figure 12: Average complexity of Production, Services and Professional, Technical and Finance Occupations calculated on the basis of metropolitan rather than state regions. Metropolitan data are based on the drawn from the IPUMS database (METAREA variable) (Ruggles et al., 2017), which at the time of writing was only available from 1970 onwards.

Finally, we investigate whether the observed structural shift in OC scores can be accounted for by the change in employment shares of these broad categories over time. In Figure 13, Panel A shows the employment shares falling into the occupational broad categories and Panel B shows the employment shares falling into the

industrial broad categories. While we do observe a decline in employment in Production and Manufacturing categories, and a rise in both services and professional, technical and finance employment, the changes occur relatively steadily over the studied time period. These smoother changes do not account for the distinct shift in OC and IC scores observed over the 1970-1980 period.

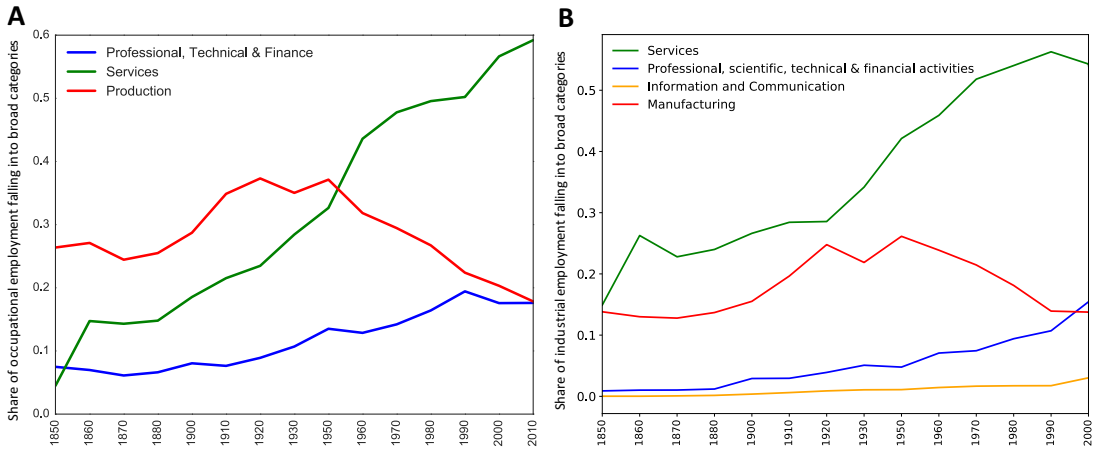


Figure 13: National employment shares over 1850-2010 Panel A shows employment shares in broad occupational categories. Panel B shows employment shares in broad industrial categories.

To summarise, these results suggest that the observed OC dynamics over the 1970-1980 period are not just an artefact of using occupation or state level data. Moreover, they cannot be well accounted for by changes in employment shares over time.

4.8.6 Tradable and non-tradable analysis

Distinguishing between tradable and non-tradable industries or occupations is a notoriously imprecise activity. A number of approaches aim to use an activity's

geographical concentration as a proxy for its 'tradability' (Porter, 2003, Jensen and Kletzer, 2010, Spence and Hlatshwayo, 2012). The key intuition is that traded activities tend to either take advantage of increasing returns to scale or particular inputs such as natural resources or particular types of skills, while non-traded activities tend to be more equally distributed across space, or at least *"follow the geographical distribution of the goods-producing population"* (Krugman, 1991d, p 65).

We apply a similar approach that also draws on the work of Youn et al. (2016) who examined the tendency for the number of establishments in different industries (which tends to be closely correlated to employment) to systematically change with city size. They noted that non-tradable activities tended to scale *linearly* (i.e increase in a linear proportion) with a place's population, while tradable activities either scaled sub-linearly (such as agriculture, mining and utilities) or super-linearly (such as professional, scientific and technical and managerial services). Along these lines, a very simple approach for classifying tradable and non-tradable activities is just to measure the extent to which employment in a given activity is linearly correlated with population.

In Table 6, we show the top 20 and bottom 20 occupations ranked by the R-squared for the relationship between employment in an occupation and total population across states for the year 2010. Top ranked occupations closely align with what one would intuitively expect to be non-traded activities, while bottom ranked occupations tend to relate to more tradable manufacturing or agricultural activities.

In Figure 14 we plot the R-squared values for each occupation falling into a particular 2-digit classification and in Figure 15 we show a histogram of the entire

distribution of R-squared values across occupations in 2010.

Selecting an appropriate cut-off (in this case, the appropriate R-squared value) to distinguish between a traded and non-traded activity is always relatively ad-hoc (Jensen and Kletzer, 2010). As shown in the histogram in Figure 15, a significant proportion of occupations tend to have a R-squared values greater than 0.8, with a long tail of occupations having lower R-squared values. But it is difficult to justify a threshold of 0.8 over 0.7 or 0.9.

The key purpose of this exercise is to determine whether the SOC indices perform better (in terms of more accurately distinguishing between rich and poor places) when non-traded activities are removed. As such, we experiment with three different thresholds (0.7, 0.8 and 0.9) and compare them to the original dataset, where all occupations are included. Specifically, we construct three modified W matrices of binary state occupational employment concentrations in which occupations which have an R-squared greater than the given threshold are omitted. We then calculate new SOC indices on the basis of each modified W matrix and examine the correlation between the resulting SOC index and log income per capita across states.

Figure 16 compares the Pearson R correlation coefficient for the relationship between state income per capita and the new SOC indices calculated for each threshold. We find that the correlation coefficient is highest for the SOC index calculated on the basis of *all* occupations, and declines as we remove additional occupations in accordance with the given thresholds. This indicates that the SOC index is better able to distinguish between rich and poor places when non-tradable occupations are considered.

R2	Occupation	2 digit classification
0.99	First-Line Supervisors of Sales Workers	Sales and related
0.99	Cashiers	Sales and related
0.99	First-Line Supervisors of Office and Administrative Support Workers	Office and administrative support
0.99	General and Operations Managers	Management, business, science and art
0.99	Bookkeeping, Accounting, and Auditing Clerks	Office and administrative support
0.99	Receptionists and Information Clerks	Office and administrative support
0.99	Stock Clerks and Order Fillers	Office and administrative support
0.99	Sales Representatives, Wholesale and Manufacturing	Sales and related
0.99	Food Service and Lodging Managers	Management, business, science and art
0.99	Retail Salespersons	Sales and related
0.99	Automotive Service Technicians and Mechanics	Installation, maintenance and repair
0.99	Office and administrative support workers, nec	Office and administrative support
0.98	Hairdressers, Hairstylists, and Cosmetologists	Personal care and service
0.98	Education Administrators	Management, business, science and art
0.98	Electricians	Construction and extraction
0.98	File Clerks	Office and administrative support
0.98	Managers, nec (including Postmasters)	Business operations specialist
0.98	Accountants and Auditors	Financial specialists
0.98	Transportation, Storage, and Distribution Managers	Management, business, science and art
0.98	Data Entry Keyers	Office and administrative support
...	...	...
0.17	Tool and Die Makers	Production
0.16	Tool Grinders, Filers, and Sharpeners	Production
0.16	Riggers	Installation, maintenance and repair
0.16	Derrick, rotary drill, and service unit operators, and roustabouts, oil, g	Construction and extraction
0.15	Gaming Services Workers	Personal care and service
0.15	Atmospheric and Space Scientists	Life, physical and social science
0.14	Metal Furnace Operators, Tenders, Pourers, and Casters	Production
0.14	Gaming Managers	Management, business, science and art
0.14	Economists and market researchers	Life, physical and social science
0.13	Railroad Brake, Signal, and Switch Operators	Transportation
0.13	Fishing and hunting workers	Farming, fishing and forestry
0.11	Forging Machine Setters, Operators, and Tenders, Metal and Plastic	Production
0.08	Model Makers and Patternmakers, Metal and Plastic	Production
0.08	Rolling Machine Setters, Operators, and Tenders, metal and Plastic	Production
0.07	Mining Machine Operators	Construction and extraction
0.06	Gaming Cage Workers	Office and administrative support
0.04	Textile Knitting and Weaving Machine Setters, Operators, and Tenders	Production
0.03	Tire Builders	Production
0.02	Textile Winding, Twisting, and Drawing Out Machine Setters, Operator	Production
0.01	Explosives Workers, Ordnance Handling Experts, and Blasters	Construction and extraction

Table 6: Top and bottom 20 occupations ranked in terms of the R-squared for the relationship between employment in an occupation and total population across states for the year 2010.

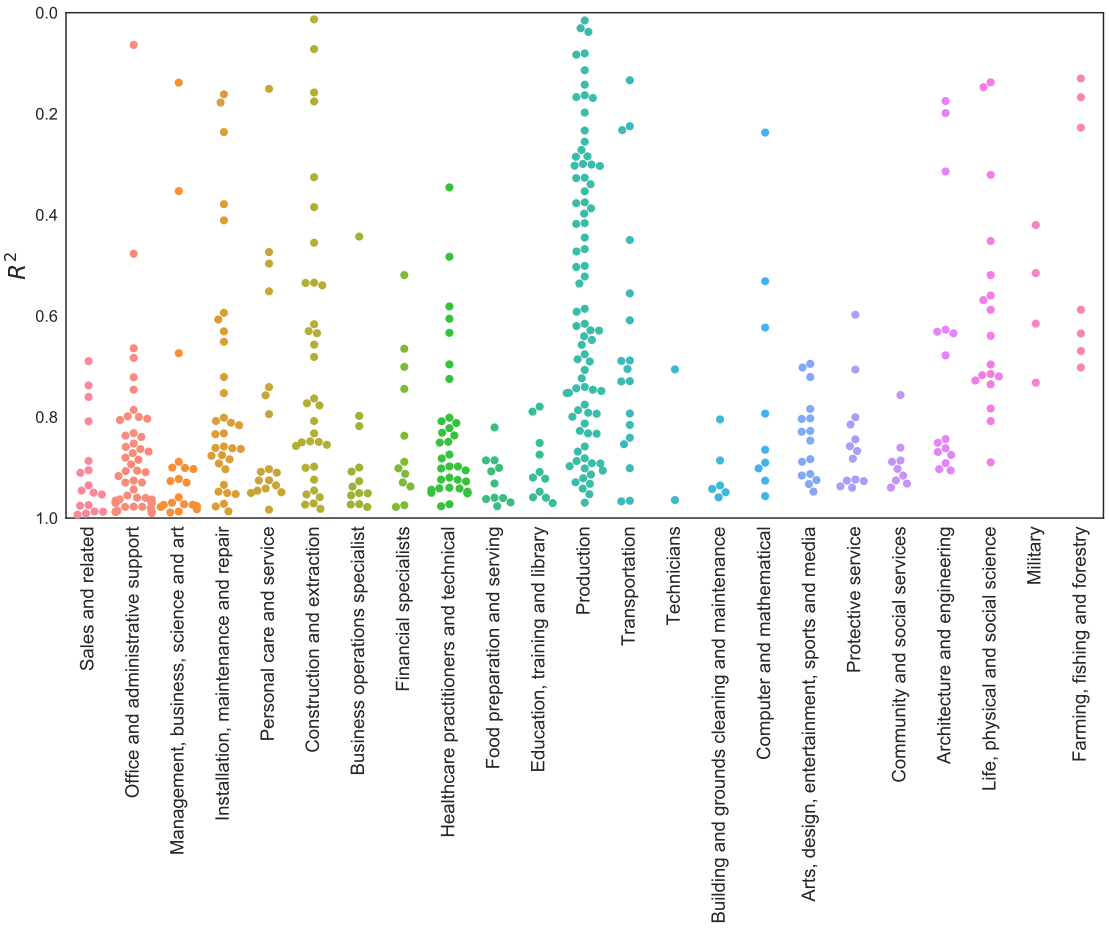


Figure 14: Distribution of R squared values for each detailed occupation within broader 2-digit classifications.

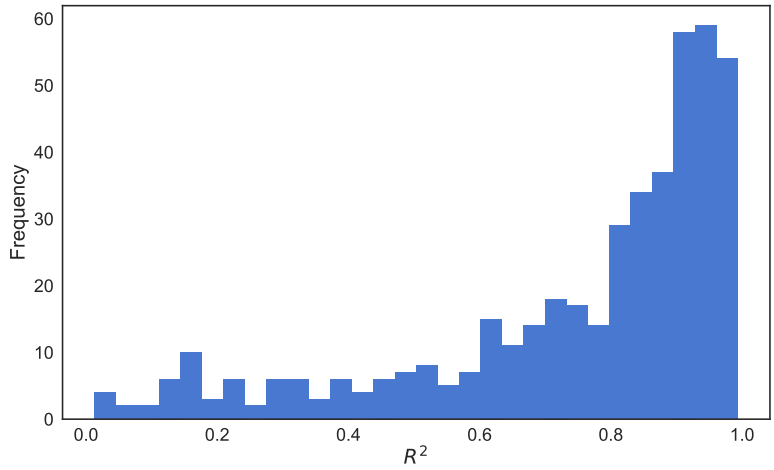


Figure 15: Histogram showing the distribution of R-squared values across occupations for 2010.

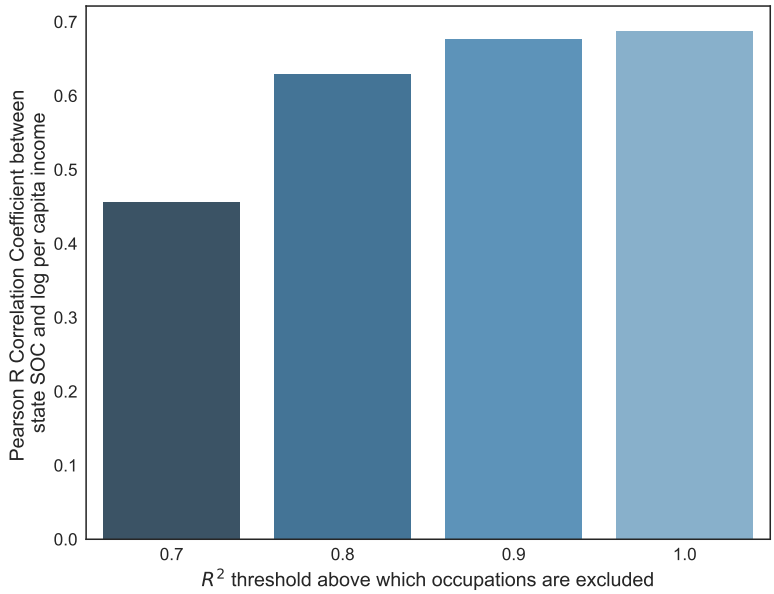


Figure 16: Comparison of the Pearson R correlation coefficient for the relationship between state income per capita and new SOC indices calculated for each threshold.

4.8.7 Comparisons to the Krugman Index

Some studies have utilised the Krugman Index of 'dissimilarity' (Krugman, 1991d, 1993) to examine regional convergence in the US (Kim, 1998) and the UK (Martin et al., 2018). This index quantifies the difference in economic structure between one region s and another region s^* by summing up the differences in their employment shares across sectors and is given by

$$KI_{ss^*} = \sum_i |s_i - s_i^*|, \tag{15}$$

where s_i represents the employment share of region s in sector i.

While this would generate $N \times N$ pairwise comparisons for N regions, the index can also be calculated by comparing each region's employment share s_i to the national average employment share in sector i. In Figure 17, we apply both versions of the Krugman Index to state employment shares over the 1850-2010 period. The orange line shows the average of pair-wise state dissimilarities calculated for each year, while the blue line shows the average of state-dissimilarities to the national average occupational structure. With the exception of a peak in dissimilarity in 2000, the measures have shown a declining trend since 1870, suggesting that the distribution of state employment shares have becoming more similar over time.

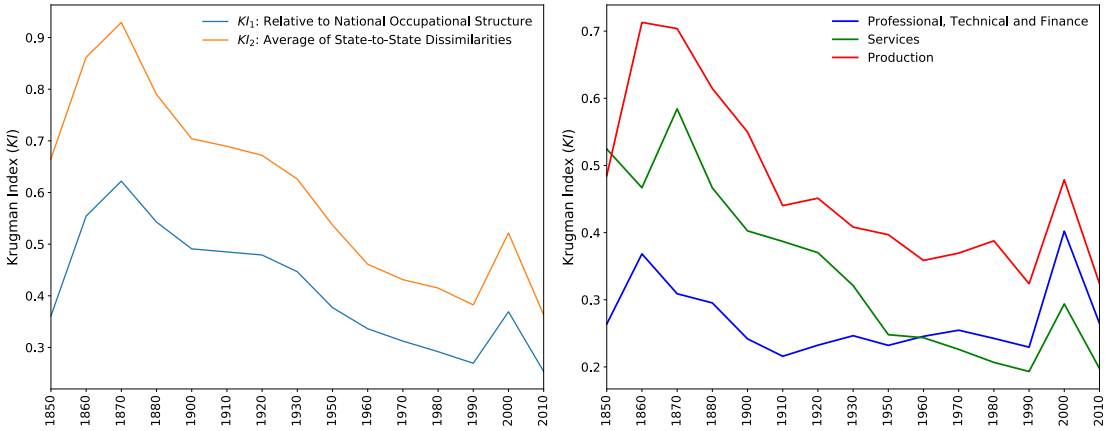


Figure 17: Panel A: Krugman Index for US state occupational employment shares. Panel B: Krugman Index for occupational broad categories.

In Panel B, we calculate the Krugman Index (based on state comparisons to the national average employment share) for the same broad occupational categories shown in Figure 3.

While the same peak occurs in 2000, the trends suggest that states are more dissimilar in relation to production-related occupations, followed by high-skill service professions followed by other services. The decreasing trend in production-related occupations has slowed over time, while the trend for high-skill service professions has been slowly increasing.

While the Krugman Index provides a useful indication of the convergence in distributions of state employment shares, we suggest that the metrics described in this paper could provide an additional complementary lens on the changing nature of regional employment concentrations.

5 What you do at work matters: New lenses on labour

Abstract:[1] How is work distributed across individuals within society? And what can this tell us about career transition possibilities and job switching opportunities? This paper investigates the network structure of the division of labour by analysing discrete work activities that people undertake in different occupations. We find that what people do in their current job matters for their future job - people are significantly more likely to transition into occupations sharing similar work activities. Moreover, we find that our measure of occupational work-activity similarity is more predictive of job-to-job transitions than existing benchmark measures. We also highlight how our new networks-based lenses on labour can illuminate a range of labour market topics, including the gendered division of labour and the future of work.

[1]We are grateful to Mika Straka, Dominic Dent, Renaud Lambiotte and the INET Complexity Research Group for their valuable discussions and feedback in the development and writing of this paper.

5.1 Introduction

How is work organised in society? And how might this influence people's future career options? Although more than two centuries have passed since Adam Smith famously articulated that wealth is linked to the division of labour, there has been relatively little effort to empirically investigate how discrete tasks are distributed across society. Even less is known how specialisation in particular types of work may influence an individual's future career transition possibilities. And yet many individuals are beginning to grapple with the implications of specialising in particular professions. While there is considerable debate about the likely impacts that automation (Autor, 2015, Arntz et al., 2016, Frey and Osborne, 2017), offshoring (Antràs et al., 2006, Blinder et al., 2009) and even environmental policies (Greenstone, 2002, Morgenstern et al., 2002, Walker, 2011) may have on employment, there is a strong consensus that understanding career transition possibilities for displaced workers is likely to become increasingly important.

This paper provides a new lens on the labour force by investigating specific tasks that people working in different occupations undertake. By drawing on data from the Occupational Information Network (O*NET), which stipulates discrete work activities associated with given occupations, we measure the similarity between occupations in terms of the work activities they have in common. We find that people are significantly more likely to transition into occupations that involve similar work-activities to their current profession. We also find that our measure of occupational similarity is more predictive of job transitions than existing benchmarks, and is robust to the inclusion of a variety of control variables.

We then investigate the network structure of the division of labour. We create two related networks. The first is the *Job Space* network, which is a network linking occupations together if they undertake a high degree of similar and specialised tasks. We find nine distinct clusters of occupations in Job Space, which highlight the relatedness in work activities within and across broad Standard Occupational Classification (SOC) 2-digit codes (often referred to as 'job families').

Unlike occupations, no standard classification (such as SOC) exists for classifying the nature of work undertaken by individuals. Therefore, we construct the *Work Activity Space* - a network in which work activities are nodes connected to each other if they are frequently employed in the same occupation. We identify eleven communities of activities and name each cluster by examining the frequency of words in each work activity descriptions.

These network lenses provide a new way to investigate a range of labour market phenomena. Using data on the US labour force, we find stark divides relating to wage, education and gender. Highly skilled and well-paid occupations cluster in a particular region of the Job Space, and there are distinct differences in the work activities undertaken by males and females. While the clustering patterns are relatively stereotypical (for example, females are over-represented in office administration and healthcare, while males are over-represented in construction and engineering), the lack of overlapping work activities (and by extension career-transition possibilities) underscores the difficulty in overcoming social mobility and gender-segregated workforces.

We also demonstrate how this approach complements the task-based literature, which has focused on the extent to which occupations involve abstract, manual and

routine tasks (Autor et al., 2003, Autor, 2013). By visualising each occupation's task intensity score[2] in the Job Space, we provide an additional perspective on the work activity similarity and job transition possibilities within and across abstract, manual and routine task intensity constructs.

Finally we show a very preliminary view into how our new lenses on labour could help address policy questions relating to the future of work. Drawing on estimates of each occupation's susceptibility to automation (Frey and Osborne, 2017), we find strongly segregated clustering of high and low risk occupations in the Job Space. Such divisions in labour not only reiterate current concerns about the distributional consequences of automation, it also highlights potential challenges for workers seeking to transition into jobs with lower automation risk. In contrast, visualising the position of pollution-intensive 'brown' occupations (Vona et al., 2017) in the Job Space paints a more positive picture for workers that may be adversely affected by the transition to a green economy.

5.2 Results

5.2.1 Predicting job transitions

While several studies have analysed career switching across industries (Neffke et al., 2017, Neffke and Henning, 2013, O'Clery et al., 2016) and firms (Guerrero and Axtell, 2013), comparably few studies have examined job transitions at the level of occupations. To characterise the empirical probability of transitioning from

[2]Abstract, manual and routine task intensity scores are based on continuous scale measures, while work activities are discrete activities either are or are not associated with a given occupation

one occupation to another, we draw on monthly panel data from the US Current Population Survey (CPS) for the period January 2010-January 2017. By examining the 4-digit occupation codes of surveyed respondents, we calculate the probability of transitioning from occupation i to occupation j as

$$P_{ij} = \frac{T_{ij}}{T_i}, \tag{1}$$

where T_{ij} is the number of people that switched from occupation i to j, and T_i is the number of people that transitioned out of occupation i over the sampled time period.

We then develop a measure of each occupation i's similarity to every other occupation j in terms of the discrete work activities that occupation i and j have in common. While previous studies on the task-related occupations have employed *continuous-scale* task intensity scores from O*NET and the Dictionary of Occupational Titles (DOT) and (see for example Autor et al. (2003), Acemoglu and Autor (2011), Lin (2011), Michaels et al. (2013)) , here we draw on an alternative O*NET dataset that stipulates 332 *discrete* work activities that people working in different occupations undertake (see Figure 5).

Some work activities, such as 'direct organisational operations' are fairly general and undertaken by many occupations, while other activities, such as 'research healthcare issues' are quite specialised and only associated with a few occupations (see section 5.4.1). We note that O*NET has a wide range of other occupational descriptors, including more detailed work activities. However, as we show in the Appendix (section 5.5), O*NET's intermediate work activities explain the great-

est variance in job transition probability. In addition, as O*NET occupations are more disaggregated than the American Community Survey (ACS) classification of occupations for which we calculate job transition probabilities, we aggregate O*NET occupations to the 4-digit ACS classification following a procedure outlined in section 5.4.2.

Our measure of occupational similarity exploits both the proportion of overlapping work activities between two occupations, and the extent to which these shared work activities are specialist (or scarce) in nature. Specifically, we organise the intermediate work activities for each occupation in a matrix A, where each of the i rows is associated with an occupation and each of the w columns is associated with an intermediate work activity. Letting $A_{iw} = 1$ if occupation i undertakes work activity w and 0 otherwise, we define an occupational similarity measure γ_{ij} as

$$\gamma_{ij} = min(\frac{\sum_w A_{iw}A_{jw}s_w}{\sum_w A_{iw}}, \frac{\sum_w A_{iw}A_{jw}s_w}{\sum_w A_{jw}}) \tag{2}$$

where s_w represents how scarce (or rare) each work activity is, and is given by

$$s_w = \frac{1}{\sum_i A_{iw}}. \tag{3}$$

Similar approaches have been taken to calculate the similarity between exports in terms of their co-exporting countries. In particular, our γ_{ij} functional form is identical to the approach taken by Zaccaria et al. (2014) to construct a 'taxonomy network' of traded products on the basis of countries that exported them. It also

bears resemblance to the approach taken by Hidalgo et al. (2007) to construct the 'product space', but differs in the inclusion of the s_w weight, which allows γ_{ij} to be increasing in the scarcity of work activities that two occupations share.

We find that the similarity in two occupations' work activities γ_{ij} is significantly positively correlated with their corresponding job transition probability (Pearson $\rho = 0.301$, p-value $< 2.2 \times 10^{-16}$), suggesting that people are more likely to transition into occupations that involve similar work activities to their current occupation. While we find that γ_{ij} on its own only explains around 9% of the variation, it is important to note that there are a range of factors influencing job transitions that are not captured by work activity similarity (such as wages, education, location, supply and demand of labour). Moreover, as shown in Figure 1, γ_{ij} performs considerably better than the most relevant benchmark: the O*NET career changers matrix.

The O*NET's career changers matrix aims to identify jobs that people in a target occupation could pursue with minimal additional preparation (Allen et al., 2012). It matches each occupation to 10 related (or similar) occupations the basis of two procedures: (1) an analytical exercise, which considers each occupation's Euclidean distance to other occupations in terms of several O*NET occupational descriptors including work activities, knowledge, skills and work-context, and (2) a 'rational review' exercise, where analysts subjectively modify any inappropriately matched occupations. Despite only drawing on O*NET's intermediate work activities data, Panel a) of Figure 1 shows that γ_{ij} can explain substantially more variance in job transition probability.

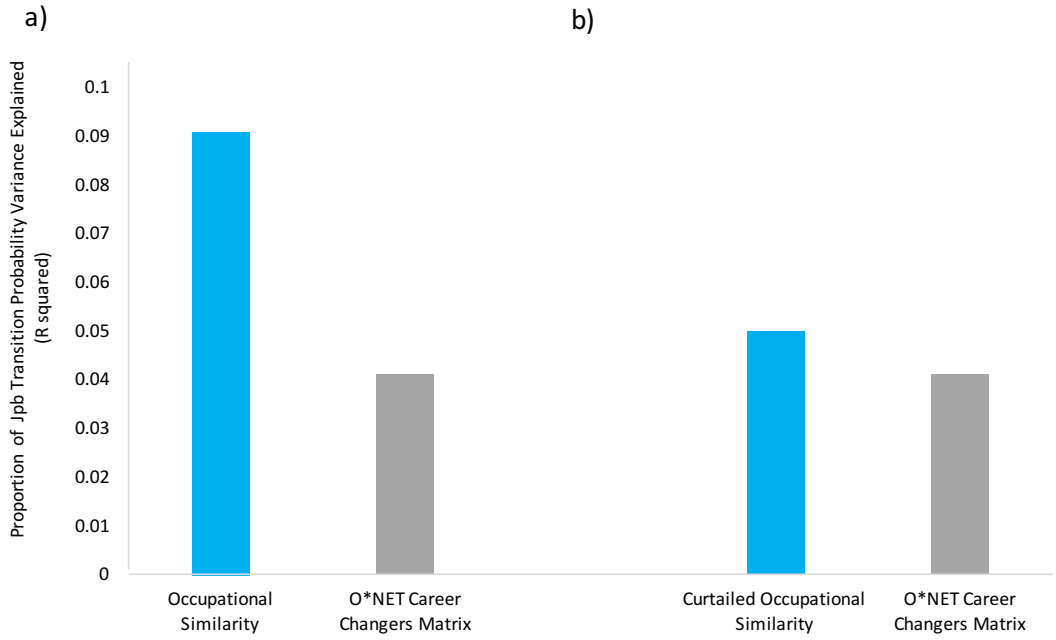


Figure 1: **Panel a):** Comparison in the variance in job transition probability explained by the occupational similarity measure γ_{ij} and the O*NET Career Changers matrix **Panel b):** Comparison in the variance in job transition probability explained by the curtailed occupational similarity measure $\widetilde{\gamma}_{ij}$ and the O*NET Career Changers matrix.

However, this may not be a fair comparison as the O*NET career changers matrix (denoted CC_{ij}) is a binary matrix ($CC_{ij} = 1$ if occupation i is considered to be 'related' to occupation j and $CC_{ij} = 0$ otherwise), which only maps each occupation to 10 other occupations. In contrast, γ_{ij} is a non-binary measure that is defined across all i and j occupation combinations. As such, in Panel b) of Figure 1, we show a comparison between the O*NET career changers matrix and a binary $\widetilde{\gamma}_{ij}$, in which $\widetilde{\gamma}_{ij} = 1$ if occupation j is in the top 10 most proximate occupations to occupation i and 0 otherwise. Here we find that even the curtailed $\widetilde{\gamma}_{ij}$ still outperforms the O*NET career changers matrix. In Appendix 5.5 we make

comparisons with measures based on other O*NET occupational descriptors (such as 'knowledge', 'abilities', 'skills', 'generalised work activities' and 'detailed work activities'). We find that the γ_{ij} of occupational similarity based on intermediate work activities consistently explains the highest proportion of variance in job transition probability by a considerable margin. Detailed work activities are able to explain at most, 3.8% of the variance, while all other measures fail to explain more than 0.05%.

To ensure the relationship between γ_{ij} and job transition probability is robust to the inclusion of relevant control variables, we undertake a regression analysis, shown in Table 1. For each of the regression models, we regress the log of job transition probability on different combinations of explanatory variables and compare their explanatory power. In all models, we also include a SOC 2-digit (job family) dummy to control for occupation-specific effects of job switching for particular job families.

Table 1: Job Transition Regression Results

	Dependent variable:					
	Log Job Transition Probability					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Occupational Similarity	1.045***					1.015***
	(0.033)					(0.032)
Average Education Attainment		−0.0001***			−0.0002***	−0.0001***
		(0.00003)			(0.00004)	(0.00004)
Difference in Education Attainment		−0.00005***			−0.0001***	−0.0001***
		(0.00001)			(0.00002)	(0.00002)
Average Earnings			−0.005***		0.015***	0.010***
			(0.002)		(0.002)	(0.002)
Difference in Earnings			−0.002***		0.008***	0.007***
			(0.001)		(0.001)	(0.001)
Average Employment				0.432***	0.434***	0.410***
				(0.007)	(0.007)	(0.006)
Difference in Employment				0.215***	0.216***	0.218***
				(0.004)	(0.004)	(0.003)
Constant	0.0003***	0.003***	0.002***	0.0005***	0.001***	−0.0005
	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0003)	(0.0003)
SOC 2-digit Dummy?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	215,296	215,296	215,296	215,296	215,296	215,296
Adjusted R^2	0.094	0.0002	0.00005	0.111	0.112	0.201

Note: *p<0.1; **p<0.05; ***p<0.01

Educational attainment, wage and employment data is for the year 2010 and sourced from the IPUMS-USA database (Ruggles et al., 2017).

Educational attainment is based on the IPUMS 'EDUC' variable, which relates to the highest year of school or degree completed by surveyed respondents. Responses are coded from 0= 'NA or no schooling' to 11 = '5 + years of college'. For each occupation, we calculate the average EDUC value of surveyed respondents in a given ACS occupation code. **Wage** is based on IPUMS 'INCWAGE' variable, which relates to the total pre-tax wage and salary income for the previous year. We calculate the average 'INCWAGE' value of surveyed respondents in given ACS occupation. **Employment** is based on the total number of respondents in each ACS occupation. Employment and wage variables are scaled to be in millions.

Model 1 compares the explanatory power of γ_{ij} to the educational attainment of occupations (Model 2), wages (Model 3), and the number of people employed in occupations (Model 4) and all control variables combined (Model 5). In Model 6, we include all explanatory variables and show that γ_{ij} remains a very strong and significant predictor of job transition probability, and can explain variance that is

not accounted for by wage, education or employment variables.

Our regression analysis also highlights additional findings relating to job switching in the US labour force. In Model 2, which includes the average education attainment of occupation i and j and the difference between occupations (occupation j's education minus occupation i's education), the negative and statistically significant coefficient on both variables suggests that job switching is more likely to occur within occupations requiring less education. Moreover, workers are less likely to transition into a job requiring higher education than their current occupation. A similar result is found for Model 3, which looks at occupational wages. Interestingly, the sign of the wage coefficient changes from negative to positive once we include education and employment variables (Model 5). This change suggests that once we account for the educational requirement and total employment in occupations, workers are more likely to switch to occupations that provide higher wages.

Model 4 shows that (not surprisingly) more switching takes place in occupations that employ a large number of people, and workers are more likely to transition into jobs that hire a larger number of people than their current occupation. In contrast to education and wage, total employment accounts for a comparably large proportion of the variance in job transition probability.

5.2.2 The network structure of the division of labour

To examine the structure and arrangement of the division of labour, we construct two related networks. The first is the *Job Space*, which is a network in which

occupations are nodes linked to each other if they have a high degree of specialised work activities in common. The second is the *Work Activity Space*, which instead represents work activities as nodes linked to each other if they tend to be undertaken in the same occupation.[3] In identifying key clusters within these networks, we shed light on how work tends to be divided amongst professions in society - and how professions relate to each other in terms of the similarity in activities that they undertake.

5.2.2.1 The Job Space

To visualise the Job Space network, we follow existing approaches (Hidalgo et al., 2007) and first construct a maximum spanning tree of the γ_{ij} adjacency matrix. We then add additional links with γ_{ij} greater than a given threshold τ. This approach ensures we only connect occupations that share a high degree of specialised occupations. In this case, $\tau = 0.005$ [4], however alternative values of τ give similar results.

We present the Job Space in Figure 2. In Panel a), we colour each node by their SOC 2-digit job families (also labelled in black). In Panel b), we use the Louvain community detection algorithm (Blondel et al., 2008, Bastian et al., 2009) to identify nine clusters of occupations that share similar work activities. Comparing panel a) and b) shows a close association between the job families and identified job clusters in the network. However, the Job Space network clusters provide additional information in providing insights into the occupational relatedness and

[3] These networks can be seen as one-mode projections of the weighted bipartite network associated with A_{iw}.

[4] $\tau = 0.005$ is approximately 1 standard deviation higher than the mean γ_{ij} value

career transition potential *across* job families.

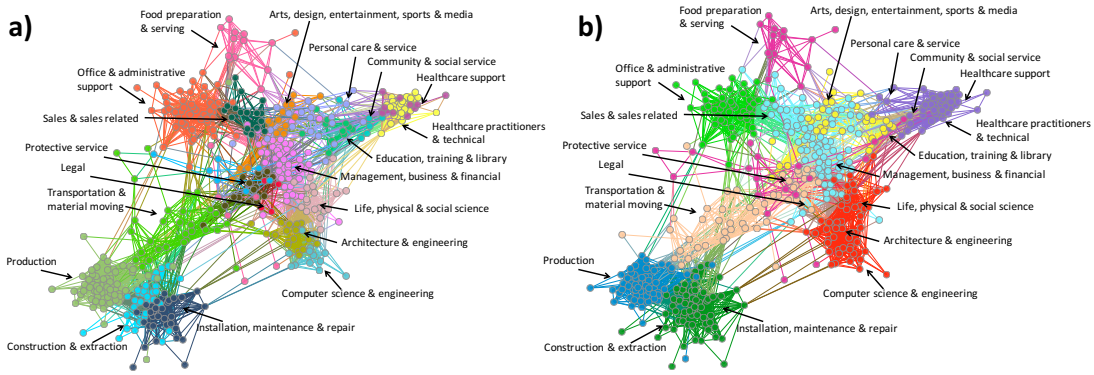


Figure 2: The Job Space: **Panel a)** Nodes are coloured by their 2-digit SOC job family. Job family titles are labelled in black. **Panel b)** Nodes are coloured in accordance with the nine job clusters identified with the Louvain algorithm.

5.2.2.2 The Job Space

While standard classification schemes exist for occupations, no such codification currently exists for classifying the type of work undertaken by individuals. To this end, we construct the Work Activity Space, which is a network of work activities linked to each other if they are more likely to co-occur within occupations.

The Work Activity Space is closely related to the Job Space. We calculate links between work activities by applying the γ formula (equation 2) to the transpose of the A matrix.[5] We then construct the Work Activity Space network using the same maximum spanning tree and thresholding approach that was used to construct the Job Space.[6]

[5]Effectively, we interchange the w and i indices in the γ formula
[6]The threshold t used for the Work Activity Space network is 0.005

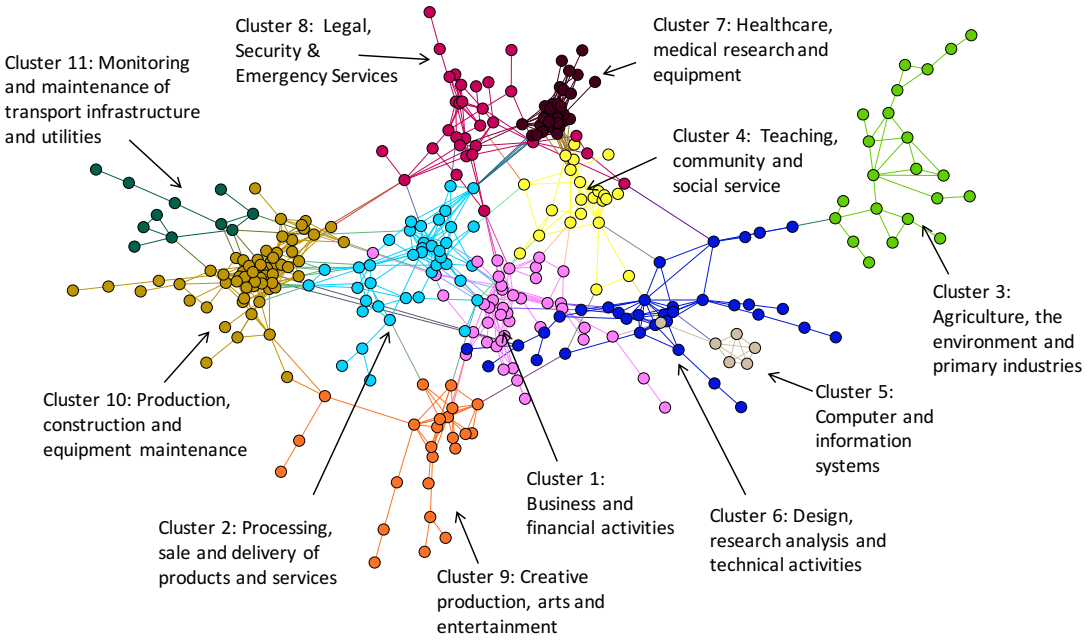


Figure 3: The Work Activity Space: Nodes are coloured in accordance with the 11 identified clusters. Labels are cluster names assigned by inspecting the frequency of words associated with work activity descriptions in each cluster (see section 5.4.3).

Like the Job Space, the Work Activity Space exhibits a high degree of community structure. In Figure 3, we again use the Louvain algorithm (Blondel et al., 2008, Bastian et al., 2009) to identify 11 clusters of work-activities that tend to be undertaken together. As no existing classification exists, we name each cluster by examining the relative word frequency in the work activity descriptions (see section 5.4.3). In showing how society tends to organise work-activities and allocate tasks to individual professions, this network provides a new way to visualise the structure and arrangement of the division of labour.

5.2.3 New lenses on labour

What can our new network lenses tell us about labour and the way work is distributed across society? In Figure 4, we highlight several insights arising from the Job Space network.

Panel a) and b) colour nodes in accordance with the average occupational wage and educational requirements. Here we have drawn on US census data from the Integrated Public Use Microdata Series (IPUMS) for the year 2010 (Ruggles et al., 2017). Highly skilled occupations are fairly tightly clustered in a particular region of the Job Space, suggesting that the type of work highly skilled professions undertake tends to involve fairly similar work activities. Given the well documented positive relationship between education and wages, it is not surprising to see highly paid occupations also clustering in similar locations of the Job Space. However, there are some interesting exceptions, such as the tendency for some mining-related occupations located in the bottom cluster (e.g. derrick, rotary drill, and service unit operators, rustabouts, earth drillers and extraction workers) to receive a comparatively high wage in relation to their average education level.



Figure 4: The Job Space provides a new lens on labour. **Panel a, b and c:** Occupation nodes are coloured in accordance with their average wage, educational requirements and gendered proportions. Data is drawn from the US census data (available from IPUMS (Ruggles et al., 2017)) in 2010. **Panel d, e and f:** Nodes are coloured by their abstract, manual and routine task intensity, based on DOT task measures made available by Autor and Dorn (2013). **Panel g:** Nodes are coloured by their susceptibility of computerisation as estimated by Frey and Osborne (2017). **Panel h:** Node colours denote the set of pollution-intensive ('brown') jobs, which were identified by (Vona et al., 2017) as occupations that are seven times more likely to work in industries in the 95th percentile of pollution intensity.

Panel c) which shows the gendered proportion of US workers in each occupation, highlights the distinct gender-based division of labour in the US workforce. The plot reflects stereotypical divides in work across the sexes. Females are concentrated in occupations relating to office and administrative support, food preparation and serving, healthcare and community service, while males are concentrated in construction, manufacturing and engineering professions. Although sex-based occupational segregation is known to be a persistent phenomenon in many labour markets (Blau et al., 1998, Blau and Kahn, 2000, Preston and Whitehouse, 204, Seron et al., 2016), the marked gendered clustering *across job families* in the Job Space highlights the endemic divisions in the nature of work activities undertaken by males and females. Such divides also suggest that career paths are likely to be fairly male and female-centric, underscoring the systemic difficulties in achieving more gender-balanced work places.

Since we know which work activities are performed by each occupation, we can explore the type of work undertaken at the top and bottom ends of the wage, education and gender spectra in more detail. In Table 2 we rank work activities by calculating the average wage, educational requirement or female proportion of the occupations it is associated with. So for example, if a work activity is undertaken by two occupations i and j, which are paid \$30,000 and \$50,000 respectively, the average wage of the work activity will be \$40,000.

In each column, we find work activities that align closely with expectations about the nature of work at the top and bottom of each spectrum. Highly paid activities relate to legal, analytical managerial work, while low paid activities involve less sophisticated tasks like preparing and providing food an beverages. A similar

pattern is evident for educational attainment, although activities associated with highly educated workers have a stronger slant towards medicine and healthcare.

The gender spectrum of work activities is particularly interesting as it gives us clues for understanding and addressing the persistent gender imbalance across professions. Tasks most commonly undertaken by males tend to relate to work requiring physical strength, where on average, males may have traditionally held a comparative advantage. However, automation and the increasing integration of robots within industry could substantially lower this barrier. On the other end of the spectrum, tasks currently dominated by females tend to relate to healthcare, administration and teaching. The fact that there is no physical barrier to males performing these roles highlights the importance of addressing long-held gendered stereotypes and other socio-cultural factors that exacerbate sex-segregations in the division of labour (England and Li, 2006, Cotter et al., 2011, Seron et al., 2016).

Rank	Wage	Educational Attainment	Female Proportion
1	Draft legislation or regulations.	Make legal decisions.	Assess student capabilities, needs, or performance.
2	Make legal decisions.	Research laws, precedents, or other legal data.	Transport patients or clients.
3	Analyze scientific or applied data using mathematical princi	Diagnose health conditions or disorders.	Prepare health or medical documents.
4	Manage control systems or activities.	Evaluate scholarly work.	Assist individuals with special needs.
5	Design structures or facilities.	Draft legislation or regulations.	Serve on organizational committees.
6	Evaluate project feasibility.	Develop health assessment methods or programs.	Teach life skills.
7	Develop scientific or mathematical theories or models.	Develop scientific or mathematical theories or models.	Prepare medical equipment or work areas for use.
8	Analyze business or financial risks.	Order medical tests or procedures.	Clean medical equipment or facilities.
9	Research laws, precedents, or other legal data.	Advise others on healthcare or wellness issues.	Schedule appointments.
10	Assess characteristics or impacts of regulations or policies.	Teach academic or vocational subjects.	Set up classrooms, facilities, educational materials, or equipment.
[...]	[...]	[...]	[...]
323	Take physical measurements of patients or clients.	Perform general construction or extraction activities.	Operate construction or excavation equipment.
324	Process animal carcasses.	Dispose of waste or debris.	Install energy or heating equipment.
325	Sew garments or materials.	Remove workpieces from production equipment.	Install plumbing or piping equipment or systems.
326	Develop recipes or menus.	Set up protective structures or coverings near work areas.	Repair vehicle components.
327	Hunt animals.	Operate construction or excavation equipment.	Repair electrical or electronic equipment.
328	Train animals.	Cut trees or other vegetation.	Set up protective structures or coverings near work areas.
329	Apply hygienic or cosmetic agents to skin or hair.	Sew garments or materials.	Fabricate devices or components.
330	Prepare foods or beverages.	Operate agricultural or forestry equipment.	Perform general construction or extraction activities.
331	Groom or style hair.	Process animal carcasses.	Climb equipment or structures.
332	Provide food or beverage services.	Hunt animals.	Install commercial or production equipment.

Table 2: Work activities ranked in accordance with the average wage, educational attainment and female proportion of occupations that undertake them.

The Job and Work Activity Spaces also complement the growing literature on tasks. Introduced by Autor et al. (2003), the 'task framework' focuses on how

skills are allocated tasks, and has been used to investigate a wide range of labour market questions relating to the impact of automation and technological change (Autor et al., 2003, Spitz-Oener, 2006, Autor, 2015, Berger and Frey, 2016), offshoring (Antràs et al., 2006, Grossman and Rossi-Hansberg, 2008, Blinder et al., 2009, Blinder and Krueger, 2013)), immigration (Peri and Sparber, 2009, Basso et al., 2017), structural employment shifts (Autor et al., 2006, Goos and Manning, 2007) and the evolution of wages (Acemoglu and Autor, 2011, Autor et al., 2008). While our methodology employs O*NET's dataset on *discrete* work activities, the task-approach literature usually draws on continuous scale task intensity scores from the DOT or O*NET to quantify occupations in terms of the extent they involve manual, routine and abstract (or cognitive) tasks. As shown in panel d), e) and f) of Figure 4, our networks-based approach offers a new way to visualise the distribution of manual, routine and abstract tasks across the professions. Moreover, it has the added advantage of providing additional information on task similarity and job transition possibilities.

Finally, we briefly highlight the potential value of our new lens on labour for addressing policy questions relating to future labour transitions. While it is beyond the scope of this paper to delve into labour market impacts in any detail, we provide a cursory view of two phenomena that will likely affect workers in coming years. The first is automation, and the potential impact that the widespread adoption of computers and robots will have on the labour force. The second is climate policy, and the potential impact that transitioning to a carbon-constrained world may have on jobs in pollution-intensive industries.

There is currently a healthy debate on the potential impact of automation on

employment. Using an innovative method based on machine learning, Frey and Osborne (2017) estimate that 47% of employment in the US workforce has high susceptibility to computerisation, and could be at risk of being automated over the next decade or two. Other studies using related methodologies have arrived at estimates of 9% (Arntz et al., 2016) and 45% (McKinsey, 2015). Since Frey and Osborne (2017)'s estimates are publicly available, we draw on this work to colour each occupation by its estimated probability of computerisation in Panel g) of Figure 4 (see section 5.4.1.4 for more detail about how these probabilities were calculated). It is important to emphasise that an occupation's susceptibility to computerisation does not mean its workers will be necessarily displaced (Autor, 2015, Arntz et al., 2016). Further, Frey and Osborne (2017)'s analysis is silent on the possibilities for new job creation. However, the segregated clustering of highly susceptible occupations in the Job Space underscores existing concerns about the adverse distributional impacts of automation (Autor et al., 2003, Goos and Manning, 2007, Michaels et al., 2014, Hémous and Olsen, 2016). Moreover, the fact that there are such distinct differences in the type of work undertaken by high and low risk occupations highlights the potential difficulties displaced workers may find in transitioning into roles with a lower automation risk.

The impact of environmental policies on employment has also been a subject of debate. Some studies have found environmental regulations have had negative impacts on jobs (Greenstone, 2002, Liu et al., 2017) and wages (Walker, 2011), while others document small positive (Morgenstern et al., 2002) or negligible impacts (Berman and Bui, 2001, Gray et al., 2014). In Panel g) of Figure 4 we preview a different way to look at this question. We draw on the set of 'brown jobs' identified

by Vona et al. (2017), which relate to occupations that have a high probability of being employed in pollution intensive industries. These include occupations like petroleum pump system operators, dredge operators and well pumpers (see section 5.4.1.5 for more detail on how these occupations were identified). Interestingly, we find that most jobs that could be threatened by pollution-reduction policies occupy a relatively localised region of the Job Space. While the geographical concentration of brown jobs could mean that job losses are amplified by local multiplier effects, it is encouraging to see that there appears to be a number of nearby non-brown jobs that share similar work-activities. Further analysis could examine whether these nearby occupations could provide alternative employment for displaced workers.

5.3 Discussion

This paper has advanced a new lens on labour by investigating the network structure of activities that people working in different occupations undertake. We find that the Job Space and Work Activity Space both have informative community structures that highlight similarities (and divisions) in the nature of work undertaken by different professions. Moreover, we find that our measure of occupational similarity is more predictive of job transitions than existing benchmark measures.

Our network lenses provide a new way to analyse a range of labour market topics. In addition to demonstrating how the Job Space can offer a complementary perspective on the growing literature on tasks (Autor et al., 2003, Autor, 2013), we also highlighted the value of our approach for investigating issues ranging from

gender to environmental policy to automation. Future work could explore each of these areas - and others - in more detail.

Beyond labour markets, our work also contributes to the broader literature on path-dependence. A commonly documented finding is that a place (countries, cities and regions) is more likely to develop new specialisations in industries products, technologies or occupations that are similar to what already exists in that place (Hidalgo et al., 2007, Neffke and Henning, 2008, Boschma et al., 2013, Neffke and Henning, 2013, Muneepeerakul et al., 2013, Hausmann et al., 2014, Delgado et al., 2014, O'Clery et al., 2016, Neffke et al., 2017). This study provides empirical evidence to show that this phenomenon also operates on the level of people - individuals are more likely to transition into new jobs that undertake activities that are similar to what people have done in the past.

5.4 Methods

5.4.1 Data

5.4.1.1 Work activities

O*NET work activities database is constructed on the basis of 19,450 unique task descriptions, which are each associated with one of the 974 SOC occupations classified at the 8-digit level. O*NET have aggregated these tasks into two levels of activities: 2,071 'Detailed Work Activities' (DWA) and the more aggregated 332 'Intermediate Work Activities' (IWAs). Examples of tasks, DWAs and IWAs are given in Table 3.

Element	Description
IWA	• Direct organizational operations, activities, or procedures.
DWA	• Direct sales, marketing, or customer service activities. • Direct organizational operations, projects, or services.
Task	• Appoint department heads or managers and assign or delegate responsibilities to them. • Direct or coordinate activities of businesses or departments concerned with production, pricing, sales, or distribution of products. • Direct non-merchandising departments, such as advertising, purchasing, credit, or accounting.

Table 3: Example of task, detailed work activity (DWA) and intermediate work activity (IWA) descriptions.

While tasks are occupation-specific (each task is associated with exactly one occupation), DWAs and IWAs have some degree of overlap across occupations. Figure 5 shows an example of the intermediate work activities (in green) associated with different occupations. More general work activities, (such as 'direct organisational operations') overlap across many occupations, while other specialised work activities (such as 'research healthcare issues') only relate to a few occupations. As we show in Figure 6, the number of work activities that occupations undertake is also reasonably heterogeneous, with some occupations such as engineering technicians linked to 78 IWAs, while other occupations such as medical transcriptionists are only linked to 5 IWAs.

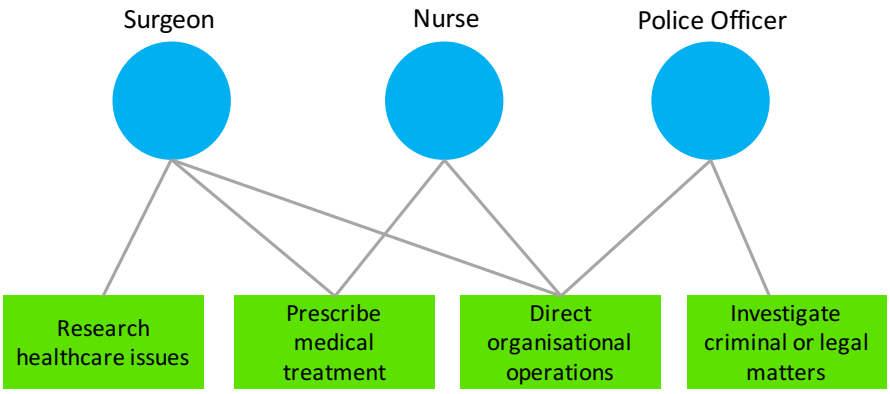


Figure 5: Illustration of intermediate work activities associated with different occupations.

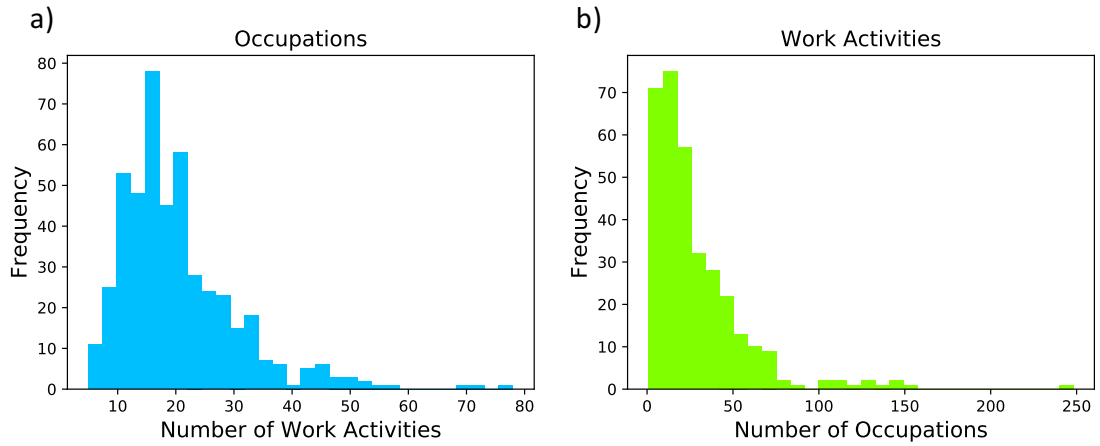


Figure 6: **Panel a):** Distribution of the number of work activities associated with occupations. **Panel b):** Distribution of the number of occupations associated with work activities.

5.4.1.2 Job transitions data

Data on job transitions are drawn from the US Current Population Survey (CPS), which is available from IPUMS-CPS. We use monthly panel data on the occupations of surveyed respondents over the period January 2010-January 2017. The CPS surveys around 60,000 US households every month in a 4-8-4 sampling scheme. This means that households participate in the survey for four consecutive months, are removed from the sample for eight months, and then return for four months before being removed from the sample permanently.

5.4.1.3 Abstract, manual and routine task intensity data

Data on abstract, manual and routine task intensity scores for each occupation (shown in Panel d, e and f of Figure 4) is drawn from Autor and Dorn (2013)'s

analysis of the Dictionary of Occupational Title (DOT) task constructs. Abstract task intensity scores are based on the simple average of the DOT task variables DCP (Direction, control and planning of activities) and GED-MATH (quantitative reasoning requirements). Routine task intensity is calculated as the average of STS (adaptability to work requiring set limits, tolerances, or standards) and FINDEX (finger dexterity). Manual task intensity is captured by EYEHAND (eye, hand and foot coordination).

5.4.1.4 Automation probability data

Data on automation probability is sourced from Frey and Osborne (2017) estimates of each occupation's susceptibility of computerisation. These were calculated by first subjectively hand-labelling 70 occupations as 1 if automatable and 0 if not. They then drew on machine learning to classify the automatability of 702 occupations using the assigned labels and several feature vectors based on continuous-scale O*NET variables corresponding to bottlenecks to computerisation (variables related to perception and manipulation, creativity and social intelligence).

5.4.1.5 Brown jobs data

Data on pollution intensive jobs is drawn from Vona et al. (2017). They first identified 62 4-digit NAICS industries in the 95th percentile of pollution intensity (measured in terms of emissions per worker) for at east three of the following pollutants: CO_2, CO, VOC NO_x, SO_2, PM10, PM2.5 and lead. They then identify 'brown' jobs as occupations that have a probability of working in polluting

industries seven times higher than other sectors.

5.4.2 Aggregation procedure

The IPUMS CPS data on job transitions are available at the American Community Survey (ACS) 4-digit occupational level, while the O*NET data is available at the SOC 8-digit level. To examine the relationship between ONET occupational descriptors and job transition probability, we map the O*NET 8-digit occupations to the more aggregate ACS 4-digit occupations.

We first aggregate SOC 8-digit codes (974 occupations) to 6-digit codes (775 occupations) using the O*NET-SOC2010 occupations to 2010 SOC occupations available from https://www.onetcenter.org/crosswalks.html. We then use the 2016 National Employment Matrix/SOC to ACS Crosswalk available from https://www.bls.gov/emp/ep_crosswalks.htm to map each O*NET 6-digit occupation to the ACS 4-digit occupation. Each 6 digit SOC occupation is linked to exactly one 4-digit ACS occupation.

We specify intermediate work activities associated with each ACS occupation as the union of intermediate work activities associated with its corresponding SOC occupations in the SOC-ACS occupation mapping. That is, we assume that a 4-digit ACS occupation undertakes all the work activities associated with its corresponding 8-digit SOC occupations.

We use a similar approach when aggregating the automation probability and brown jobs data to the IPUMS 4-digit level. As (Frey and Osborne, 2017)'s automation probability estimates are available at the SOC 6-digit level, we use the 2016 Na-

tional Employment Matrix/SOC to ACS Crosswalk and assign an automation probability to each ACS 4-digit occupation based on the simple average of the SOC 6-digit occupation automation probabilities it is linked to. Similarly, as Vona et al. (2017)'s brown job identification is available for 6-digit SOC occupations, we also use the 2016 National Employment Matrix/SOC to ACS Crosswalk and identify a ACS 4-digit occupation as brown if it is linked to at least one brown 6-digit SOC occupation. (Assigning the status on the basis of the majority of the 6-digit SOC occupations has very little impact on our results).

5.4.3 Naming work activity clusters

O*NET Intermediate work activities are a one-line description containing a single verb (see Table 3 for example). To name the Work Activity Space clusters, we examine the frequency of words in the description of work activities associated with each identified cluster. Specifically, we remove all stop-words (commonly used words such as 'the' or 'and'), and consider the top five most frequent words arising in each cluster. While no classification currently exists for tasks undertaken by individuals in occupations, our approach provides one possible methodology for providing a taxonomy of work undertaken by society.

5.5 Appendix

5.5.1 Comparison of Alternative functional forms and different O*NET descriptors for predicting job transitions

In constructing the measure of occupational similarity used in this paper, we also considered different functional forms and alternative occupational descriptors available on the O*NET database. Here we compare how each functional form, applied to alternative O*NET occupational descriptors, performs in explaining variance in job transition probability. We show that the occupational similarity measure (γ_{ij}) based on intermediate work activities has a much higher ability to predict job transitions than the considered alternatives.

5.5.2 O*NET occupational descriptors

O*NET is an online database providing a vast amount of information on occupational characteristics. It collects and maintains its database surveying expert analysts and job incumbents across a sample of US firms. We explored to types of O*NET occupational descriptors: (1) discrete descriptors, which describe the activities undertaken by each occupation and (2) continuous scale descriptors, which assign an importance rating on a continuous scale indicating how important the given element to a given occupation.

5.5.2.1 O*NET discrete descriptors on tasks and work activities

O*NET maintains a list of $19,450$ specific task descriptions which are uniquely

associated with a particular 8-digit SOC occupation. As part of the O*NET Work Activities project (Hansen et al., 2014), these tasks were aggregated into two higher level descriptions of 2,071 'detailed work activities' (DWAs) and 372 'intermediate work activities' (IWAs). Unlike task descriptions, DWAs and IWAs are not unique to each occupation, but have some overlap - so occupations can have DWAs and IWAs in common with other occupations.

5.5.2.2 O*NET continuous scale occupational descriptors

We also examined O*NET's 'knowledge', 'abilities', 'skills' and 'generalised work activities' occupational descriptors. These descriptors each contain a unique set of elements (for example, knowledge encompasses elements such as 'Economics and accounting' or 'Sales and marketing', while abilities include elements such as 'Deductive reasoning' and 'Number facility'). For each occupation, continuous scale importance ratings are assigned to each element, signifying how important it is to a given occupation. Importance ratings range from (1-5).

5.5.3 Occupational similarity measures

We experiment with three different functional forms for measuring the similarity between occupation i and j.

- ω_{ij} Scaled Pearson correlation measure (used in Hausmann et al. (2014))
- ϕ_{ij} Proximity measure (used in Hidalgo et al. (2007)), which considers the proportion of elements that two occupations have in common

- γ_{ij} Scarcity-adjusted proximity measure (used in Zaccaria et al. (2014)) which is similar to ϕ_{ij}, but includes an additional weight reflecting the scarcity of elements that two occupations have in common.

5.5.3.1 Scaled Pearson correlation measure

The Scaled Pearson correlation measure is given by

$$\omega_{ij} = \frac{(1 + corr(Y_i Y_j)}{2}). \tag{4}$$

For continuous-scale descriptors, Y_i represents the vector of importance values assigned to occupation i across all descriptor elements, while for discrete descriptors, Y_i is a binary vector of activities which = 1 if an occupation undertakes the activity and 0 otherwise.

This metric is scaled so that characteristic vectors that are perfectly negatively correlated will have a value of 0, and perfectly positively correlated vectors will have a value of 1.

5.5.3.2 Proximity measure

For discrete descriptors the Proximity measure is given by

$$\phi_{ij} = min(\frac{\sum_w A_{wi} A_{wj}}{\sum_w A_{wi}}, \frac{\sum_w A_{wi} A_{wj}}{\sum_w A_{wj}}) \tag{5}$$

where the A_i column of the matrix can be seen as a binary vector of w work

activities in which elements = 1 if occupation i undertakes work activity w and 0 otherwise.

To apply the Proximity measure to the set of continuous-scale descriptors, we create binary versions of the continuous scale importance ratings. Here, we introduce a 'relative importance indicator' (RII_{ie}) which represents the ratio of the importance of descriptor element e compared to the average importance ratings given to all descriptor elements for a given occupation i and the average importance ratings given to the descriptor element e compared to the average importance ratings across all occupations and elements of that descriptor.

$$RII_{ie} = \frac{M_{ie}/\langle M_i \rangle}{\langle M_e \rangle / \langle M \rangle}, \tag{6}$$

where M_{ie} is the importance rating assigned to descriptor element e for occupation i, $\langle M_i \rangle$ is the average importance rating across all elements of a given descriptor for an occupation i, $\langle M_e \rangle$ is the average importance rating across all occupations for the descriptor element e and $\langle M \rangle$ is the average importance rating assigned across all occupations and elements within a given descriptor (like 'knowledge').

We then create binary descriptor vectors for a given occupation i by assigning a 1 to the occupation-element pair if $RII_{ie} > 1$ and 0 otherwise. On the basis of these binary descriptors, we can then apply the ϕ_{ij} measure described in equation 5.

5.5.3.3 Scarcity-adjusted proximity measure

While the Proximity measure is relatively straight-forward and intuitive, it is lim-

ited in that it does not account for the heterogeneous prevalence of particular work activities (or the relative importance of continuous-scale descriptors). For example, some 'generalist' work activities, such as 'maintain operational records' or 'direct organizational operations' are undertaken by a large number of occupations, other 'specialist' work activities, such as 'draft legislation or regulations' are only taken by a small number of occupations. As two occupations having a 'specialist' work activity in common may be considered to be more closely related than two occupations having a 'generalist' work activity in common, we propose an alternative scarcity-adjusted proximity measure, which weights the proximity between two occupations by the scarcity of the work activities that they share in common.

Specifically γ_{ij}, which represents the weighted occupational relatedness between occupations i and j is given by

$$\gamma_{ij} = min(\frac{\sum_w A_{iw}A_{jw}s_w}{\sum_w A_{iw}}, \frac{\sum_w A_{iw}A_{jw}s_w}{\sum_w A_{jw}}), \tag{7}$$

where s_w represents how scarce (or rare) each work activity is, and is given by

$$s_w = \frac{1}{\sum_i A_{iw}}. \tag{8}$$

To apply this scarcity-adjusted proximity measure to the scale-continuous descriptors, we apply the same methodology of converting the descriptors into binary values as discussed in the previous section.

5.5.4 Correlation with job transition probability

Applying each of the three functional forms to DWAs, IWAs, skills, abilities, knowledge, generalised work activities and an '*all* continuous descriptor' variable which encompasses all of the continuous scale variables (skills, abilities, knowledge and generalised work activities) gives 21 possible occupational similarity metrics. In this section, we test which measure has the greatest ability to explain the variance in job transition probability.

As discussed in the main paper, we characterise the empirical probability of job switching by drawing on monthly panel data from the US Current Population Survey (CPS) for the period January 2010-January 2017. By examining the 4-digit occupation codes of surveyed respondents, we calculate the probability of transitioning from occupation i to occupation j as

$$P_{ij} = \frac{T_{ij}}{T_i}, \tag{9}$$

where T_{ij} is the number of people that switched from occupation i to j, and T_i is the number of people that transitioned out of occupation i over the sampled time period.

Table 4 shows that the intermediate work activities descriptor has the highest correlation with the job transition probability variable - across all three measures. Applying the Scarcity Adjusted Proximity measure to IWAs gives the highest correlation, with a Pearson R of 0.3. Detailed work activities give the next highest set of correlations, while the continuous scale variables - including the variable including *all* continuous scale descriptors tend to correlate fairly weakly across all

measures. All variables are statistically significant at the 1% level.

Table 4: Correlation of each metric with job transition probability

Metric	Pearson R
Intermediate Work Activities	
Scaled Pearson Measure	0.234
Proximity Measure	0.240
Scarcity Adjusted Proximity Measure	0.301
Detailed Work Activities	
Scaled Pearson Measure	0.165
Proximity Measure	0.176
Scarcity Adjusted Proximity Measure	0.196
Skills	
Scaled Pearson Measure	0.041
Proximity Measure	0.012
Scarcity Adjusted Proximity Measure	0.027
Abilities	
Scaled Pearson Measure	0.031
Proximity Measure	0.026
Scarcity Adjusted Proximity Measure	0.034
Knowledge	
Scaled Pearson Measure	0.060
Proximity Measure	0.036
Scarcity Adjusted Proximity Measure	0.060
Generalised Work Activities	
Scaled Pearson Measure	0.063
Proximity Measure	0.010
Scarcity Adjusted Proximity Measure	0.022
All Continuous Scale Descriptors combined	
Scaled Pearson Measure	0.044
Proximity Measure	0.015
Scarcity Adjusted Proximity Measure	0.036

5.5.5 Correlation across different occupational similarity metrics

In Figure 7, we show the Pearson correlation between all occupational similarity metrics. Three things are worth mentioning. Firstly and not surprisingly, each descriptor tends to have high internal correlation across the three different proximity functional forms. Secondly, the occupation similarity metrics based on continuous scale descriptors tend to be fairly highly correlated with each other. This means that occupations that have similar knowledge will also have similar abilities and skills etc. Thirdly, the occupation similarity metrics based on detailed and intermediate work activities descriptors are not very strongly correlated to the continuous scale proximity metrics. This result suggests that these detailed and intermediate work activity descriptors capture different information that, as shown by Table 4, is particularly relevant for predicting job transitions.

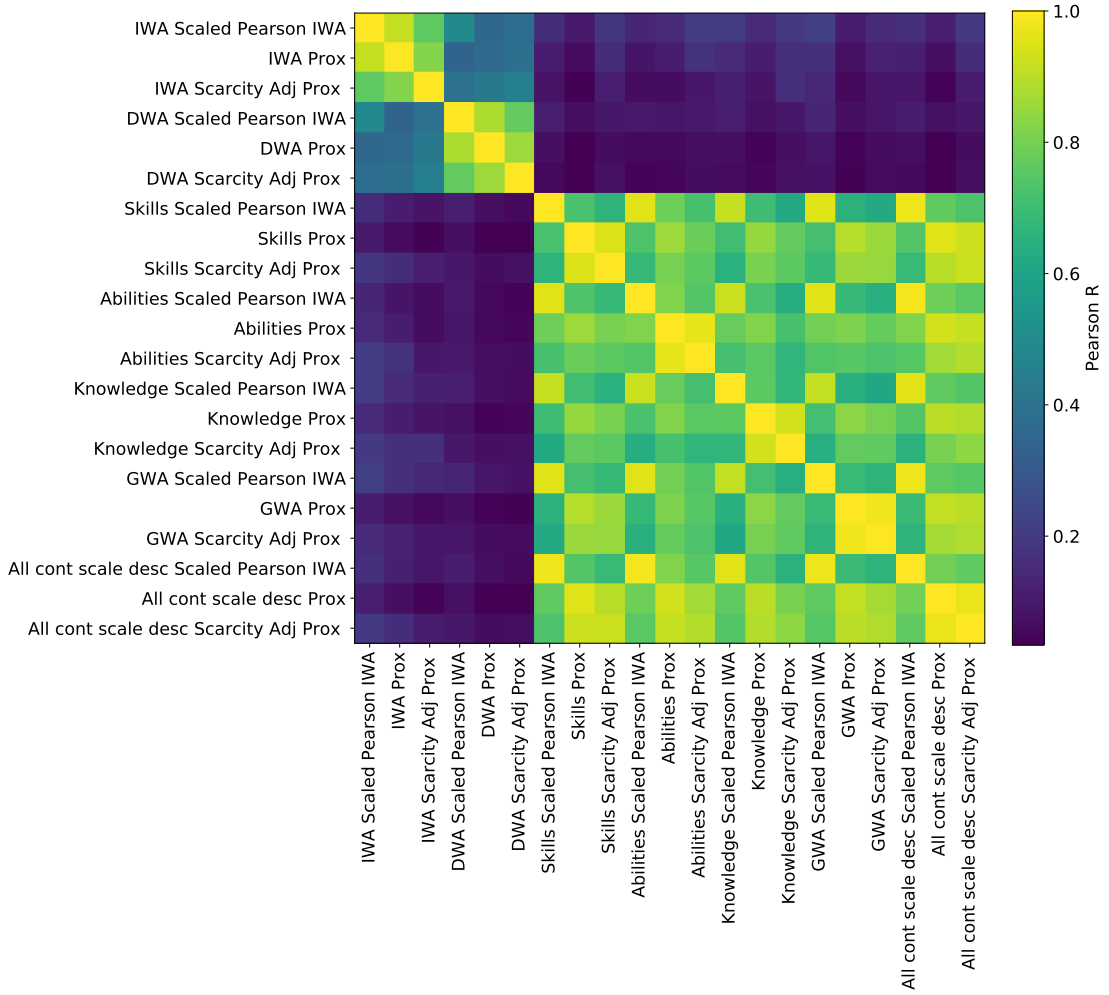


Figure 7: Heatmap showing Pearson R correlation between all occupation proximity descriptor metrics.

6 Economic Complexity and the Green Economy

Abstract:[1] This paper studies countries' productive capabilities relevant to the green economy.We pool all existing environmental goods classifications to create a new, comprehensive dataset of 293 green products traded between 1995 and 2014. We match these products to export data and analyse resulting patterns in the trade network using the *economic complexity* methodology. We construct the *Green Complexity Index* (GCI), which ranks countries in accordance with the number and complexity of green products they export competitively and show that countries with higher GCI have higher environmental patenting rates, lower CO_2 emissions, and more stringent environmental policies. We then look at each country's *Green Adjacent Possible* (GAP), which represents the set of technologically proximate green products that a country could potentially become competitive in. Using the GAPs, we construct a measure of each country's *Green Complexity Potential* (GCP), which we show is predictive of countries' future competitiveness in green products. The strong correlation between GCP and GCI suggests path dependence in the accumulation of green production capabilities and shines a light on discussions around green industrial policy.

[1]We would like to thank Simon Angus, Diane Coyle, Neave O'Clery, Cameron Hepburn, César Hidalgo, Helen Schweiger, Rick van der Ploeg, and seminar participants at the Oxford OxCarre seminar for helpful conversations and feedback.

6.1 Introduction

The transition to the green economy will undoubtedly involve a transformation of production structures and economic activities around the world. This shift has the potential to alter the global competitive landscape and reshape countries' comparative advantages in production (Fankhauser et al., 2013). In this paper, we tackle two questions: (1) which countries are currently best placed to become leaders in the green economy, and (2) how might countries re-orient their industrial structure to become more competitive in green (or environmentally friendly) products.

To address these questions, we draw on country export data to systematically estimate countries' current *capabilities* (e.g. productive know-how, market and institutional infrastructure) and future *potential* (diversification opportunities) for success in the green economy.

A key challenge in measuring green production capabilities on the basis of exports is the current lack of a universally accepted definition of environmental goods and services. Although efforts have been underway since 2001 when the World Trade Organisation (WTO) instigated a mandate to reduce or eliminate tariffs on environmental goods and services, progress towards this has slowed since the 2016 negotiations on the Environmental Goods Agreement stalled in December 2016.

To address these data gaps, we pool all existing environmental goods classifications available from the WTO, the Organisation for Economic Cooperation and Development (OECD), and the Asia-Pacific Economic Cooperation (APEC) into a comprehensive dataset of green traded products. Merging this with export data allows us to analyse country trade in green products over the 1995-2014 period.

On the basis of this dataset, we investigate countries' green production capabilities by drawing on the economic complexity methodology (Hidalgo et al., 2007, Hidalgo and Hausmann, 2009, Hausmann et al., 2014).[2] This approach aims to infer information about countries' productive capabilities and industrial structure by making relative comparisons across country export baskets. We first calculate the Product Complexity Index (PCI) for each green product. The PCI aims to estimate the complexity of capabilities required to export a product *competitively*,[3] and has been used as a measure of technological sophistication (see, for example, Felipe et al., 2012, Poncet and de Waldemar, 2013). We find that green products relating to concentrated solar power, environmental monitoring and analysis tend to have the highest PCI scores, while environmentally preferable products (many of which are made from vegetable material) receive lower PCI scores. Interestingly, we find that green products tend to be more complex than average – suggesting that part of the challenge in addressing climate change may relate to the fact that many clean energy technologies are fairly sophisticated and require substantial investment and expertise to scale for production.

To estimate which countries are currently best placed to become leaders in the green economy, we develop a new measure, called the Green Complexity Index (GCI), which is increasing in the number and complexity of green products that a country is competitive in. It is important to distinguish the GCI from the Economic Complexity Index (ECI) (Hidalgo and Hausmann, 2009, Hausmann et al.,

[2]This methodology has received significant attention in recent years and there is a healthy scientific debate about the best way to infer capabilities relevant to economic growth from export data.(Tacchella et al., 2012, 2013, Cristelli et al., 2013, Zaccaria et al., 2014, Cristelli et al., 2015, Albeaik et al., 2017, Gabrielli et al., 2017, Albeaik et al., 2017)

[3]We say that a country is *competitive* in a product if its revealed comparative advantage (RCA) for this product is greater than 1 (Balassa, 1965). See Equation 2 below.

2014). The ECI estimates the complexity of countries' production capabilities on the basis of the similarity in country export baskets (Kemp-Benedict, 2014, Mealy et al., 2017). The measure is calculated using a clustering algorithm (a dimension reduction method) that sorts countries into an ordering that places countries with similar exports near each other. What makes this ordering interesting is that it can explain more variation in income per capita and economic growth than other variables traditionally employed in growth equations.

The GCI differs from the ECI in two important ways. First, while the ECI is estimated on the basis of all traded products, the GCI is estimated on the basis of a *subset* of *green* traded products. Second while the ECI represents the *average* PCI of all products a country is competitive in, the GCI represents the *sum* of the PCI of all green products a country is competitive in. Note that while we have applied the GCI to a specific subset of green traded products, the measure is completely general and can be applied to any subset of products (such as biotech).

We rank countries according to the GCI and show that while it is positively correlated with the ECI, the GCI is able to explain more variation in environmentally-relevant variables. In particular, we find that controlling for countries' per capita GDP and ECI, countries with higher GCI tend to have significantly higher environmental patenting rates, lower CO_2 emissions and more stringent environmental policies.

We investigate country GCI rankings over the 1995-2014 period, and find relatively little variance in the highest rankings – top ranked countries such as Germany, Italy, US and Denmark have managed to remain leaders over the 20 year period. Some countries such as China, Vietnam and Uganda have made significant gains

in their GCI scores, while other countries such as Australia show notable declines.

Turning to the second question of how countries may re-orient their industrial structure to become more competitive in green products, we exploit the fact that industrial development tends to be path-dependent. Countries or regions are more likely to diversify into products or industries that require production capabilities similar to what they currently possess (Patel and Pavitt, 1997, Weitzman, 1998, Hidalgo et al., 2007, Neffke et al., 2011a, Boschma et al., 2013).

Drawing on measures developed in Hidalgo et al. (2007), which estimate the similarity in production capabilities between two products *(proximity)* and between a country and a product *(proximity density)* by considering countries' conditional probability of being competitive in one product given competitiveness in another, we construct each country's *Green Adjacent Possible (GAP)*. The GAP shows the set of green products that are proximate to a country's current production capabilities – i.e. the new green industrial opportunities that are likely to be the easiest to transition into, given what a country already knows how to do.

We then aggregate the information contained in each country's GAP into a single, comparable metric, which we call Green Complexity Potential (GCP). GCP measures each country's average proximity to complex green products that it is not yet competitive in. We show that GCP is able to significantly predict future increases in a country's GCI, green export share and the number of green products that a country is competitive in, even after controlling for each country's GDP per capita and ECI. We also find a strong positive correlation between GCP and GCI suggesting that countries that currently export a significant number of green complex products are generally well placed to diversify into other green complex

products in the future.

Our results offer several contributions to policy discussions. First, we establish an extensive set of products relevant to the green economy and identify which countries currently have the capabilities to produce them competitively. Our findings complement Fankhauser et al. (2013)'s related efforts to analyse "who will win the green race", but provides a broader coverage of countries and an alternative analytical lens based on the economic complexity methodology. Second, our GCI measure allows policy makers to assess a country's green production capabilities relative to other countries and also consider how its green competitiveness has been changing over time. The path dependence in green diversification suggests that earlier and more aggressive action to establish green production capabilities is required in order to succeed in the future green economy (Aghion et al., 2014, 2016). Third, by identifying the GAPs we show clearly which products countries are best placed to gain a competitive edge in, informing policymakers about the optimal direction of government interventions and green industrial policy (Aghion et al., 2011, Huberty and Zachmann, 2011, Hallegatte et al., 2013, Rodrik, 2014). We also present some preliminary results on the effect of recent green stimulus policies on green exports and green complexity for a selection of countries.

This paper is organised as follows. Section 6.2 describes the data sources and presents some key empirical patterns associated with trade in green products. Section 6.3 examines the complexity of green products and develops the GCI. Section 6.4 looks at future green diversification opportunities and discusses product proximity, proximity density, GAP, and GCP. Section 6.5 briefly discusses some tentative policy implications and Section 6.6 concludes. The Appendix contains

more information about the data (6.7.1), selected green products (6.7.2), countries (6.7.3), gives further regression robustness checks (6.7.4) and reviews how various measures used in the paper are derived (6.7.7 and 6.7.8).

6.2 Green trade data and empirical patterns

At the opening of the Doha Development Round of negotiations in 2001, the World Trade Organization received a mandate aimed at the "reduction or as appropriate, elimination of tariff and non-tariff barriers to environmental goods and services" (WTO, 2001, 33(iii)). We refer to these "environmental goods" as "green products" in this paper.

A number of international organisations have proposed lists of green products, but a universally accepted classification of green products is still not available. This is largely due to the difficulty of conceptually determining what products should be considered "green" and the practical challenges of classifying products within the existing trade classification system (Bucher et al., 2014).[4]

To construct a dataset of green products used in this paper, we draw on existing lists and classifications developed by the World Trade Organization (WTO, 2010, 2011), the Organisation for Economic Cooperation and Development (OECD, 1999, Sauvage, 2014), and Asia-Pacific Economic Cooperation (APEC, 2012) (see Table 5 in Appendix 6.7.1). We combine all available lists to construct a dataset totalling 543 products classified at the 6-digit level in HS1992. We then combine this dataset with COMTRADE data to analyse environmental trade across

[4]The latest round of talks on WTO Environmental Goods Agreement which promised to deliver a list of green products stalled in December 2016.

countries for the period 1995-2014.[5]

While our dataset of 543 products represents a useful benchmark of *potentially* green products, the environmental status of a number of products included the broad-reaching WTO Reference List may be questionable.[6] In order to arrive at a robust set of products that share wide expert endorsement—and are useful to policymakers—we develop two main product lists. The first is a list of 293 green products, which we obtain by taking the union of the WTO Core list, OECD list, and the APEC list.[7] This refined list of green products has the advantage that each product has either been endorsed by a large number of WTO or APEC member countries, or its environmental benefits have been determined by the (rather selective) OECD. This list represents a wide range of environmental categories, such as air pollution, waste water management, and recycling. We use this green product list for our empirical analysis throughout the paper.[8]

We also develop a smaller list of 57 renewable energy products (a subset of the products on the green product list). This list includes all products falling under the WTO Renewable Energy Products category, under the OECD's Renewable Energy Plant categories, as well as two additional APEC renewable energy products (solar heliostats and parts for solar heliostats) that were not included on either the WTO or OECD lists. The renewable energy product list focuses on low-carbon

[5] Huberty and Zachmann (2011) undertook a detailed study of six green products within the context of the European Union. Hamwey et al. (2013) identified 11 green products in the product space, focusing on the case of Brazil. Fraccascia et al. (2018) studied the complexity of 41 green products in 141 countries between 2005 and 2013.

[6] For example 848210 (ball bearings) submitted by Saudi Arabia with the rationale that they are used in carbon capture and storage applications.

[7] While the original set of green products included 295 goods, we had to remove Profile Projectors (903110) and Exposure meters (902740) due to data quality issues.

[8] All our data are available upon request.

technologies that are key for addressing climate change.

6.2.1 Data advantages and limitations

The green product classifications we use to construct our two product lists offer a number of advantages. First, for each proposed product in the WTO and OECD lists, it is possible to identify one (or more) environmental category that the product falls under. Although the WTO and OECD differ in the structure of their environmental categories, they are still broadly consistent and helpful for identifying a product's environmental purpose (such as renewable energy, waste water management, energy efficiency etc.) Second, the APEC and WTO lists also include specific information about each product's environmental benefits. This information was provided by member countries of the respective organisations as rationale for a proposed product's environmental endorsement. Thirdly, the APEC and WTO lists also indicate the set of member countries endorsing a given product as green. This information is useful for helping gauge the level of consensus associated with each product's environmental status.

A number of limitations are also important to keep in mind. First, the HS system (which classifies products for the purpose of trade and tariffs) was not set up to account for the environmental benefits of products. This can sometime result in poor alignment between a recognised environmental product (such as a wind turbine) and its most relevant HS code.[9] Second, many products are dual use, which means they can have both environmental and non-environmental purposes.

[9]For example, a key relevant HS code for identifying wind turbine towers is a very broad HS category - 730820 - which relates to "Towers and lattice masts, iron or steel".

Although WTO and APEC classifications provide "ex-outs" (a further description to identify relevant environmental products classified under the HS code), it can be very challenging to identify the precise environmental trade flow associated with a particular ex-out for a given HS category. As such, our analysis (which is based trade volumes for entire HS-6 commodity codes) will tend to somewhat over-estimate environmental trade volumes. Finally, our dataset does not provide information about the production process of a given product, only its use-oriented benefits. Consequently, our data do not allow us to examine the environmental impact of product production and use (e.g. lifecycle emissions of a product).

6.2.2 Empirical patterns

We now present a few empirical patterns associated with trade in green and renewable energy products. Figure 1 illustrates the change in trade volumes of green and renewable energy products between 1995 and 2014. Panel A shows that green and renewable energy products have exhibited steady growth in trade volumes–particularly over the 2000-2011 period–with a levelling off before 2014. However, when examining green trade as a proportion of total trade, we see relatively flat trajectories over the 20-year period.

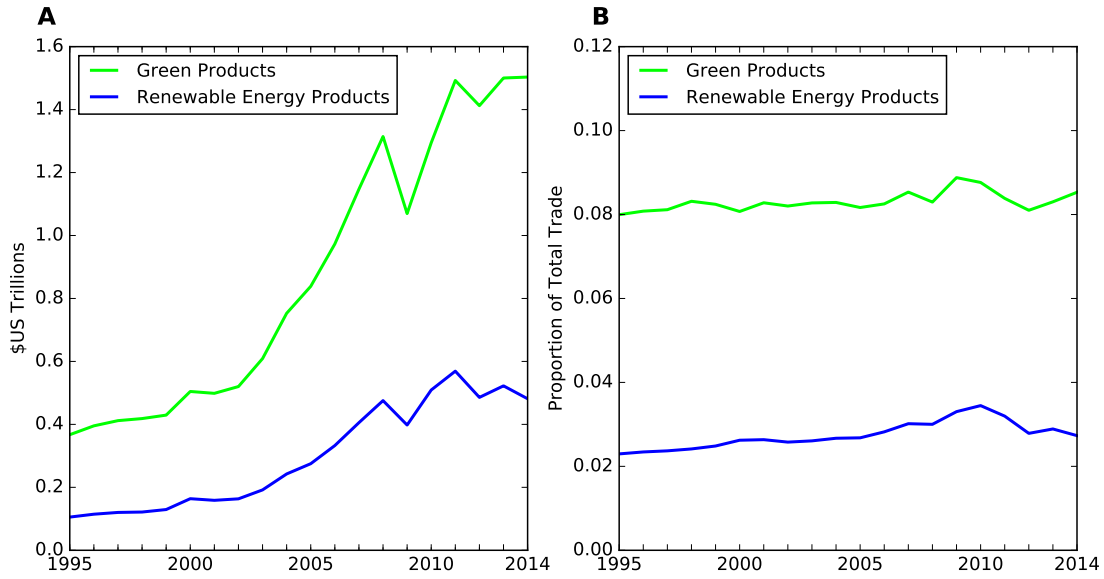


Figure 1: Growth in Green and Renewable Energy Products

Figure 2 shows the top exporters of all green products (by trade volume). Panel A presents leaders in absolute terms. While the US was the largest green exporter from 1995-2003, Germany took over in 2004 but was in turn displaced by China in 2010. Panel B shows the green exports of these same countries, but instead as a proportion of each country's total exports. Denmark has had the highest relative share of green exports – peaking over the financial crisis period at around 14 per cent. Of all these countries, South Korea has seen the largest "greening" of its export basket – its green exports increased as a percentage of total exports from around 6 percent in 2002 to around 12 percent in 2010.

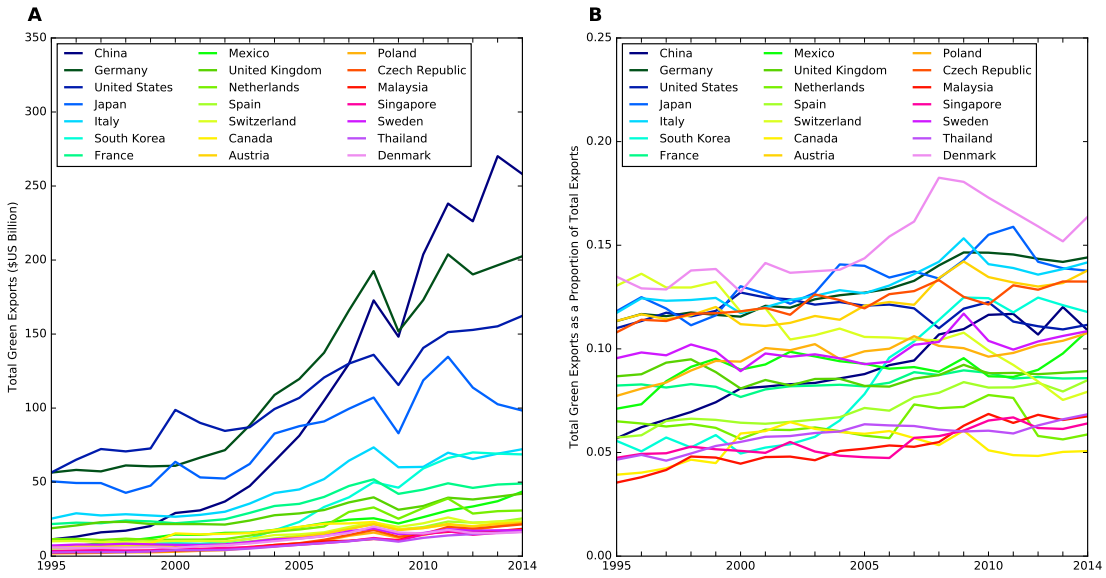


Figure 2: Top 20 Exporters of Green Products

Figure 3 shows the top exporters of renewable energy products. Again, Panel A presents the leading countries in absolute terms. As before, China has become the largest exporter of renewable energy products and its export dominance in renewable energy products (in some years exceeding $20 billion) is even greater than its dominance in all green products. In Panel B, we show the same countries' renewable energy exports relative to each nation's total exports. Here, South Korea's and Denmark's rapid patterns of green export growth become even more prominent.

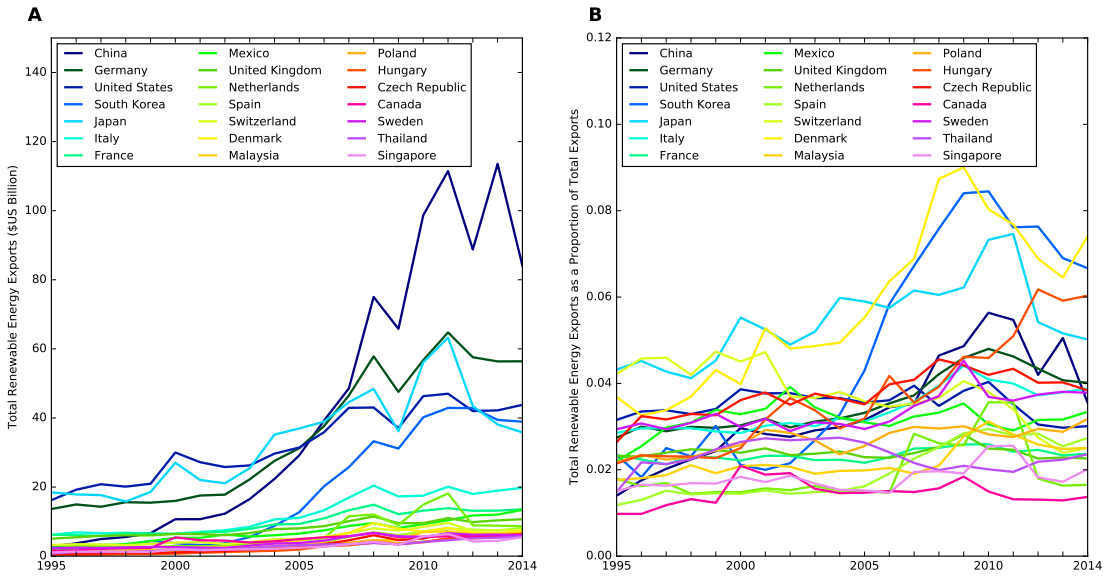


Figure 3: Top 20 Exporters of Renewable Energy Products

Next, we examine the typical income level of countries that export and import green and renewable energy products. Here we use *trade-adjusted GDP per capita* (denoted Y_X) that calculates the weighted average of GDP per capita using countries' trade shares:

$$Y_X = \frac{\sum_i \frac{X_c}{\sum_c X_c} GDP_c}{\sum_i \frac{X_c}{\sum_c X_c} POP_c} \tag{1}$$

where

$X \in$ {green exports, green imports, RE exports, RE imports, total exports, total imports},

X_c relates a particular trade measure to country c, GDP_c is country c's total GDP, and POP_c is country c's total population.[10]

[10] If $\frac{X_c}{\sum_c X_c} = \frac{1}{N}$ (where N is the number of countries), we would simply calculate the world GDP per capita.

Figure 4 shows that in 1995, green and renewable energy product exporters were much richer than an average exporter. By 2014, this was no longer the case. Figure 5 shows that the trade-adjusted GDP per capita of green and renewable energy importers has increased slightly towards the income of an average importer throughout the period.

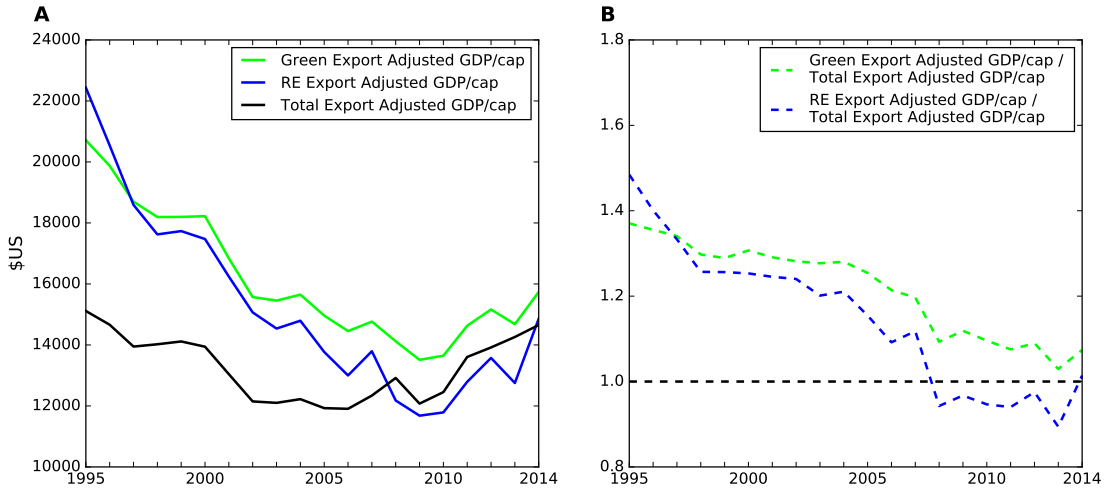


Figure 4: Green and Renewable Energy Export-Adjusted GDP/capita

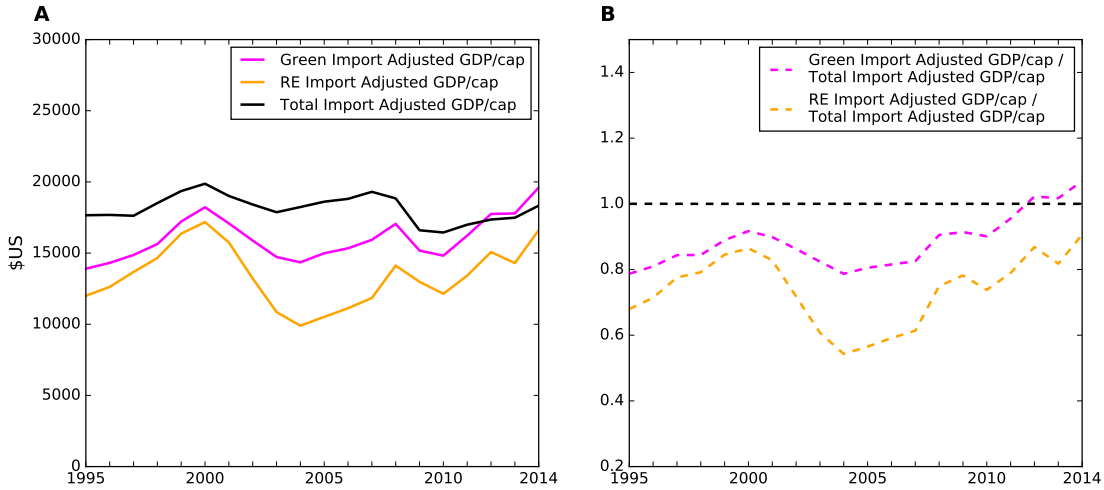


Figure 5: Green and Renewable Energy Import-Adjusted GDP/capita

An alternative way to look at the income associated with green products is to draw on a measure developed by (Hausmann et al., 2007) called *PRODY*. This measure estimates the income level associated with countries that can export a particular product *competitively*. *PRODY* for a given product p is calculated as the weighted average per capita income of countries exporting p. The weights are based on countries *revealed comparative advantage* (RCA) (Balassa, 1965). This ratio represents the size of a country c's export share in product p, relative to that product p's share of world trade, that is:

$$RCA_{cp} = \frac{x_{cp} / \sum_p x_{cp}}{\sum_c x_{cp} / \sum_c \sum_p x_{cp}} \tag{2}$$

where x_{cp} is country c's exports in product p.

Hence, the *competitiveness-adjusted GDP per capita* associated with a given product p is:

$$PRODY_p = \sum_c \frac{RCA_{cp}}{\sum_c RCA_{cp}} Y_c \tag{3}$$

where Y_c represents the per-capita GDP of country c.[11]

Figure 6 shows how the average *PRODY* of green exports (green line) and renewable energy exports (blue dotted line) has changed over time relative to the average *PRODY* of green imports (pink line) and renewable imports (dotted purple line). Competitive exporters of green and renewable energy products are much richer on average than countries that import disproportionate amounts of green products. The gap has narrowed somewhat since 1995, but remains substantial in 2014.

[11] To calculate *PRODY* based on the per capita incomes of countries *importing* a product, we use the same formula, but let x_{cp} represent country c's imports of product p.

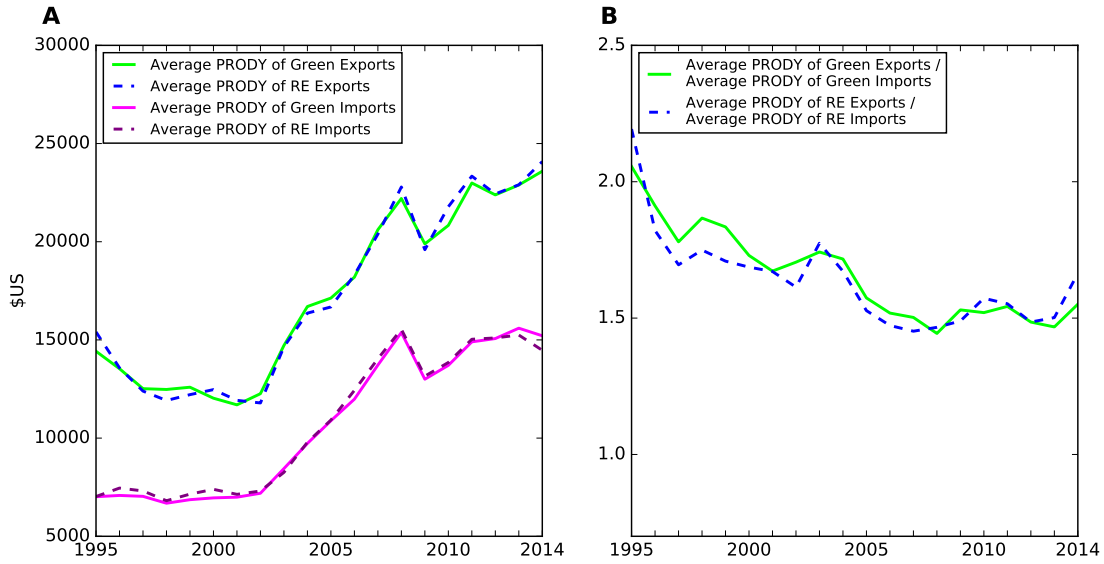


Figure 6: *PRODY* of Green and Renewable Exports and Imports

6.3 Economic complexity and the green economy

Having looked briefly at some key empirical patterns associated with green trade, we now turn to considering the nature of green production capabilities. This section first reviews the literature on capabilities and existing measurement strategies, and then applies these analytical frameworks to capture information relevant to the green economy.

6.3.1 Capabilities and complexity

The notion of capabilities has strong ties to the development and growth literature. In the development context, capabilities are often discussed with reference to the technologies, production know-how, infrastructure and institutions that enable a

country to improve its productivity and achieve higher growth rates (Lall, 1992, Bell et al., 1995, Sutton and Trefler, 2016). Here, we consider capabilities in a similar spirit – but with a distinct focus on the set of capabilities that are relevant to the green economy.

Despite its conceptual appeal, precisely defining and measuring capabilities has historically proved challenging. A number of efforts have aimed to infer information about countries' productive capabilities on the basis of their exports, with the key assumption that if a country is able to export a product, it must have the capabilities (i.e. productive know-how, infrastructure, and institutional capacity) to produce it competitively (Lall, 2000, Lall et al., 2006, Hausmann et al., 2007, Hidalgo and Hausmann, 2009, Hausmann et al., 2014). Trade data is also advantageous in offering a rich source of detailed and comparable information on what countries are able to produce competitively.

While strategies to measure capabilities relevant for growth and development have taken various forms (see Verspagen and Kaltenberg (2015) for a review), here we focus on the most recent analytical approaches, which operate on the network structure of trade data to arrive at measures of the relative complexity of capabilities associated with a given product or country (Hidalgo and Hausmann, 2009, Tacchella et al., 2012, Hausmann et al., 2014). While the term "complexity" is somewhat ambiguous, the measures are conceptually closely tied to more traditional notions of technological sophistication (Lall et al., 2006) and product quality (Sutton and Trefler, 2016).

The Economic Complexity Index (ECI) and the Product Complexity Index (ECI), which were proposed by Hidalgo and Hausmann (2009) and refined in Hausman

et al. (2011) have attracted significant interest from researchers and policy makers as they are able to explain more variation in country income per capita and economic growth than other variables commonly employed in growth equations (such as governance, institutional quality, education, and competitiveness; see Sala-i-Martin, 1997). Although there has been some confusion as to what the measures capture and how they work, Mealy et al. (2017) recently showed that they, in fact, correspond to a standard clustering algorithm that partitions a similarity graph into two groups, such that entities in one group are similar to each other and dissimilar to the other group.

Viewed from this perspective, the ECI indirectly captures the "complexity" of countries' capabilities by exploiting the pattern of similarity in country export data. Countries receiving a high ECI have more similar export baskets to other countries with a high ECI, and these tend to be advanced economies that are able to export technologically sophisticated products. In contrast, countries assigned a low ECI have greater export similarity to other countries with low ECI, and these countries' export baskets tend to be characterised by less technologically sophisticated products. The PCI operates in exactly the same way, but instead provides an ordering over products in terms of the similarity in the countries that export them. That is, products receiving a higher PCI tend to be exported by a similar group of countries to other products receiving a high PCI, and tend to be more technologically sophisticated – and vice versa. More detail about the ECI and PCI can be found in Appendix 6.7.7.

While this paper primarily focuses on the measures proposed by Hausmann et al. (2014), we note that alternative complexity measures have also been proposed by

Tacchella et al. (2012). These measures are not related to a clustering algorithm, but calculated as the fixed-point solution of a non-linear mapping function, which instead exploits the pattern of export diversity (the number of products a country is able to export competitively). Tacchella et al. (2012)'s Country Fitness measure (an alternative to the ECI) can be thought of as a weighted-diversity measure, where each product that a country exports competitively is weighted by its complexity. Tacchella et al. (2012)'s corresponding Product Complexity measure is a non-linear function that is inversely related to the number of countries that can export the given product competitively. We have also performed the analysis in this paper using this alternative approach and find fairly consistent results (see Appendix 6.7.8), suggesting that our findings are robust to the chosen complexity metric.

6.3.2 The complexity of green products

Drawing on Hausmann et al. (2014)'s methodology, we first calculate the PCI for the entire set of HS1992 6-digit products. From the universal set of almost 5,000 products, we then identify products falling into our green and renewable energy product lists.

We present the key descriptive statistics for PCI of all products, green products, and renewable energy products in Table 1. The PCI values are normalised such that the set of all HS1992 6 digit products have a mean of 0 and standard deviation of 1. We find that green and renewable energy products have mean PCI value of 0.48 and 0.49 respectively. This suggests that, on average, green and renewable energy products require more complex capabilities for production than typical

products.[12] In Figure 7, we present the distribution of the PCI values of all green and renewable energy products.

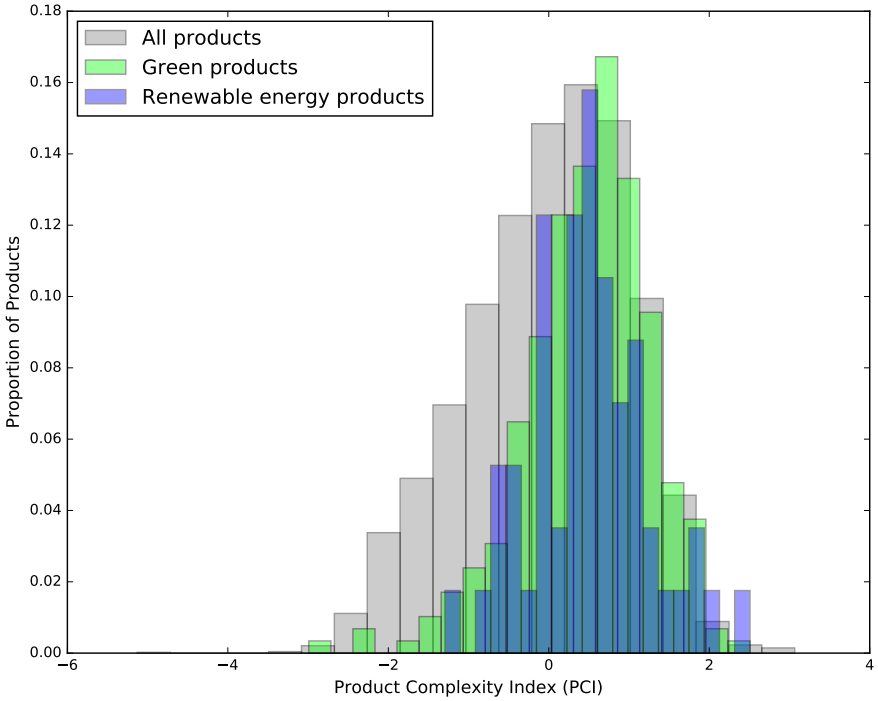


Figure 7: PCI distribution for all HS6 products, green products and renewable energy products

Table 1: Product PCI distribution descriptive statistics

Product Set	Number of Products	Mean PCI	Std PCI
All HS6 Products	4857	0	1
Green Products	293	0.48	0.79
Renewable Energy Products	57	0.49	0.72

In Appendix 6.7.2, we present the top 10 and bottom 10 green products ranked

[12]The Kolmogorov-Smirnov 2-sample tests reject the null hypothesis that green product PCI distributions are different from all product PCI distributions (KS-Statistic for green products vs all products = 0.242, p-value = 1.11×10^{-14}). The Kolmogorov-Smirnov 2-sample test fails to reject the null hypothesis that the green and renewable energy products are drawn from the same distribution (KS-Statistic = 0.096, p-value = 0.747)

according to their PCI values for the year 2014. We also present analogous PCI rankings for renewable energy products. We include the environmental benefit or category associated with each product as well as the environmental list each product is included in. The most complex green products relate to devices used for environmental monitoring and analysis, and concentrated solar technologies (Table 6), while the most complex renewable products represent a range of different mechanical devices and associated parts used in concentrated solar, wind and gas turbines (Table 8). Less complex green goods tend to relate to environmentally-friendly products – many of which are made from vegetable material (Table 7). Less complex renewable products represent a mixed bag of electrical components, biofuels, mirrors for solar PV and masts for wind turbine towers (Table 9).

6.3.3 The green complexity index

Which countries have the most complex green production capabilities? To answer this question, we introduce the *Green Complexity Index (GCI)*, which is an increasing function of both the number and complexity of green products that a country is able to export competitively. Specifically, the GCI of country c is given by

$$GCI_c = \sum_g \rho_g \widetilde{PCI}_g, \tag{4}$$

where ρ_g is a binary vector in which a 1 corresponds to a country having $RCA > 1$ in green product g and 0 otherwise, and $\widetilde{PCI}_g$ is the Product Complexity Index of g normalised to take a value between 0 and 1.

While existing measures such as Hausmann et al. (2014)'s ECI and Tacchella et al. (2012)'s Country Fitness measure operate on the entire set of traded products, the GCI is a function of the *subset* of green products. This definition is of course completely general, and could be applied to any product subset of interest.[13]

In Figure 8, we show how the GCI ranks of countries have changed between 1995 (left axis) and 2014 (right axis). Interestingly, there has been relatively little variance in the rankings of the top 10 countries. Impressively, Germany has managed to keep its top spot for the entire 20 year period. Some countries, such as China, Vietnam and Uganda have made significant gains in their green production capabilities, while other countries, such as Australia, have seen a substantial decline in their GCI rankings.

[13]We compare the GCI calculated using the Hausmann et al. (2014) PCI and Tacchella et al. (2012)'s Product Complexity measure. As shown in Appendix 6.7.5, both formulations give very similar results, suggesting that the GCI is robust to the choice of product complexity measure. It is also important to note that for this particular set of traded products, the complexity scores are fairly homogenous (see Figure 7), particularly when normalised to take a value between 0 and 1. As such, the GCI score is very strongly correlated to a country's green diversity (the number of green products it is competitive in). However, this will not necessarily be the case for different product subsets, where there is greater variation in product complexity.

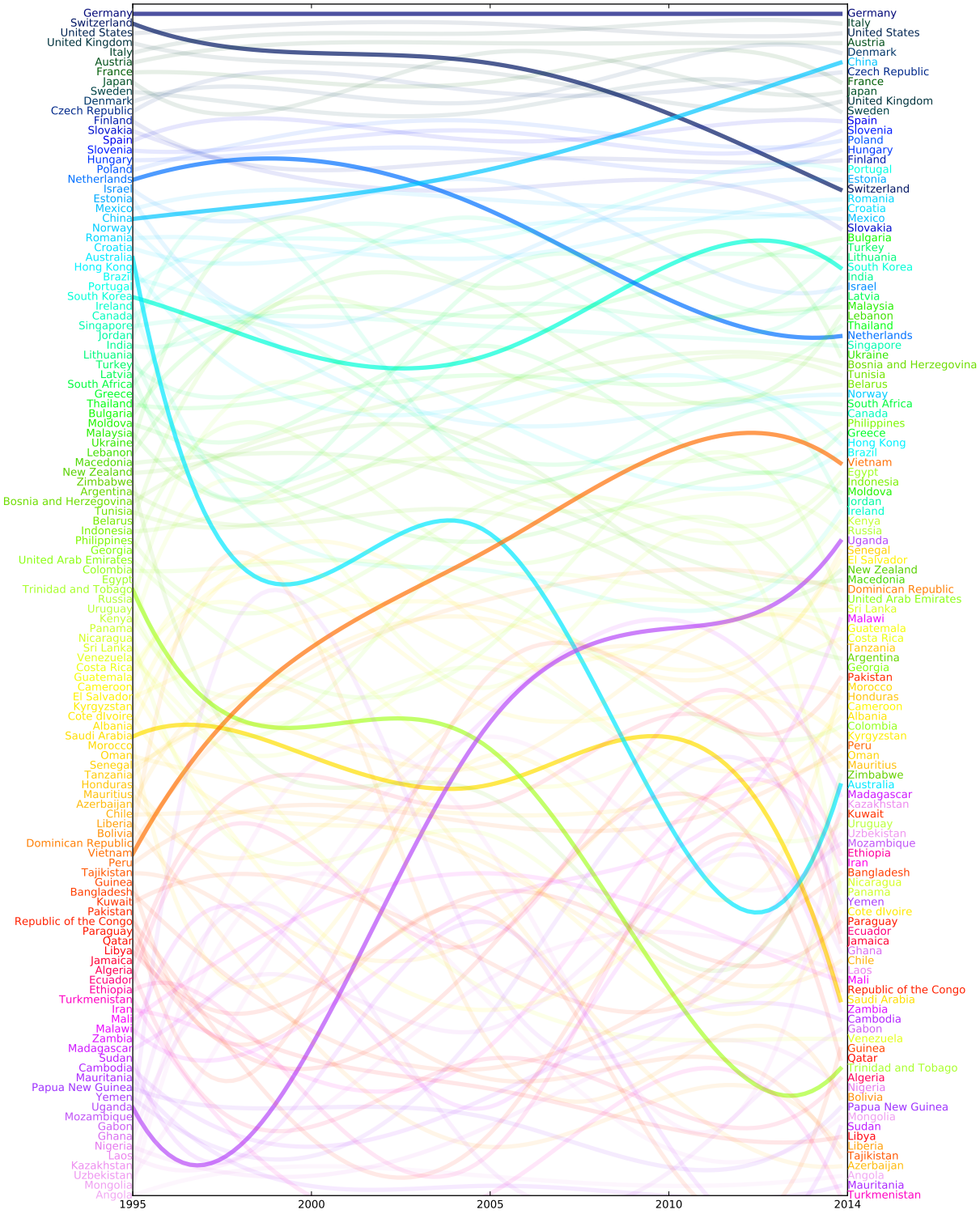


Figure 8: Green Complexity Index rankings over time

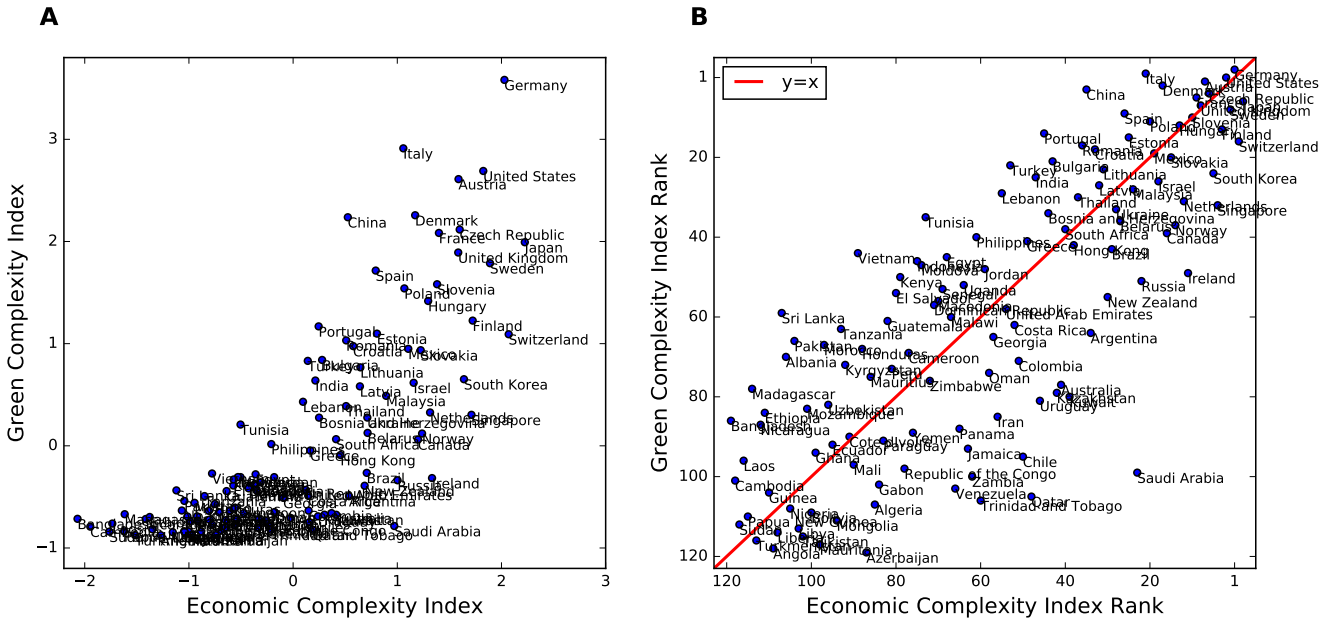


Figure 9: GCI and ECI comparisons for 2014

GCI is positively correlated with ECI, as shown in Figure 9. Panel A shows the correlation between GCI and ECI scores[14], while Panel B shows the relationship between country rankings.[15] The positive correlations are unsurprising given that richer countries have higher economic complexity and green products tend to be more complex on average.[16] However, the deviations in the ranks and values of ECI and GCI are informative about the differences in the orientation of countries' export baskets. For example, countries that are heavily focused on exporting oil and petroleum products, such as Saudi Arabia, Trinidad and Tobago, and Qatar, have lower GCI compared to ECI. In contrast, Tunisia, China, and Italy have much higher GCI scores relative to their ECI, suggesting that their production

[14]Pearson correlation coefficient $= 0.766$, p-value$= 0.5 \times 10^{-25}$)

[15]Spearman's rank correlation coefficient $= 0.79$, p-value$= 5.2 \times 10^{-27}$.

[16]The correlation between ECI and GCI has remained relatively stable over the 1995-2014 period. These results are available upon request.

capabilities may be more aligned to green products. It may also suggest that if a green transition substantially increased demand for green goods, these countries could stand to benefit, relative to other countries having similar levels of economic complexity.

In Figure 10, we show the relationship between GCI and log GDP/capita for 2014. Again, the positive relationship is not surprising,[17] but the variance in the relationship provides additional insights into the current orientation of countries' economies. Consistent with Figure 9, a number of resource-rich countries have low GCI scores given their income. Germany, Italy, China, and India stand out as having much higher GCI scores given their income per capita, suggesting that their production capabilities are more oriented to the green economy than other countries with a similar standard of living.

[17] Pearson's correlation coefficient $= 0.55$, p-value $= 7.52 \times 10^{-11}$.

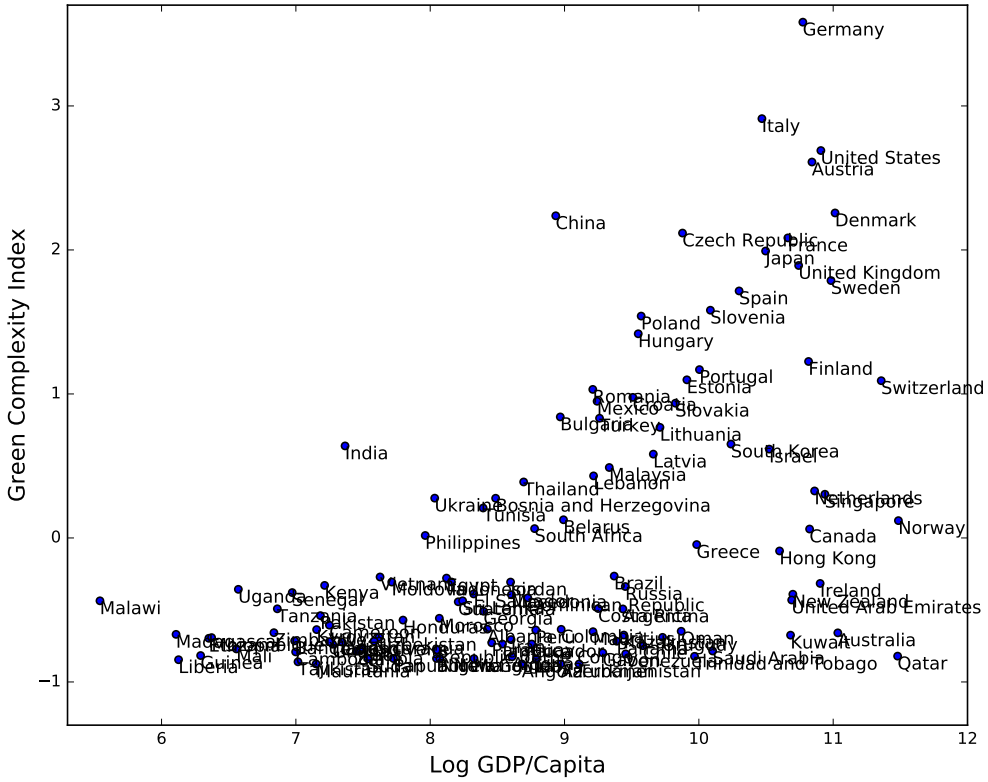


Figure 10: GCI vs log GDP per capita for 2014

Finally, if GCI really captures the green economy capabilities of countries, we should expect it to explain variation in environmentally-relevant country characteristics. We examine GCI's ability to explain cross-country variation in environmental patents, CO_2/capita emissions, and the OECD measure of stringency of environmental policy (EPS), after controlling for each country's ECI and per capita income.

Since GCI and ECI can fluctuate year on year due to variability in trade data, we use simple regressions on time-averaged explanatory variables as follows:

$$\bar{y}_i = \bar{x}_i\beta + \epsilon_i$$

where $y_i \in \{\text{Log Env. Patents, } CO_2\text{/capita, Log EPS}\}$, $\bar{y}_i = \frac{1}{N}\sum_{t=t_0}^{t=t_N} y_{it}$, $\bar{x}_i = \frac{1}{N}\sum_{t=t_0}^{t=T_N} x_{it}$ are time-averaged explanatory variables for N available periods, and ϵ_i is the error term.

Table 2 shows GCI's ability to explain variation in environmentally-relevant variables over the twenty year period covered by our data. We find that GCI is strongly positively correlated with the number of environmental patents across countries, even after controlling for GDP/capita and ECI. This provides further validation that the GCI captures green-economy relevant know-how across countries. We also find that countries with higher GCI tend to have lower CO_2 emissions. This relationship is particularly interesting, given our dataset does not account for the emissions intensity of each product's production process. Additionally, we find a positive relationship between GCI and the OECD's Environmental Policy Stringency Index, suggesting there is some association between the environmental policies in place in a country and its green production capabilities. While the results in Table 2 reflect the GCI's explanatory power over the long run, we also run regressions for different years in Appendix 6.7.4 and find consistent results.

Table 2: Green Complexity Index

	Log Env. Patents	Log CO_2/cap	Log EPS
GCI	1.009***	−0.307***	0.100***
	(0.215)	(0.102)	(0.029)
ECI	1.158***	0.290*	−0.115**
	(0.286)	(0.157)	(0.053)
Log GDP/Cap	0.116	0.850***	0.213***
	(0.128)	(0.086)	(0.034)
Intercept	1.593	−6.196***	−1.093***
	(1.125)	(0.734)	(0.315)
Observations	1220	2318	558
Adjusted R^2	0.766	0.765	0.7532

Robust standard errors in parenthesis.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Environmental patents data covers 2000 and 2005-2013, available from http://stats.oecd.org/. CO_2 (metric tons per capita) data covers 1995-2013, available from https://data.worldbank.org/indicator/EN.ATM.CO2E.PC. Environmental Policy Stringency (EPS) data covers 1995-2012, available from http://stats.oecd.org/. For all regressions, we take country averages over all available time periods.

6.4 The green product space and future green industrial transitions

In providing an estimate of countries' capabilities to export green, complex products, the GCI gives us an idea of which countries are *currently* likely to be best placed to be leaders in the green economy. However, successfully transitioning to a greener economic growth model will no doubt require many countries to re-orientate their existing industrial structure and cultivate new green industries. As such, we now turn to investigating the question: How might countries re-orient

their industrial structure to become more competitive in green products in the future?

To answer this question, we draw on a theory, advanced in both the economic complexity and economic geography literatures, which proposes that economic development is path-dependent due to the underpinning knowledge accumulation process. The underlying intuition is that if a country has the capabilities to produce T-shirts, it is relatively easy for the country to diversify into trousers because much of the requisite production capabilities (e.g. in sewing techniques, factory layout, textile supply chains, design of clothes) and factors of production are similar. However, it is much more difficult to diversify *directly* from T-shirts into automobiles because the country would have to acquire a large amount of new production know-how and invest into completely new factors of production. By estimating the similarity of production know-how underpinning different products, and combining this with information on the country's *current* production capabilities (inferred from the export data), we identify the industries and products that countries would be more likely to be able to easily diversify into in the future.

We focus on two measures developed by Hidalgo et al. (2007), which were applied specifically to export data. They argued that if products share similar capabilities, they are more likely to be exported competitively together. On this basis, they suggested that the *proximity* (similarity in capabilities) between two exported products could be calculated by their pairwise conditional probability of being co-exported competitively. Further, to estimate how proximate a product is to a country (or, how easy a product might be to export competitively given what the country already exports competitively), Hidalgo et al. (2007) proposed a *proximity*

density metric, which calculates the average proximity of a country's exports to the given product. Importantly, Hidalgo et al. (2007) were able to provide empirical support for the path-dependence of knowledge accumulation as measured by proximity density. They showed that over time, countries were much more likely to become competitive in products that were proximate to their existing capabilities.

In this section, we apply proximity and proximity density measures to the green product dataset. First, we explore green products in terms of the similarity of their requisite capabilities and visualise this by constructing a *green product space*. Second, we estimate how proximate countries are to green goods on the basis of their current export basket. Third, we develop a new metric to estimate countries' *potential* to become competitive in green products in the future.

6.4.1 The green product space

Understanding what capabilities green products share helps us estimate how easy it is to export one green product (e.g. batteries) competitively if a country already has the capabilities to export another green product (e.g. solar panels) competitively.

Using Hidalgo et al.'s 2007 proximity measure given below we calculate the proximity between each green product i and green product j (denoted ϕ_{ij}) on the basis of their conditional probability of being co-exported competitively:

$$\phi_{ij} = \min(P(RCA_i > 1|RCA_j > 1), P(RCA_j > 1|RCA_i > 1)) \tag{5}$$

where $P(RCA_i > 1|RCA_j > 1)$ is the conditional probability of being competitive

in green product i given you are competitive in green product j. Taking the minimum ensures that the proximity matrix is symmetrical ($\phi_{ij} = \phi_{ji}$)

To visualise the similarity in capabilities underpinning green products, we also follow Hidalgo et al. (2007) and construct a hierarchically clustered network where green products are linked to other green products if they have a high probability of being co-exported. To create this network, we construct a maximum spanning tree[18] from the weighted matrix ϕ and add additional edges with proximity greater than a given threshold (here we use a proximity threshold = 0.37).[19] This ensures we only connect green products that have a high probability of being co-exported. We show the resulting network – *the green product space* – in Figure 11.

[18] A spanning tree of a given graph is a tree (contains no cycles) that connects all vertices with the minimum possible number of edges. A *maximum* spanning tree is a spanning tree of a weighted graph that has the maximum weight. That is, it connects nodes by adding edges with the largest weight until the graph is fully connected.

[19] Alternative thresholds give similar results.

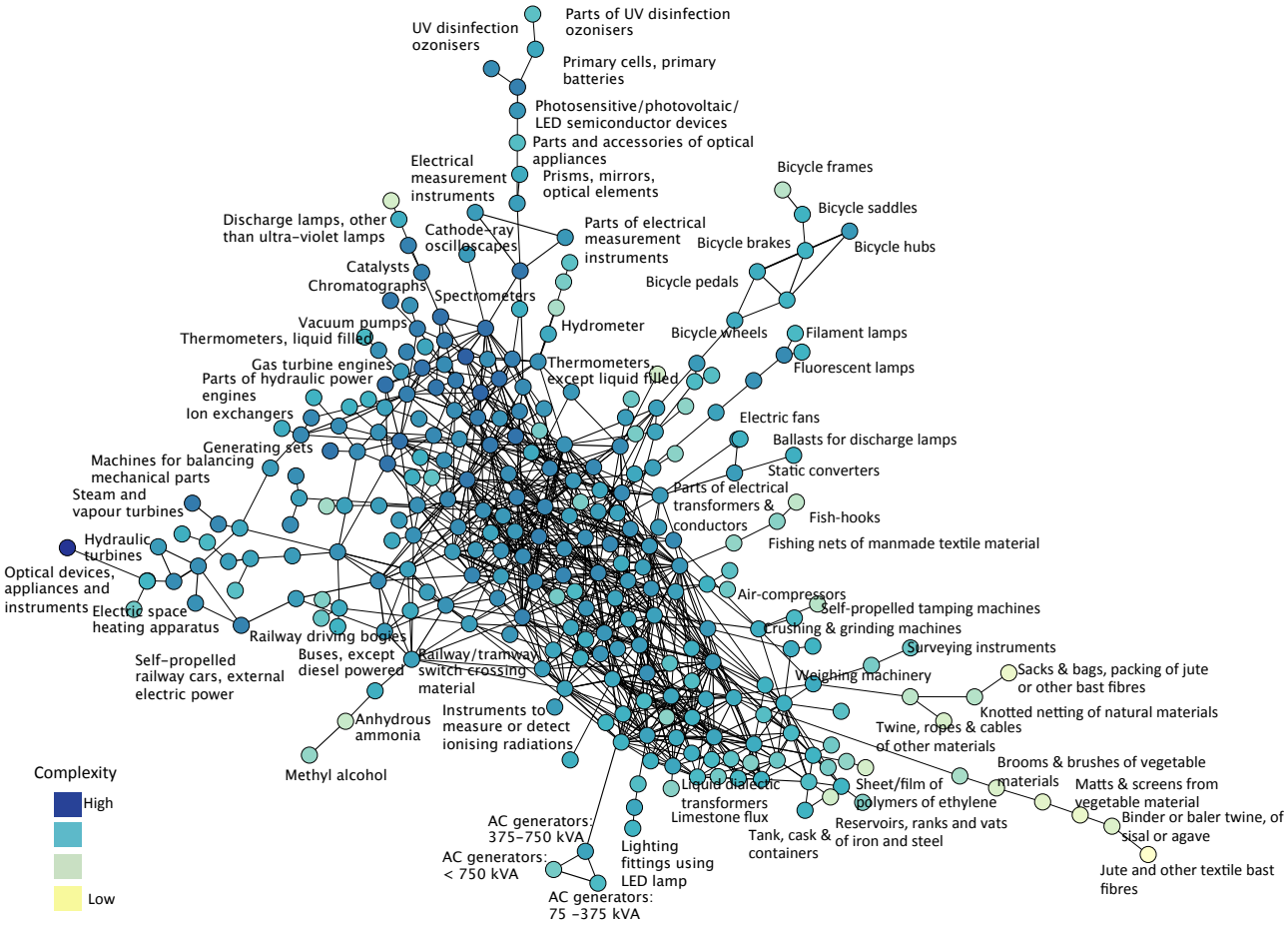


Figure 11: The Green Product Space. Nodes represent green products linked to other green products they are the most proximate to. Product nodes are coloured in accordance with their PCI value.

Similar to Hidalgo et al.'s (2007) product space for the entire set of traded products, we find that less complex green products tend to be located in the periphery of the green product space, while more complex products are located in the core. This is interesting from a green diversification-oriented development perspective: while it may be relatively easy to export green products with lower complexity competitively, the accumulated capabilities may have limited spillover opportuni-

ties into other green products. However, as more complex products tend to be technologically linked to many other green products, gaining capabilities to export more complex green products could provide greater future green industrial development possibilities.

The green product space also provides a new way to visualise each country's competitive green exports. We show a selection of different countries in Figure 12. Holding the underlying network fixed, we colour (in green) products that a given country exports competitively. While the most striking aspect of Germany's export basket is the sheer abundance of competitive green products, it is interesting to note that the majority of these are located in the core of the green product space. South Korea also competitively exports a number of complex green products located in the green product space core, but specialises in a distinct branch of green products relating to solar photovoltaics and batteries. As a developing country, Uganda currently exports fewer green products - many of which are less complex and tending relate to vegetable materials. Finally and unsurprisingly, Saudi Arabia currently exports very few green products - all located around the periphery of the green product space. In line with the results above, this signals that countries with production capabilities too narrowly focused on resource extraction activities may find that their green production capabilities are underdeveloped and their competitive advantage is less aligned with the direction of the future green economy.

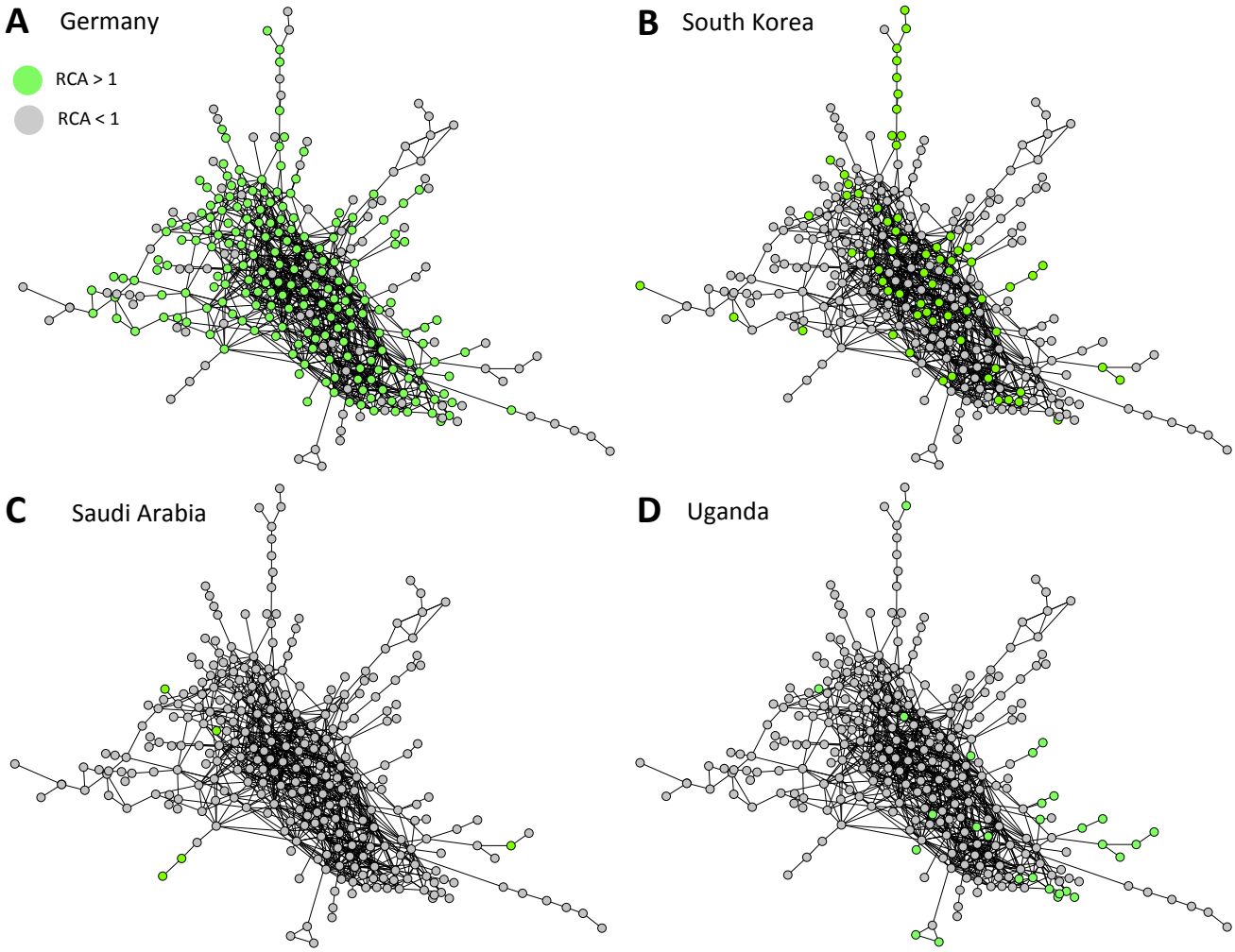


Figure 12: Comparison of a selection of country positions in the green product space. Green coloured nodes represent green products that countries are competitive in.

6.4.2 Proximity density and the green adjacent possible

Visualising countries' positions in the green product space gives us an indication of countries' current and future green diversification potential. An alternative quantitative way to explore green transition possibilities is to draw on Hidalgo et al.'s

second proximity measure: *proximity density*. While *proximity* (described in the previous section) estimates the similarity in green productive capabilities underpinning two products i and j, *proximity density* estimates the similarity between a country c's current set of green productive capabilities and the capabilities required to produce a given green product j. Put another way, *proximity density* estimates how easy it is to diversify into a new product j, given all the products country c can export already competitively.

Proximity density (given below) is a function of the *proximity* metric ϕ_{ij} defined above and is calculated as the average *proximity* between a given green product j and *all* the other products country c exports competitively:

$$\omega_j^c = \frac{\sum_i \rho_i \phi_{ij}}{\sum_i \phi_{ij}} \tag{6}$$

where ρ_i is a vector of i products for which country c has $RCA > 1$.

In Figure 13, we show the same selected countries depicted in Figure 12. These plots characterise each country's *Green Adjacent Possible* (GAP) – the set of green industrial opportunities that are proximate to a country's current production capabilities. In each panel, dots represent green products that countries do not currently export competitively. The x-axis shows the proximity density value for each green product for a given country and the y-axis measures each product's PCI.

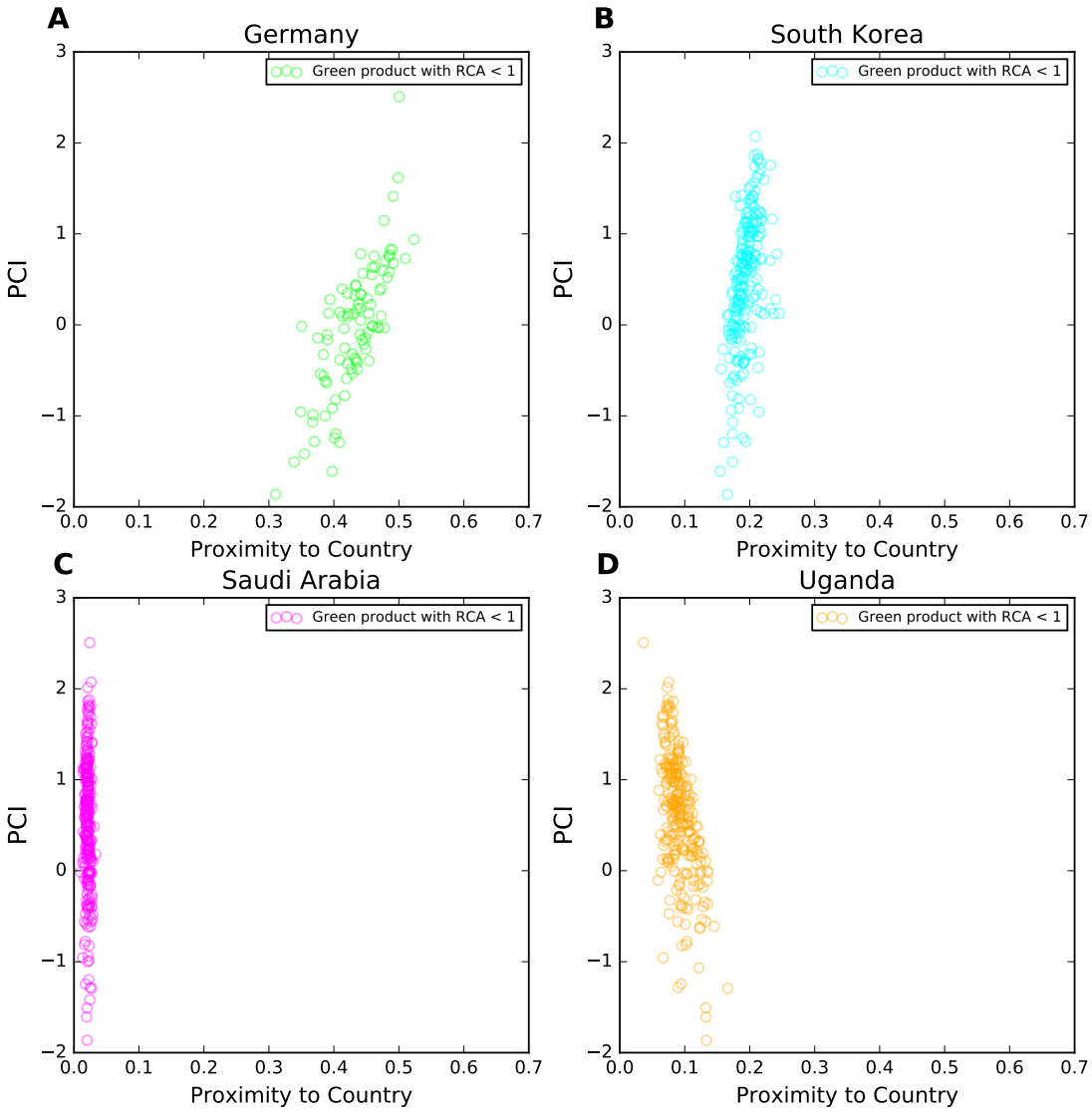


Figure 13: Characterising the Green Adjacent Possible for different countries. In each plot, the circles represent green products that the denoted country is not yet competitive in. The y-axis represents the complexity (as given by the PCI) of each product, and the x-axis represents that product's *proximity density* to a country (an estimate of how close it is to the country's current capabilities).

A number of things are interesting to note. As we would expect, Saudi Arabia is much less proximate to the set of green products because its productive know-how

is more closely focused on extracting fossil fuel resources. Uganda, a developing country, is also less proximate to more complex green products because it currently has fewer production capabilities overall. In contrast, being more economically advanced and specialising in a number of complex green products, South Korea's industrial structure is proximate to a number of green products. A slight positive slope is also evident, suggesting that South Korea's productive capabilities are more oriented towards complex green products than less complex green products. Finally, Germany's advanced manufacturing base and significant expertise in green products is reflected in its greater positive slope and high proximity to very complex green products.

6.4.3 Green complexity potential

Thus far our the GAP analysis has only been illustrative and focused on four countries. We now summarise the GAP of each country into a single aggregate number that can be compared across countries and over time. To this end, we develop the *Green Complexity Potential* metric, which which measures the *average* proximity to green complex products a country currently is not competitive in:

$$GCP_c = \frac{1}{|1-\rho_g|}\sum(1-\rho_g)\omega_g^c\widetilde{PCI}_g \tag{7}$$

where $1-\rho_g$ is the vector of green products a country currently does not have RCA >1 in, ω_g^c is the proximity of product g to country c, and $\widetilde{PCI}_g$ is the PCI of product g, normalised to take a value between 0 and 1. GCP is similar to *Complexity Outlook Index* (Hausmann et al., 2014) and the *Complexity Potential*

measure (O'Clery et al., 2016). However, while these measures are applied to the entire set of traded products (Hausmann et al., 2014) or industries (O'Clery et al., 2016), the GCP is specific to the subset of green products.

Figure 14 shows the relationship between GCP and GCI for 2014. Panel A shows the relationship between GCP and GCI scores, while Panel B shows the relationship between GCP and GCI ranks. In both cases, we find a strong correlation which indicates that the more green production capabilities a country has, the easier it is to diversify into additional new green products.[20]

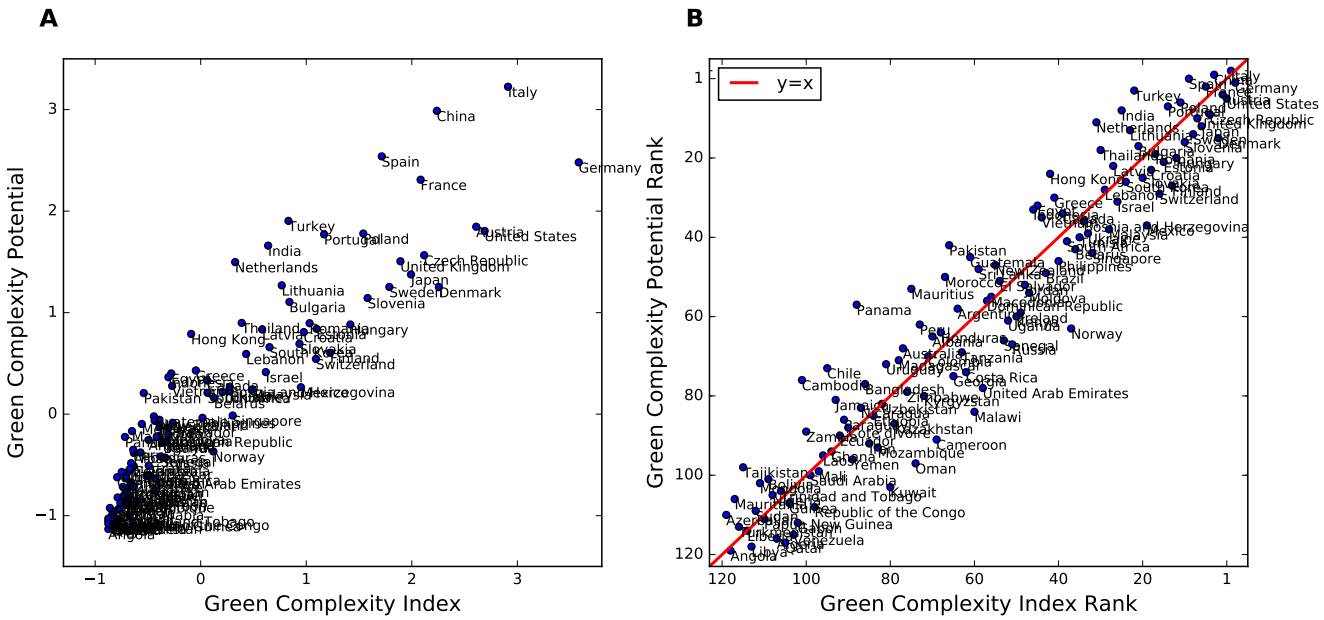


Figure 14: GCP and GCI comparison for 2014

However, the differences between GCI and GCP provide additional information about future growth: countries including China, Spain, Turkey, India and the Netherlands have higher significantly higher GCP than GCI, suggesting that these

[20]Panel A: Pearson correlation coefficient $= 0.921$, p-value$= 3.49 \times 10^{-51}$, Panel B: Spearman correlation coefficient $= 0.951$, p-value$= 2.11 \times 10^{-63}$.

countries may be particularly well-positioned for fast development of future green capabilities. In contrast, while countries like the US, Japan and Denmark currently have very strong green production capabilities; their lower GCP score indicate that future expansion into new green product markets could be relatively slower.

In Table 3, we explore how predictive a country's GCP is for future increases in its green capabilities (as measured by GCI), the number of green products it is able to export competitively, and the share of green exports in its total export basket. Specifically, we regress the countries' GCP at the beginning of the period (averaged over 1995-2000) on the change in countries' GCI, number of competitively exported green products and green export trade ratio at the end of the period (averaged over 2009-2014) i.e.

$$\Delta \bar{y}_i = \bar{x}_i \beta + \epsilon_i$$

where $y_i \in \{\text{GCI, \#Green exported products, Green exports}\}$, $\Delta \bar{y}_i = \frac{1}{5} \sum_{t=2009}^{t=2014} y_{it} - \frac{1}{5} \sum_{t=1995}^{t=2000} y_{it}$, $\bar{x}_i = \frac{1}{5} \sum_{t=1995}^{t=2000} x_{it}$ are explanatory variables averaged at the beginning of the sample, and ϵ_i is the error term. This specification is similar to the approach taken by O'Clery et al. (2016). However, to ensure our results are robust to year on year trade data fluctuations, we take 5-year averages.

Controlling for countries' current incomes and ECI, we find that countries with higher GCP scores are more likely to have greater future increases in their GCI, green export trade ratio and number of green products they are able to export competitively.

Table 3: Green Complexity Potential

	Δ GCI $(t+\delta)$	Δ #Green exported products $(t+\delta)$	Δ Green export trade ratio $(t+\delta)$
Log GCP(t)	0.172***	7.118***	0.012***
	(0.038)	(1.678)	(0.003)
Log GDP/Cap(t)	−0.005	−0.448	0.001
	(0.024)	(1.135)	(0.002)
ECI(t)	−0.143**	−7.450***	−0.006*
	(0.043)	(2.112)	(0.004)
GCI(t)	−0.060		
	(0.051)		
Green exported products(t)		0.084	
		(0.057)	
Green export trade ratio(t)			−0.075
			(0.158)
Intercept	0.577**	29.715***	0.039**
	(0.245)	(11.110)	(0.016)
Observations	1220	1220	1220
Adjusted R^2	0.203	0.212	0.169

Robust standard errors in parenthesis.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
t relates to country averaged values over years 1995-2000 and $t + \delta$ relates to country averaged values over years 2009-2014. #Green exported products refers to the number of green exports in which the country has $RCA > 1$ (i.e. diversity).

In Appendix 6.7.4, we show that the predictive power of GCP is robust to different time-averaging specifications.

6.5 Policy implications

Our results have a number of policy implications.

First, we are able to identify and measure trade in a much more extensive set of

green products by drawing on a number of international agreements and independent policy sources. Our dataset is therefore a robust, consensus-driven definition of green products that can be used in research and policy. Our results in Section 6.3 also indicate that green and renewable energy products tend to be complex and are consequently likely to require substantial investment and expertise to scale for production.

Second, we estimate which countries are currently best positioned to thrive in the green economy - shedding light on increasingly important question for policy makers. Our work complements that of Fankhauser et al. (2013), who investigated a similar question by focusing on the trade competitiveness and patenting rates in 110 manufacturing sectors across 8 countries. We provide a more extensive coverage of countries and green products, and also advance an alternative analytical framework based on the economic complexity methodology.

Third, we show how a country's capabilities relevant to the green economy can evolve over time. A country that finds itself sliding down the GCI or GCP ranking may want to strengthen policies aimed at increasing its green capabilities. In fact, we can present some preliminary evidence that direct government intervention can improve green production capabilities. We use data on green stimulus packages in 19 countries over the early years of the global financial crisis (Barbier et al., 2010). Many countries embarked on stimulus programmes to boost their weak economies and green spending formed a significant part of the stimulus. Table 4 shows that even after controlling for GDP per capita, the size of the stimulus packages is positively associated with increases in GCI, in the number of green exports that the country is competitive in, and in the ratio of green exports to total exports

between 2008 and 2011 (this holds both for stimulus and stimulus per capita, see Appendix 6.7.6).

Finally, the path dependence of green capabilities tentatively indicates a role for industrial policy (see, for example, Aghion et al., 2011, Hallegatte et al., 2013, Rodrik, 2014). However, we cannot recommend a specific green industrial policy based on our results. By identifying the GAP, we can precisely indicate where the next competitive green opportunities for each country are likely to be. Stimulating competitiveness in some green industries – particularly in more complex green products – is likely to make it easier for countries to transition to other green industries in the future. But whether a transition to these technologies requires interventionist industrial policy or regulatory reform needs to be decided on a case by case basis. Our green stimulus results only provide some indication that government policy can have an effect on green capabilities. It is also important to stress that GCI, GCP, and GAP only reflect the particular orientation of countries' current export baskets and does not account for domestic or service-based green production capabilities that may exist within these countries. A proper green industrial policy should take all green capabilities and all relevant domestic policy objectives and constraints into account.

Table 4: Green Stimulus Analysis

	Δ GCI $(t+\delta)$	Δ #Green exported products $(t+\delta)$	Δ Green export trade ratio $(t+\delta)$
Green Stimulus	0.970***	41.565***	0.027*
	(0.205)	(9.699)	(0.015)
Log GDP/Cap(t)	0.000**	0.000**	0.0000
	(0.000)	(0.000)	(0.0000)
GCI(t)	0.054*		
	(0.027)		
#Green exported products(t)		0.076**	
		(0.029)	
Green export trade ratio(t)			0.082*
			(0.043)
Intercept	0.056	−0.337	−0.007*
	(0.047)	(2.204)	(0.003)
Observations	19	19	19
Adjusted R^2	0.495	0.495	0.265

Robust standard errors in parenthesis.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
t relates to the year 2008 and $(t+\delta)$ relates to the year 2011.
Green Stimulus units are US '000 per capita. Green Stimulus data are from Table 1 of Barbier et al. (2010), and relates to low carbon support for renewable energy, carbon capture and sequestration, energy efficiency, public transport and rail, and improving electrical grid transmission. #Green exported products refers to the number of green exports in which the country has $RCA > 1$ (i.e. diversity).

6.6 Conclusion

This paper has advanced an approach to systematically analyse countries' current capabilities and future potential to succeed in the green economy.

Our comprehensive new dataset revealed a number of important patterns. First, green and renewable products have not grown as a fraction of total trade in the last

twenty years. Given the urgency of the transition towards a green economy, agreements to advantage trade in these products might play an important role. Second, using the economic complexity methodology, we showed that how countries' capabilities inferred from export data can explain a number of environmentally-relevant variables. Third, we identified particular green products that countries could gain a competitive edge in as they transition towards green-oriented growth models.

There are plenty of fruitful areas for further work. First, we have only considered capabilities based on export data. While the GCI and GCP explain variation in environmentally-relevant measures across countries, we cannot account for capabilities embodied only in services (Stojkoski et al., 2016, OECD, 2017) or in goods sold only domestically. Second, we do not account for occupation-specific skills relevant for new green economy products (Neffke and Henning, 2013, Mealy et al., 2018). Third, we do not look at regional or city-level variation (Boschma et al., 2013, O'Clery et al., 2016). Fourth, we have not considered channels for green technology diffusion across neighbouring countries (Bahar et al., 2014). Fifth, we have not explored regional green industrial policy or regional specialisation (e.g. Pearl River Delta, Silicon Valley): government policy and research might want to focus directly on the competitiveness of regional production clusters (Delgado et al., 2014). Finally, it would be worth understanding what roles services, skills, and regional specialisation play in a more expanded definition of the green economy by looking closely at green patents, green research and development, carbon emissions and environmental protection.

6.7 Appendix

6.7.1 Description of data sources

Table 5: Green Product Data Sources

List	Description	Source
WTO Reference Universe	408 products that represent a universe of potentially green products proposed by different WTO Member States	• WTO Report by the Chairman to the Trade Negotiations Committee on the Committee and Trade and Environment in Special Session TN/TE/19 (22 March 2010) • WTO Report by the Chairman to the Trade Negotiations Committee on the Committee and Trade and Environment in Special Session TN/TE/20 (21 April 2011)
WTO Sample Core List	26 products with wide endorsement from WTO Member States	• WTO Report by the Chairman to the Trade Negotiations Committee on the Committee and Trade and Environment in Special Session TN/TE/20 (21 April 2011)
APEC List of Environmental Goods	54 green products for which APEC Member states agreed to reduce applied tariff rates to 5% or less by the end of 2015	• 2012 APEC Leaders Declaration Annex C
OECD (1999) Illustrative Product List of Environmental Goods	List of 121 illustrative environmental products developed by the OECD/Eurostat Informal Working Group	• OECD (1999), "Future Liberalisation of Trade in Environmental Goods and Services: Ensuring Environmental Protection as well as Economic Benefits" • A Comparison of the APEC and OECD Lists", *OECD Trade and Environment Working Paper No. 2005-04* Table A1.
List of 257 customised products developed by the OECD	List of 257 customised products developed by the OECD	• Sauvage (2014), "The Stringency of Environmental Regulations and Trade in Environmental Goods", *OECD Trade and Environment Working Papers*, 2014/03

6.7.2 Products

Table 6: Top 10 Green Products by PCI

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
1	2.5073	901380	Optical devices, appliances and instruments, nes	Solar Heliostats (Heliostats orient mirrors in concentrated solar power systems to reflect sunlight on to a CSP receiver)	APEC, OECD (2014)
2	2.0716	902790	Microtomes, parts of scientific analysis equipment	Microtomes are devices that prepare slices of samples for analysis - used in environmental monitoring	APEC, OECD (1999), OECD (2014)
3	2.0134	847989	Machines and mechanical appliances, nes	Machines and appliances designed for a wide range of areas of environmental management including waste, waste water, drinking water production and soil remediation	WTO Sample, APEC, OECD (1999), OECD (2014)
4	1.8805	902730	Spectrometers, spectrophotometers, etc using light	Used in a wide range of environmental applications, including identification of unknown chemicals, toxins and trace contaminants, environmental control, water management, food processing, agriculture and weather monitoring	WTO Sample, APEC, OECD (1999), OECD (2014)
5	1.8625	902780	Equipment for physical or chemical analysis, nes	Used to measure, record, analyse and assess environmental samples or environmental influences	APEC, OECD (1999), OECD (2014)
6	1.8291	680690	Mineral heat or sound insulating materials and articles	Used for heat and energy management	OECD (2014)
7	1.8119	902720	Chromatographs, electrophoresis instruments	Used to monitor and analyse air pollution emissions, ambient air quality, water quality, etc.	APEC, OECD (1999), OECD (2014)
8	1.8077	902710	Gas/smoke analysis apparatus	Used for monitoring and analysing environmental pollution.	APEC, OECD (1999), OECD (2014)
9	1.7945	847990	Parts of machines and mechanical appliances nes	Parts for environmental management devices (Machines and appliances designed for a wide range of areas of environmental management including waste, waste water, drinking water production and soil remediation)	APEC, OECD (2014)
10	1.7795	848360	Clutches, shaft couplings, universal joints	Used for initial assembly, repair, and maintenance of wind energy systems	OECD (2014)

Table 7: Bottom 10 Green Products by PCI

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
284	−1.2445	871200	Bicycles, other cycles, not motorized	Cleaner or more resource efficient technology or product	OECD (2014)
285	−1.2826	871411	Motorcycle Saddles	Cleaner or more resource efficient technology or product	OECD (2014)
286	−1.2935	220710	Undenatured ethyl alcohol > 80% by volume	Renewable Energy Plant	OECD (1999)
287	−1.4189	560790	Twine, cordage, ropes and cables, of other materials	Environmentally preferrable product	OECD (2014)
288	−1.5074	960310	Brooms/brushes of vegetable material	Waste collection equipment (solid waste management)	OECD (1999)
289	−1.6088	560721	Binder or baler twine, of sisal or agave	Environmentally preferrable product	OECD (2014)
290	−1.864	460120	Mats, matting and screens, vegetable plaiting material	Environmentally preferrable product	WTO Sample

Continued on next page

Table 7 – continued from previous page

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
291	−2.1905	530599	Vegetable fibre nes, processed not spun, tow and waste	Environmentally preferrable product	OECD (2014)
292	−2.2365	630510	Sacks and bags, packing, of jute or other bast fibres	Environmentally preferrable product	OECD (2014)
293	−2.9908	530310	Jute and other textile bast fibres, raw or processed but not spun	Environmentally preferrable product	OECD (2014)

Table 8: Top 10 Renewable Energy Products by PCI

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
1	2.5073	901380	Optical devices, appliances and instruments, nes	Solar Heliostats (Heliostats orient mirrors in concentrated solar power systems to reflect sunlight on to a CSP receiver)	APEC, OECD (2014)
2	2.0134	847989	Machines and mechanical appliances nes	Machines and appliances designed for a wide range of areas of environmental management including waste, waste water, drinking water production and soil remediation	WTO Sample, APEC, OECD (1999), OECD (2014)
3	1.7795	848360	Clutches, shaft couplings, universal joints	Used for initial assembly, repair, and maintenance of wind energy systems	OECD (2014)
4	1.7554	848340	Gearing, ball screws, speed changers, torque converte	Gearboxes transform the rotation of the blades of wind turbines into the speed required to produce renewable electricity	OECD (2014)
5	1.7419	841199	Parts of gas turbine engines except turbo-jet/prop	Parts for gas turbines, which generate electrical power from recovered landfill gas, coal mine vent gas, or biogas	APEC, OECD (2014)
6	1.4104	841181	Gas turbine engines nes of a power < 5000 kW	Gas turbines for electrical power generation from recovered landfill gas, coal mine vent gas, or biogas (clean energy system)	WTO Sample, OECD (2014)
7	1.2239	840619	Steam and vapour turbines nes	Turbines designed for the production of geothermal energy (renewable energy) and co-generation ((CHP) which allows for a more effective use of energy than conventional generation)	WTO Sample, OECD (2014)
8	1.216	903289	Automatic regulating/controlling equipment nes	Used in renewable energy and smart grid applications, as well as other process control instruments and apparatus for temperature, pressure, flow and level, and humidity	APEC, OECD (1999), OECD (2014)
9	1.1399	841950	Heat exchange units, non-domestic, non-electric	Provide cooling effect to heat exchangers in solar collector or solar system controllers to avoid overheating. Heat exchangers are also used in geothermal energy systems.	WTO Sample, OECD (1999), OECD (2014)
10	1.1216	840690	Parts of steam and vapour turbines	Parts for turbines designed for production of geothermal energy (renewable energy) and co- generation ((CHP) which allows for a more effective use of energy than conventional generation	APEC, OECD (2014)

Table 9: Bottom 10 Renewable Energy Products by PCI

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
48	−0.14782	761100	Aluminium reservoirs,vats, tanks, etc, volume >300l	Containers for the production of biogas, waste water management, drinking water production and solar thermal energy purposes.	OECD (2014)
49	−0.31929	850432	Transformers electric, power capacity 1-16 KVA, nes	Renewable Energy Plant	OECD (2014)

Continued on next page

Table 9 – continued from previous page

Rank	PCI	HS6 Code	Product Description	Environmental Benefits	Environmental Lists
50	−0.36528	850421	Liquid dielectric transformers < 650 KVA	Used for initial assembly, repair, and maintenance of wind energy systems	OECD (2014)
51	−0.38168	850161	AC generators, of an output < 75 kVA	Used in conjunction with boiler and turbines to generate electricity in renewable energy plants	OECD (2014)
52	−0.41086	850720	Lead-acid electric accumulators except for vehicles	Provides for energy storage in off-grid PV system	OECD (2014)
53	−0.56155	700992	Glass mirrors, framed	Renewable Energy Plant	OECD (2014)
54	−0.61976	730820	Towers and lattice masts, iron or steel	Used to elevate and support a wind turbine for the generation of renewable energy	WTO Sample, OECD (2014)
55	−0.63559	290511	Methyl alcohol	Renewable Energy Plant	OECD (1999)
56	−0.81094	850431	Transformers electric, power capacity < 1 KVA, nes	Renewable Energy Plant	OECD (2014)
57	−1.2935	220710	Undenatured ethyl alcohol > 80		

6.7.3 Countries

In Table 10, we present each country's GCI, GCP and ECI ranks for 2014. We also identify each country's most proximate green product that they are not yet competitive in and show the proximaty density of that product to the given country.

Table 10: Country Rankings and most proximate green product for 2014

Country	GCI Rank	GCP Rank	ECI Rank	Most proximate green product	Proximity Density
Germany	1	4	3	Webs, mattresses, other nonwoven fibreglass products	0.523527
Italy	2	1	24	Multiple-walled insulating units of glass	0.551191
United States	3	8	5	Vacuum pumps	0.393836
Austria	4	7	10	Manostats	0.407972
Denmark	5	18	20	Mineral and aerated waters not sweetened or flavoured	0.338434
China	6	2	38	Jute and other textile bast fibres, raw or retted	0.547997
Czech Republic	7	12	9	Parts of wash, filling, closing, aerating machinery	0.357498
France	8	5	12	Valves, pressure reducing	0.4277
Japan	9	15	1	Railway maintenance-of-way service vehicles	0.355749
United Kingdom	10	13	11	Compression refrigeration equipment with heat exchange	0.335603
Sweden	11	17	4	Gas/smoke analysis apparatus	0.316329
Spain	12	3	29	Mineral and aerated waters not sweetened or flavoured	0.486129
Slovenia	13	19	13	Domestic iron/steel solid fuel appliances, not cooker	0.299162
Poland	14	9	23	Mineral and aerated waters not sweetened or flavoured	0.398532
Hungary	15	23	16	Manostats	0.267572
Finland	16	30	6	Mufflers and exhaust pipes for motor vehicles	0.228582
Portugal	17	10	48	Brooms/brushes of vegetable material	0.422045
Estonia	18	24	28	Mineral and aerated waters not sweetened or flavoured	0.284672
Switzerland	19	32	2	Clutches, shaft couplings, universal joints	0.230932
Romania	20	22	39	Liquid dielectric transformers < 650 KVA	0.282028
Croatia	21	26	36	Building blocks, bricks of cement, or artificial ston	0.307701
Mexico	22	40	22	Gas supply/production/calibration meters	0.170645
Slovakia	23	28	18	Prefabricated structural items of cement or concrete	0.241184
Bulgaria	24	20	46	Mineral and aerated waters not sweetened or flavoured	0.327624
Turkey	25	6	56	Brooms/brushes of vegetable material	0.462755
Lithuania	26	16	34	Cans, iron/steel, capacity <50l closed by crimp/solde	0.336749
South Korea	27	29	8	Bicycle brakes, parts thereof	0.245542
India	28	11	50	Brooms/brushes of vegetable material	0.388543
Israel	29	34	21	Surveying, etc instruments nes	0.1938
Latvia	30	25	35	Tank, cask or container, iron/steel, capacity 50-300l	0.280853
Malaysia	31	41	27	Parts and accessories of optical appliances nes	0.182011
Lebanon	32	31	58	Building blocks, bricks of cement, or artificial ston	0.262736
Thailand	33	21	40	Mats, matting and screens, vegetable plaiting material	0.268803
Netherlands	34	14	15	Surveying, etc instruments nes	0.344208
Singapore	35	47	7	Optical devices, appliances and instruments, nes	0.183491
Ukraine	36	42	31	Sheet etc, cellular of polymers of styrene	0.189066
Bosnia and Herzegovina	37	39	47	Liquid dielectric transformers < 650 KVA	0.216554
Tunisia	38	43	76	Sacks and bags, packing, of jute or other bast fibres	0.231006
Belarus	39	46	30	Mineral and aerated waters not sweetened or flavoured	0.186029
Norway	40	66	17	Anhydrous ammonia	0.112998

Continued on next page

Table 10 – continued from previous page

Country	GCI Rank	GCP Rank	ECI Rank	Most proximate green product	Proximity Density
South Africa	41	44	43	Sacks and bags, packing, of jute or other bast fibres	0.205328
Canada	42	37	19	Railway cars nes, closed and covered	0.204049
Philippines	43	49	64	Brooms/brushes of vegetable material	0.167938
Greece	44	33	52	Sacks and bags, packing, of jute or other bast fibres	0.24185
Hong Kong	45	27	41	Bicycle hubs, free-wheel sprocket wheels	0.264135
Brazil	46	52	32	Railway cars nes, closed and covered	0.139184
Vietnam	47	38	92	Jute and other textile bast fibres, raw or retted	0.282421
Egypt	48	35	71	Sacks and bags, packing, of jute or other bast fibres	0.278484
Indonesia	49	36	78	Jute and other textile bast fibres, raw or retted	0.273387
Moldova	50	57	77	Mineral and aerated waters not sweetened or flavoured	0.159036
Jordan	51	55	62	Sacks and bags, packing, of jute or other bast fibres	0.181749
Ireland	52	62	14	Surveying, etc instruments nes	0.125117
Kenya	53	63	82	Sacks and bags, packing, of jute or other bast fibres	0.195668
Russia	54	70	25	Gas turbine engines nes of a power < 5000 kW	0.10655
Uganda	55	64	67	Undenatured ethyl alcohol > 80% by volume	0.166241
Senegal	56	69	72	Surveying, etc instruments nes	0.134271
El Salvador	57	54	83	Brooms/brushes of vegetable material	0.183812
New Zealand	58	50	33	Surveying, etc instruments nes	0.164334
Macedonia	59	58	73	Brooms/brushes of vegetable material	0.169198
Dominican Republic	60	59	74	Sacks and bags, packing, of jute or other bast fibres	0.179649
United Arab Emirates	61	81	57	Anhydrous ammonia	0.098455
Sri Lanka	62	51	110	Sacks and bags, packing, of jute or other bast fibres	0.24724
Malawi	63	87	70	Surveying, etc instruments nes	0.072619
Guatemala	64	48	85	Sacks and bags, packing, of jute or other bast fibres	0.237824
Costa Rica	65	77	55	Chlorine	0.089105
Tanzania	66	72	96	Sacks and bags, packing, of jute or other bast fibres	0.169185
Argentina	67	61	37	Methyl alcohol	0.133796
Georgia	68	78	60	Undenatured ethyl alcohol > 80% by volume	0.088708
Pakistan	69	45	107	Jute and other textile bast fibres, raw or retted	0.307798
Morocco	70	53	100	Brooms/brushes of vegetable material	0.2086
Honduras	71	67	91	Undenatured ethyl alcohol > 80% by volume	0.14617
Cameroon	72	94	80	Undenatured ethyl alcohol > 80% by volume	0.054028
Albania	73	68	109	Brooms/brushes of vegetable material	0.162402
Colombia	74	73	54	Undenatured ethyl alcohol > 80% by volume	0.101699
Kyrgyzstan	75	83	95	Sacks and bags, packing, of jute or other bast fibres	0.083056
Peru	76	65	84	Sacks and bags, packing, of jute or other bast fibres	0.149909
Oman	77	100	61	Mineral and aerated waters not sweetened or flavoured	0.044163
Mauritius	78	56	89	Brooms/brushes of vegetable material	0.186092
Zimbabwe	79	82	75	Jute and other textile bast fibres, raw or retted	0.088834
Australia	80	71	44	Methyl alcohol	0.112851
Madagascar	81	74	117	Brooms/brushes of vegetable material	0.148984
Kazakhstan	82	90	45	Anhydrous ammonia	0.07086
Kuwait	83	106	42	Methyl alcohol	0.032296
Uruguay	84	75	49	Undenatured ethyl alcohol > 80% by volume	0.0896
Uzbekistan	85	85	99	Sacks and bags, packing, of jute or other bast fibres	0.08202
Mozambique	86	96	104	Jute and other textile bast fibres, raw or retted	0.072479
Ethiopia	87	88	114	Sacks and bags, packing, of jute or other bast fibres	0.10837
Iran	88	95	59	Liquid dielectric transformers < 650 KVA	0.042025
Bangladesh	89	80	122	Brooms/brushes of vegetable material	0.14754
Nicaragua	90	86	115	Sacks and bags, packing, of jute or other bast fibres	0.111789
Panama	91	60	68	Sacks and bags, packing, of jute or other bast fibres	0.169703
Yemen	92	99	79	Sacks and bags, packing, of jute or other bast fibres	0.056184
Cote dIvoire	93	91	94	Surveying, etc instruments nes	0.066838
Paraguay	94	89	86	Sacks and bags, packing, of jute or other bast fibres	0.069216
Ecuador	95	93	98	Sacks and bags, packing, of jute or other bast fibres	0.067348
Jamaica	96	84	66	Mineral and aerated waters not sweetened or flavoured	0.082543
Ghana	97	97	102	Undenatured ethyl alcohol > 80% by volume	0.059005
Chile	98	76	53	Anhydrous ammonia	0.091139
Laos	99	98	119	Sacks and bags, packing, of jute or other bast fibres	0.070638
Mali	100	102	93	Sacks and bags, packing, of jute or other bast fibres	0.056515

Continued on next page

Table 10 – continued from previous page

Country	GCI Rank	GCP Rank	ECI Rank	Most proximate green product	Proximity Density
Republic of the Congo	101	111	81	Sacks and bags, packing, of jute or other bast fibres	0.023712
Saudi Arabia	102	103	26	Manganese oxides other than manganese dioxide	0.033556
Zambia	103	92	65	Sacks and bags, packing, of jute or other bast fibres	0.066985
Cambodia	104	79	121	Jute and other textile bast fibres, raw or retted	0.153735
Gabon	105	115	87	Surveying, etc instruments nes	0.017623
Venezuela	106	118	69	Surveying, etc instruments nes	0.011949
Guinea	107	110	113	Sacks and bags, packing, of jute or other bast fibres	0.03639
Qatar	108	120	51	Buoys, beacons, coffer-dams, pontoons, floats nes	0.009239
Trinidad and Tobago	109	107	63	Mineral and aerated waters not sweetened or flavoured	0.029218
Algeria	110	119	88	Sodium hydroxide (caustic soda) solid	0.009689
Nigeria	111	108	108	Jute and other textile bast fibres, raw or retted	0.029281
Bolivia	112	104	103	Jute and other textile bast fibres, raw or retted	0.042712
Papua New Guinea	113	114	118	Sacks and bags, packing, of jute or other bast fibres	0.02736
Mongolia	114	105	97	Sacks and bags, packing, of jute or other bast fibres	0.032406
Sudan	115	112	120	Sacks and bags, packing, of jute or other bast fibres	0.033396
Libya	116	121	106	Sacks and bags, packing, of jute or other bast fibres	0.010857
Liberia	117	117	111	Surveying, etc instruments nes	0.016883
Tajikistan	118	101	105	Jute and other textile bast fibres, raw or retted	0.046313
Azerbaijan	119	113	90	Methyl alcohol	0.020978
Angola	120	122	112	Methyl alcohol	0.003356
Mauritania	121	109	101	Sacks and bags, packing, of jute or other bast fibres	0.027647
Turkmenistan	122	116	116	Sacks and bags, packing, of jute or other bast fibres	0.018702

6.7.4 Robustness checks for regression results

There is not enough within country variation in GCI and GCP for the relatively short period covered by our dataset to run a country-fixed-effect panel regression. Instead, we present additional regression analyses for different years covered by our dataset.

Table 11: Robustness tests for the relationship between GCI and Log Env. Patents over different years

	2010	2005	2000
GCI	1.144***	1.034***	1.205***
	(0.201)	(0.213)	(0.177)
ECI	0.956***	1.147***	0.782***
	(0.307)	(0.273)	(0.189)
Log GDP/Cap	0.208	0.003	0.111
	(0.150)	(0.127)	(0.102)
Intercept	0.632	2.338**	1.089
	(1.237)	(1.067)	(0.798)
Observations	122	122	122
Adjusted R^2	0.752	0.741	0.755

Robust standard errors in parenthesis.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01
The Env. Patent variable can be found in the OECD Statistics database under Environment – Innovation in environment-related tech – Technology Development (Family Size: one or greater; Technology Domain: Selected Environment-Related Technologies). Available at: http://stats.oecd.org/

Table 12: Robustness tests for the relationship between GCI and Log CO2/cap emissions over different years

	2010	2005	2000
GCI	−0.171*	−0.333***	−0.2765***
	(0.097)	(0.100)	(0.103)
ECI	0.048	0.377**	0.398**
	(0.158)	(0.152)	(0.153)
Log GDP/Cap	0.923***	0.767***	0.746***
	(0.091)	(0.080)	(0.087)
Intercept	−7.086***	−5.432***	−5.017***
	(0.807)	(0.679)	(0.708)
Observations	122	122	122
Adjusted R^2	0.756	0.755	0.715

Robust standard errors in parenthesis.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01
CO_2/cap (metric tons per capita) is sourced from https://data.worldbank.org/indicator/EN.ATM.CO2E.PC

Table 13: Robustness tests for the relationship between GCI and Environmental Policy Stringency over different years

	2010	2005	2000
GCI	0.085*	0.100**	0.099***
	(0.049)	(0.047)	(0.028)
ECI	−0.079	−0.096	−0.091
	(0.085)	(0.083)	(0.057)
Log GDP/Cap	0.236***	0.218***	0.155***
	(0.055)	(0.038)	(0.031)
Intercept	−1.154**	−1.122***	−0.704***
	(0.538)	(0.331)	(0.242)
Observations	31	31	31
Adjusted R^2	0.527	0.556	0.647

Robust standard errors in parenthesis.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01
Environmental Policy Stringency (EPS) data is sourced from http://stats.oecd.org/

Table 14: Green Complexity Potential Regression Analysis (10 year averages)

	Δ GCI $(t+\delta)$	Δ #Green exported products $(t+\delta)$	Δ Green export trade ratio $(t+\delta)$
Log GCP(t)	0.132***	5.223***	0.009***
	(0.028)	(1.194)	(0.003)
Log GDP/Cap(t)	0.004	0.018	0.001
	(0.017)	(0.757)	(0.001)
ECI(t)	−0.124***	−6.374***	−0.005*
	(0.032)	(1.508)	(0.003)
GCI(t)	−0.032		
	(0.037)		
#Green exported products(t)		0.086**	
		(0.040)	
Green export trade ratio(t)			−0.012
			(0.130)
Intercept	0.381*	17.989**	0.025*
	(0.170)	(7.449)	(0.013)
Observations	2440	2440	2440
Adjusted R^2	0.211	0.251	0.152

Robust standard errors in parenthesis.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01
t relates to country averaged values over years 1995-2004 and $t+\delta$ relates to country averaged values over years 2005-2014.

6.7.5 GCI robustness tests using alternative complexity measures

Since there are currently two alternative approaches for calculating the complexity of products, we compare the GCI regression results using different product complexity formulations. Here we use *GCI(HH)* to denote GCI calculated on the basis of the Hausmann et al. (2014) Product Complexity Index (as specified in equation 4) and GCI(Tacch) to denote GCI calculated on the basis of the alternative

Product Complexity measure proposed by Tacchella et al. (2012).

In Table 15 we show that the relationship between environmental patents, carbon emissions and environmental policy stringency are very similar for both GCI(HH) and GCI(Tacch). This suggests that the GCI is robust to the choice of product complexity measure.

Table 15: Comparison of GCI regression results using alternative complexity measures

	Log Env. Patents		Log CO_2/cap		Log EPS	
GCI(HH)	1.551***		−0.168**		0.070***	
	(0.174)		(0.080)		(0.022)	
GCI(Tacch)		1.584***		−0.180**		0.056**
		(0.169)		(0.077)		(0.021)
Log GDP/Cap	0.524***	0.532***	0.946***	0.948***	0.168***	0.175***
	(0.109)	(0.104)	(0.060)	(0.059)	(0.021)	(0.022)
Intercept	−1.913***	−1.975**	−6.990***	−7.010***	−0.757***	−0.805***
	(0.908)	(0.864)	(0.521)	(0.510)	(0.204)	(0.215)
Observations	1220	1220	2318	2318	558	558
Adjusted R^2	0.727	0.747	0.760	0.761	0.726	0.708

Robust standard errors in parenthesis.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
GCI (HH) refers to GCI calculated using the Hausmann et al. (2014) Product Complexity Index and GCI (Tacch) refers to GCI calculated using the Tacchella et al. (2012) Product Complexity measure Environmental patents data covers 2000 and 2005-2013, available from http://stats.oecd.org/. CO_2 (metric tons per capita) data covers 1995-2013, available from https://data.worldbank.org/indicator/EN.ATM.CO2E.PC. Environmental Policy Stringency (EPS) data covers 1995-2012, available from http://stats.oecd.org/. For all regressions, we take country averages over all available time periods.

6.7.6 Green stimulus - total spend

While our green stimulus analysis in Table 4 was based per capita green stimulus spending, here we show that our results are robust to using the *total* green stimulus spend in each country.

Table 16: Green Stimulus Total Spend

	Δ GCI $(t+\delta)$	Δ #Green exported products $(t+\delta)$	Δ Green export trade ratio $(t+\delta)$
Green Stimulus ($US Bn)	0.0028***	0.1276***	0.0001***
	(0.0004)	(0.0179)	(0.0000)
Log GDP/Cap(t)	0.0000	0.0000	0.0000
	(0.0000)	(0.0001)	(0.0000)
GCI(t)	−0.0042		
	(0.0153)		
Green exported products(t)		0.0171	
		(0.0184)	
Green export trade ratio(t)			0.0605
			(0.0479)
Intercept	−0.0054	−0.3881	−0.0080*
	(0.0368)	(1.7720)	(0.002)
Observations	19	19	19
Adjusted R^2	0.586	0.628	0.236

Robust standard errors in parenthesis.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01
t relates to the year 2008 and $(t+\delta)$ relates to the year 2011.
Green Stimulus Data is from Table 1 in Barbier et al. (2010), and relates to low carbon support for renewable energy, carbon capture and sequestration, energy efficiency, public transport and rail, and improving electrical grid transmission.
#Green exported products refers to the number of green exports in which the country has $RCA > 1$.

6.7.7 Calculation of Hausmann et al. (2014) Economic Complexity Index and Product Complexity Index

The ECI and PCI are calculated simultaneously according to the following three steps.

Hidalgo and Hausmann (2009) firstly define M_{cp} as a binary matrix based on each country's RCA. $M_{cp} = 1$ if country c's Revealed Comparative Advantage (RCA)

is greater than 1 and 0 otherwise. RCA is calculated as follows:

$$RCA_{c,p} = \frac{x(c,p)/\sum_{p} x(c,p)}{\sum_{c} x(c,p)/\sum_{c,p} x(c,p)} \tag{8}$$

where $x(c,p)$ is the value of country c's exports of product p. A country is considered to be *competitive* in product p if its RCA value is > 1).

Second, Hidalgo and Hausmann (2009) define two measures: country diversity ($k_{c,0}$), which represents the total number of products for which a country has RCA > 1 and product ubiquity ($k_{p,0}$), which represents the total number of countries exporting a product (with RCA > 1). These are given in Equations 9 and 10.

$$k_{c,0} = \sum_{p} M_{cp} \tag{9}$$

$$k_{p,0} = \sum_{p} M_{cp} \tag{10}$$

In the original *Method of Reflections*, Hidalgo and Hausmann (2009) take the country diversity and product ubiquity vectors as a starting point and use one to recursively 'correct' the other, as follows:

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_{p} M_{c,p} k_{pN-1} \tag{11}$$

$$k_{p,N} = \frac{1}{k_{cp,0}} \sum_{c} M_{c,p} k_{cN-1} \tag{12}$$

They show that by inserting Equation 12 into 11 and rewriting gives

$$\begin{aligned} k_{c,N} &= \frac{1}{k_{c,0}} \sum_{p} M_{c,p} \frac{1}{k_{p,0}} \sum_{c'} M_{c',p} k_{c',N-2} \\ &= \sum_{c'} k_{c',N-2} \sum_{p} \frac{M_{c,p} M_{c',p}}{k_{c,0}, k_{p,0}} \\ &= \sum_{c'} \widetilde{M_{c,c'}} k_{c',N-2} \end{aligned} \tag{13}$$

where

$$\widetilde{M_{c,c'}} \equiv \sum_{p} \frac{M_{c,p} M_{c',p}}{k_{c,0}, k_{p,0}} \tag{14}$$

As shown by Caldarelli et al. (2012), if $\vec{k_N}$ is the vector whose c'th element is $k_{c,N}$, then

$$\vec{k_N} = \widetilde{M} \times \vec{k_{N-2}} \tag{15}$$

where $\widetilde{M}$ is the matrix whose (c, c')th element is $\widetilde{M_{c,c'}}$. Taking N to infinity gives a distribution which remains fixed up to a scalar factor:

$$\widetilde{M} \times \vec{k_{N-2}} = \lambda \vec{k} \tag{16}$$

Therefore $\vec{k}$ is an eigenvector of $\widetilde{M}$. Since the eigenvector associated with the largest eigenvalue (which is 1 since $\widetilde{M}$ is row stochastic is constant, Hausmann et al. (2014) instead look for the eigenvector associated with the second largest eigenvalue. This vector, which is normalised by subtracting the average and dividing by the standard deviation as follows, is chosen to define the Economic

Complexity Index

$$ECI = \frac{\vec{K} - < \vec{K} >}{\text{stdev}(\vec{K})} \tag{17}$$

The Product Complexity Index (PCI) is calculated symmetrically by exchanging the index of countries (c) with products (p) in the equations outlined above.

Although there has been some confusion about what this measure captures and how it works, Mealy et al. (2017) showed that the algorithm corresponds to a standard clustering algorithm used to partition a similarity graph into two parts as to minimise the links between the parts. Viewed through this lens, one can interpret the ECI as collapsing the high dimensional space of countries and their export baskets onto a single dimension)where countries with similar exports are clustered together.

6.7.8 Calculation of Tacchella et al. (2012) Country Fitness and Product Complexity

An alternative approach for estimating the complexity of productive capabilities associated with countries and exported products has also being proposed by Tacchella et al. (2012). This methodology uses the same binary M_{cp} matrix constructed on the basis of countries' RCA's, as defined in section 6.7.7. However, Tacchella et al. (2012) introduce a different formulation for arriving at a country-specific estimate (called *Fitness*) and a product-specific estimate (called *Complexity*).

The measures are calculated as the fixed-point solution of the non-linear iterative

mapping given by

$$\begin{cases} \tilde{F}_c^{(N)} = \sum_p M_{cp} Q_p^{(N-1)} \\ \\ \tilde{Q}_p^{(N)} = \dfrac{1}{\sum_c M_{cp} \dfrac{1}{F_c^{(N-1)}}} \end{cases} \rightarrow \begin{cases} F_c^{(N)} = \dfrac{\tilde{F}_c^{(N)}}{\frac{1}{C} \sum_c \tilde{F}_c^{(N)}} \\ \\ Q_p^{(N)} = \dfrac{\tilde{Q}_p^{(N)}}{\frac{1}{P} \sum_p \tilde{Q}_p^{(N)}} \end{cases}, \tag{18}$$

where $\tilde{F}_c^{(N)}$ and $\tilde{Q}_p^{(N)}$ are the N^{th} iterations for the Fitness of country c and Complexity of the product p respectively, and P is the number of products. The initial conditions are given by vectors of 1's (i.e. $\tilde{F}_c^{(0)} = 1 \forall p$ and $\tilde{Q}_p^{(0)} = 1 \forall c$), and at each iteration, the intermediate variables $\tilde{F}_c^{(N)}$ and $\tilde{Q}_p^{(N)}$ are calculated and then normalised by the average values.

As shown in Cristelli et al. (2015, 2017), the Fitness measure appears useful for predicting the growth of countries falling into a particular region in the Fitness $\times$ GDP per capita plane.

7 Transformational change: Parallels for addressing climate and development goals

7.1 Introduction

Climate change and poverty alleviation are, as Stern (2016) has coined, 'the twin defining challenges of our century'. Historically, efforts to address these two challenges have been conflicted. Adverse impacts of climate change are likely to hit the poorest of this world hardest, but traditional industrial routes out of poverty are dangerously emissions-intensive. Such tensions have been major sticking points in earlier climate negotiations and largely underpinned the failure of the 2009 Copenhagen COP to reach a global climate agreement (Nordhaus, 2010).

However, two important developments suggest a new global readiness to move beyond historical conflicts and instead take advantage of key commonalities and collective interests. The 2015 adoption of the Sustainable Development Goals (SDGs) demonstrated an acute awareness that any plan to advance living standards of present and future generations must address the inseparable links between people, the planet and prosperity (UN, 2015, Griggs et al., 2013, Brown, 2015). Further, the Paris Agreement (UNFCCC, 2015), which has been ratified by an overwhelming majority of countries, provides a promising new international platform to progress a unique collective framework for global climate cooperation. The confluence of these global agendas represent an historic opportunity to marry efforts on climate and development fronts and drive significant progress on sustainable development.

Against this encouraging backdrop, this chapter draws attention to a somewhat under-appreciated, but profoundly important commonality in the twin climate and development challenges: both require societies to navigate and manage system-wide transformative change.

Unfortunately, transformational change is a concept that is not yet well defined or understood. To our knowledge, a generally accepted definition of transformational change does not as yet exist (Mersmann et al., 2014). There is also limited consensus on how it should be meaningfully measured – particularly as transformational change often involves dynamic processes occurring at multiple scales and dimensions (Geels et al., 2016, Turnheim et al., 2015). Further, while traditional economic modeling frameworks are well-suited for studying marginal changes over short-term horizons, they are poor tools for analysing dis-equilibrium dynamics, non-linear feedbacks and emergent properties that commonly characterise transformational change processes. As a better understanding of the process of transformational change could catalyse progress on both climate and development fronts, this chapter explores parallel efforts in respective fields.

First, we examine the nature and importance of transformational change processes in climate and development contexts. Unprecedented changes in both the low carbon landscape and global economic environment have taken many by surprise. While these unfolding dynamics are invalidating traditional analytical approaches based on assumed patterns of incremental change and challenging long-held notions about growth and development, they are also offering much needed alternative possibilities for achieving climate and development goals. Indeed, given the sheer magnitude of climate and development challenges – and the rate at which change

must occur to mitigate the worst climate impacts, being able to both navigate and *drive* the process of transformational change is now seen as a critical policy imperative.

Although significant efforts are underway to understand transformative change processes in respective climate and development fields, we argue that there are also key advantages from better integration across domains. For example, by considering the dynamics of economic development processes, climate policy can better account for mitigation opportunities, particularly as countries shift from energy intensive manufacturing activities towards services. Similarly, by taking advantage of the present impetus towards low-carbon futures, developing countries could seize an unprecedented opportunity to not only accelerate their growth, but also attain a much higher *quality* of sustainable, inclusive and resilient development.

Second, we explore empirical patterns of transformational change in climate and development contexts. We find that climate and development fields share a number of things in common. Aggregate changes (e.g. in GDP/capita or emissions) are often broken down into between-sector and within-sector changes, and key driving forces of change invariably relate to technology, preferences and their endogenous evolution. While historical stylized facts associated with the development process (such as the shift from agriculture to manufacturing to services) have been well documented, empirical patterns associated with low-carbon transformations are still emerging, and also likely to vary more across countries with different productive structures.

Third, we compare and contrast four alternative methodological tools for analysing and modeling transformational change: network analysis, computable general equi-

librium (CGE) models, macro-econometric/input-output models, and agent based models. Each has key strengths and weaknesses, which are important to bear in mind when applying them to policy questions relating to transformational change. Moreover, as models influence behavior in the real world, getting these right, or at least less wrong, is not a mere academic curiosum but could be vital for making simultaneous progress on climate and development goals.

This chapter proceeds as follows: In section 7.2 we examine the importance of transformational change in climate and development contexts and highlight key areas where more amalgamation across domains could be advantageous. Section 7.3 examines existing approaches to measure transformational change and identifies commonalities across climate and development frameworks. Section 7.4 reviews existing modeling methodologies and reflects on their ability to appropriately model transformational change. Section 7.5 concludes.

7.2 Transformational change in climate and development contexts

7.2.1 The importance of transformational change in achieving climate and development goals

The past decades have seen unprecedented transformations across both economic and low carbon fronts. Rapid technological progress, population growth, urbanisation and changing global market structures are combining to shape many countries' development trajectories in new and interesting ways (Fankhauser and McDermott, 2016). Equivalently, startling reductions in renewable energy costs and mounting

international cooperation to curb emissions are drastically changing the contours of the low carbon landscape (Trancik, 2014). Such changes have prompted a re-think of long-held notions about growth and development. They also invalidate commonly used traditional analytical approaches based on assumed patterns of incremental change. However, given the scale of climate and development challenges – and the pace at which change needs to occur to mitigate the worst climate impacts, navigating and *driving* the process of transformative change is now seen as an urgent policy imperative.

Indeed, many scholars are now stressing that without radical, non-marginal change limiting global warming to 2'C will be extremely difficult (Reid et al., 2010, Burch et al., 2014, Fankhauser and Stern, 2016). The magnitude of emissions reductions required in the rapidly diminishing time frame necessitates much more than incremental change along existing developmental and technological paradigms (Perez, 2015, Zenghelis, 2015). At the same time, developing countries are encountering a significant paradigm shift of their own. With traditional industrial development routes now potentially less viable due to impinging automation and globalization forces (Rodrik, 2016), development agencies and policy makers are facing the prospects of paving profoundly different pathways to prosperity in the 21st century.

While significant research efforts are underway to understand transformative change processes in respective climate and development fields, there are also key advantages from better integration across domains. In what follows, we outline key areas where climate policy could benefit from better accounting for economic transformation processes and equally, where development policy could benefit from the present impetus to transition towards low-carbon futures.

7.2.2 Why climate mitigation efforts should better account for economic transformation processes

The Paris Agreement (UNFCCC, 2015) represented a significant step forward in achieving universal consensus and commitment to move towards a sustainable future. However, there is still a significant gap between country Nationally Determined Contributions (NDCs) and the emissions reductions required by 2030 to plausibly remain on a 2'C pathway (Rogelj et al., 2016). In order to attain a reasonable chance of avoiding a rise in global average temperatures by more than 2'C, policy makers will need to harness as many emission reduction opportunities as possible. Better accounting for non-linear dynamics associated with low-carbon technology learning curves, uptake rates and shifts in consumer behavior and societal norms is a critical first step, and improved understanding of what drives these tipping points could help us speed them up (Russill and Nyssa, 2009, Aghion et al., 2014, Boyd et al., 2015).

However, understanding emissions implications of future economic structural shifts is also key. One of the biggest drivers of errors in historical energy and emissions projections is the failure to anticipate macroeconomic changes – particularly as a country shifts from energy intensive manufacturing activities towards services (Grubb et al., 2015). Most climate-economy models take the structure of the economy as fixed and rarely incorporate macroeconomic dynamics (Zenghelis, 2015). This is particularly problematic for countries like China, whose future emissions trajectory will have important climate consequences, but is also undergoing a process of rapid socio-economic transformation (Green and Stern, 2017). For example, Grubb et al. (2015) recently reviewed projections of 89 scenarios from 12 different

models for China's emissions through to 2030. They found that not only is the range of projections extremely large indicating a high degree of uncertainty, most scenarios do not account for the economy's macroeconomic structure and the potential to shift away from its currently high share of manufacturing. Given the consistent historical trend for countries to transition towards less energy-intensive service based activities as incomes rise, incorporating structural change dynamics in climate-economy models could depict very different climate implications of China's future development path.

While this chapter is primarily focused on drawing parallels between economic transformation and the low carbon transformation, it is worth noting that climate adaptation efforts can also benefit from a better understanding of the economic transformation process. Measures to lessen adverse climate change impacts have traditionally been approached as a static concept, with efforts to reduce vulnerability in developing countries often focusing on protecting existing structures and livelihoods, such as safeguarding agricultural output. However, as many developing countries are undergoing a process of socio-economic transformation, overly static adaptation plans could hamper development progress by placing too much emphasis on sectors that are likely to become less important over time (Bowen et al., 2016, Kocornik-Mina and Fankhauser, 2015). In light of the important links between climate risks and the development process, there has recently been increased interest in understanding what might constitute 'climate-resilient development' (Denton et al., 2014). This approach seeks to better account for the dynamism of the development process and calls for adaptation efforts to become more transformational (Fankhauser and McDermott, 2016).

7.2.3 Why development should better account for low-carbon transformation processes

When it comes to the key development objective of lifting people out of poverty, climate change has traditionally been seen as a key stumbling block, with the world's poorest and most vulnerable likely to shoulder the biggest burden of climate impacts (Collier et al., 2008, Brown, 2015, Hallegatte et al., 2015, Althor et al., 2016). Further, well-worn emissions-intensive industrial pathways that led today's advanced economies to prosperity are now recognized as being incompatible with the rapidly diminishing global carbon budget (IPCC, 2007, Stern, 2008). Consequently, the ability of developing countries to 'catch-up' to advanced countries' income levels by following well established manufacturing routes (Rodrik, 2012) might now be less feasible. This leaves many developing countries with the challenging prospect of scouting out new, un-trodden and uncertain paths to prosperity.

However, with new momentum building in the low carbon space, particularly on the technological and international negotiation fronts, climate change may now present developing countries with an unprecedented opportunity for growth acceleration (ECA, 2016). Never in history have developing countries been able to leverage such compelling global pressure and financial assistance to adopt new technologies, develop better infrastructure and protect natural capital. Further, as development agencies and policy makers are now increasingly calling for new modes of sustainable, inclusive and resilient development, the 21st century could see developing countries attaining a much higher quality of development than their predecessors (Garnaut et al., 2013, Mlachila et al., 2014, UNIDO, 2016).

A particularly important case in point relates infrastructure development. The current stock of infrastructure presently accounts for 60 per cent of global greenhouse gas emissions and if existing, long-lived energy and transportation assets are not stranded, they are projected to commit the world to a significant degree of warming (Davis et al., 2010, Guivarch and Hallegatte, 2011, Pfeiffer et al., 2016) However, with the number of people living in cities projected to increase from 3.5 billion to 6.5 billion by 2050 (Fankhauser and Stern, 2016) and the overwhelming majority of this increase likely to occur in Africa and Asia (UN, 2014), developing countries have a one time opportunity to shape more clean, compact and coordinated cities and urban structures (Floater et al., 2014, NCE, 2016). Getting infrastructure right has an enormous potential to curb emissions and lock rapidly developing countries in to a low-carbon growth trajectory (Bhattacharya et al., 2015). Further, as noted by Fuller and Romer (2014),"nothing else will create as many opportunities for social and economic progress".

7.3 Empirical patterns of transformational change in climate and development contexts

Despite its widely recognized importance in achieving climate and development goals, transformational change is a concept that is not yet well defined or understood. Like notions of 'world peace' and 'living happily ever after', transformational change often inspires broad endorsement and significant approval, but is challenging to definitively pin down. To our knowledge a generally accepted definition does not as yet exist. Nevertheless, development and climate economists have examined empirical patterns of system-wide transformational change within their

respective fields. Although there are distinct differences, we highlight important commonalities that could stimulate shared learning.

7.3.1 Empirical patterns of structural change in the development process

When examining system-wide transformative change in the context of development, economists often distinguish between changes arising from two key processes. The first relates to the structural shifts between sectors, which usually occurs as resources are allocated from low productivity to higher productivity activities. The second relates to changes occurring within sectors, which are often related to economy-wide productivity increases due to improvements in key economic 'fundamentals', such as institutions or human capital (Rodrik, 2013).

In examining structural shifts, economists have found that over the last 200 years, economic growth and development has generally been associated with successive transitions across three broad sectors. Early in a country's development phase, most people and resources are engaged in subsistence agriculture, where labour productivity is very low (Herrendorf et al., 2014). However, as productivity improves, agricultural workers are released to engage in higher value-add industrial activities. Employment and value added shares in manufacturing rise begin rising as agriculture shares decline. As the economy advances further, the manufacturing share reaches a peak and begins declining, as the service sector becomes much more dominant. These empirical trends are illustrated in Figure 1, where Herrendorf et al. (2014) have plotted the empirical decline in agriculture, inverted U-shape in manufacturing and rise in services experienced by ten advanced countries.

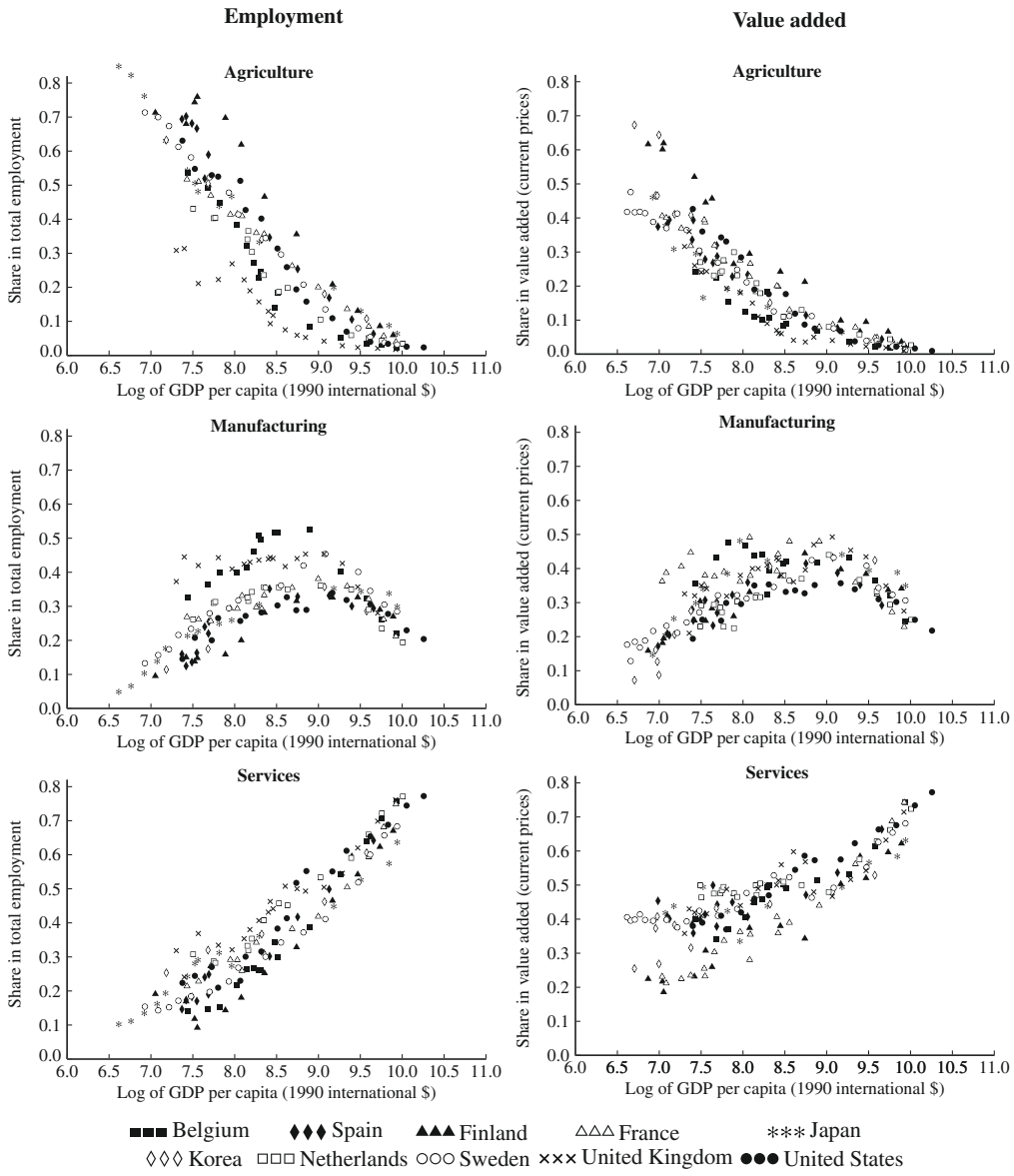


Figure 1: Sectoral shares of employment and added for selected developed countries: 1800-2000. Figure reproduced with the kind permission of Herrendorf et al. (2014).

To understand what drives the observed structural changes, researchers have focused on two key factors. The first relates to technological progress. As agricultural workers adopt new technologies and techniques, the increase in agricultural pro-

ductivity allows surplus labour to then be released to other industrial or services activities that pay higher wages (Lewis, 1954). Further, since rates of technological change differ across sectors, some sectors will generate higher improvements in growth and living standards than others (Chenery, 1986). The second factor relates to how consumer preferences change with rising income. As people become richer, they tend to spend a lower relative share of their income on food and more on durable (manufacturing) goods and services (Chenery, 1979, Cypher and Dietz, 2008). This increases domestic demand for new non-agricultural industries and economic activity.

7.3.2 Empirical patterns in the transition towards a low carbon economy

When examining shifts towards a low carbon economy, researchers have also examined empirical patterns in how countries' emissions change as the economy evolves. A sizable body of empirical research has examined the 'Environmental Kuznets Curve' (EKC) hypothesis, which postulates an inverted U shape relationship between pollutants and per capita income (Cole et al., 1997, Stern et al., 1996). One explanation for this posited relationship relates to the evolution of economic structure – as a nation progresses from relatively clean agricultural activities to dirty industrial manufacturing activities to cleaner service-based activities, one would expect emissions to rise and subsequently fall. A second explanation relates to the evolution of preferences – as societies become richer, the general expectation is that they are more likely to have a greater preference for environmental quality (Dinda, 2004). Empirical evidence strongly supports the upward sloping part of the EKC,

with many countries tending to experience an increase in emissions in the early phases of their development. However, there has been relatively inconclusive evidence to support the EKC's downward sloping section (Kaika and Zervas, 2013). A number of studies have found that urban or local air quality indicators that directly impact human health follow the hypothesized U relationship with income (Grossman and Helpman, 1993, Selden and Song, 1994, Stern and Common, 2001). For other pollutants, the trend observed among higher income countries tends to be quite mixed, with emissions tending to depend more on country-specific conditions, technologies and policies than its prosperity (Ota, 2017, Fankhauser and Jotzo, 2018).

The Kaya identify offers a different lens to view system-wide changes in countries' emissions (Raupach et al., 2007, Rosa and Dietz, 2012). This identity is specified as follows

$$C \equiv \frac{C}{E} \cdot \frac{E}{Y} \cdot Y \tag{1}$$

where C relates to emissions from human sources, E relates to energy consumption, and Y relates to economic output.

This formulation allows progress towards the low carbon economy to be analysed in terms of reductions in carbon intensity of energy ($\frac{C}{E}$), or reductions in the energy intensity of economic output ($\frac{E}{Y}$), or both. Fankhauser and Jotzo (2018) recently illustrated how countries' carbon intensity and energy intensity have evolved over time. As shown in Figure 2, energy intensity has been steadily decreasing in all countries over the 1990-2011 period. However, only high-income countries are

managing to decarbonize their energy.

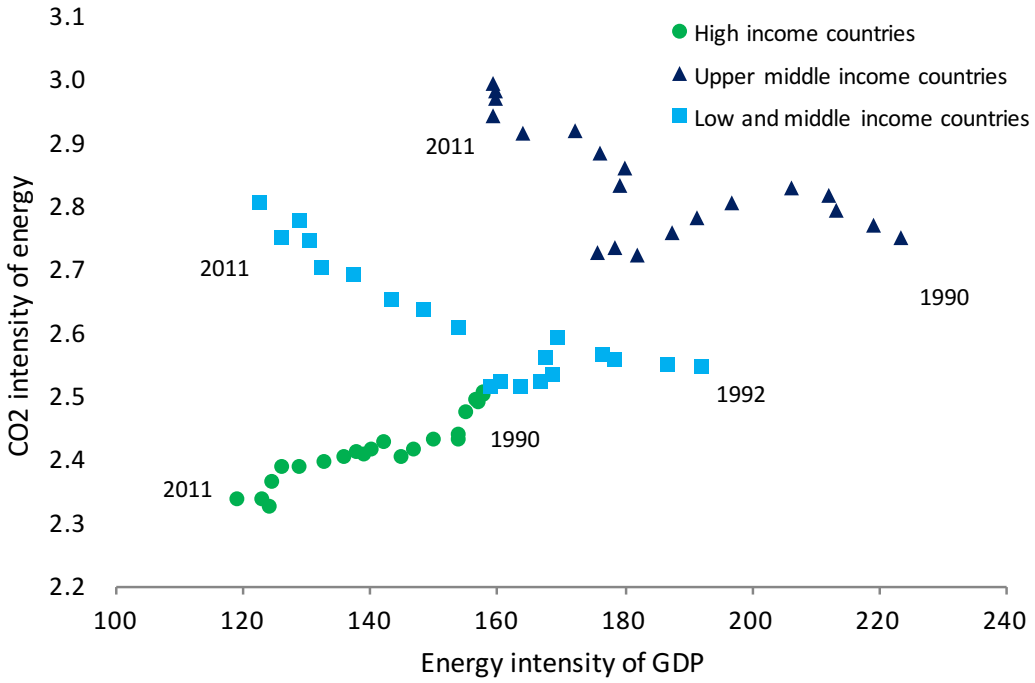


Figure 2: Long term trajectory of countries' carbon intensity of energy and energy intensity of GDP. Figure reproduced with the kind permission of Fankhauser and Jotzo (2018).

In a similar fashion to the study of empirical patterns in development economics, aggregate changes in carbon intensity and emissions intensity are also often broken down into structural shifts and system-wide efficiency improvements (Lenzen, 2016). However, unlike the relatively consistent structural dynamics and drivers observed in the economic development process, patterns in emissions and energy consumption across countries and sectors are more mixed. This in part is due to studies employing a range of different decomposition methodologies, industrial sector classifications and applying analysis to a wide diversity of regions and time periods (Su and Ang, 2012). But it is also likely that factors driving emissions are

different for different economies (Yao et al., 2015).

7.3.3 Key commonalities

In our brief overview of empirical frameworks commonly applied to understand patterns of change in development and emissions-related areas, three commonalities are apparent. First, both fields seek to understand observed dynamics in similar ways. Across the development and emissions-related literature, 'transformative change' is recognized as a phenomenon involving between sector changes and within sector changes. Key aggregates of interest are consequently decomposed into variation arising from sectoral shifts and system-wide efficiency improvements.

Second, both domains have significant commonalities in their key objective. While development economists are generally interested in what drives increases in GDP and climate economists are concerned with what reduces emissions, both tend to view system-wide progress through the lens of a productivity measure – for development, improving labour productivity is what matters, while for climate change, improving carbon productivity is key. However, it is this commonality that drives significant conflict. Traditionally, economies have increased their labour productivity by increasing their use of energy. Due to the emissions intensity of traditional energy sources, this in turn has generated greater emissions (Taylor, 2009).

The extent to which these twin objectives can be pursued together ultimately depends on our ability to decouple emissions from the development process (Hepburn and Bowen, 2013). As shown in Figure 2, many advanced economies are moving

in the right direction, particularly as increasing value add is derived from digitally enhanced material-light, intelligence-heavy activities (Baptist and Hepburn, 2012). However, current rates of progress are presently not sufficient to plausibly keep the planet on a 2'C pathway (NCE, 2014). A further critical sticking point remains for developing countries. As no advanced economy has ever become rich without undergoing an emissions-intensive industrial phase (Felipe et al., 2017), it is presently unknown whether development on the basis of 'industries without smoke stacks' is feasible (Page, 2015).

Such challenges underscore the pressing need to understand and leverage drivers of transformational change towards low-carbon development pathways. Incremental change along existing paradigms will clearly fall short (Perez, 2015). This brings us to the third commonality across climate and development research domains: in both economic and low carbon transformation studies, key driving forces of change generally relate to either technological improvements (which alter labour or carbon productivity rates) or demand-side factors relating to consumer preferences.

In relation to technological change, a key focus in the climate literature has been to better understand the non-linear dynamics associated with cost improvements in renewable energy technologies (Koh and Magee, 2008, Schilling and Esmundo, 2009). Research has also investigated the potential for policy to influence the pace of change through stimulating greater production and making directed investments in R&D (Menanteau et al., 2003, Fischer and Newell, 2008, Farmer and Lafond, 2016). The nature of knowledge spillovers is a further important factor, with research showing that low-carbon innovation tends to have greater positive benefits for the local economy than innovation incumbent, carbon-intensive technologies

(Dechezleprêtre et al., 2014a,b). Many studies have also examined the significant potential for developing countries to take advantage of new decentralized forms of energy distribution platforms, (such as mini-grids) and leap-frog the traditional centralized energy distribution systems that characterise most developed countries (Alstone et al., 2015, Levin and Thomas, 2016).

However, technological diffusion is also a critical facilitating (and limiting) factor in both development and low carbon transitions. In development research, the difficulty transferring productive knowledge from advanced countries to less developed countries is a common explanation for why some countries remain poor (Keller, 2004, Hidalgo et al., 2007, Hausmann et al., 2014, Bahar et al., 2014). Understanding the likelihood of successful technological transfer and its underpinning drivers is profoundly relevant for both development and climate policy makers seeking to facilitate transformative change. There are a number of important and encouraging developments in this field, which we will return to in the following section.

The nature of demand-side influences, particularly as they relate to underlying consumer preferences has received substantially less research attention in both climate and development fronts. In an extensive review of over 900 articles on structural change, Silva and Teixeira (2008) noted that only a small minority of articles examined demand-side factors. While there is a growing body of research relating to behavioural aspects of energy consumption and climate change awareness (Gowdy, 2008, Steg, 2008, Dietz et al., 2009, Gillingham et al., 2009) surprisingly little research seeks to address climate implications of how preferences may evolve over extensive time periods. The long time scales that development and

climate policy are often concerned with tends to invalidate standard approaches that take preferences as given and assume future preferences will look like the past. Further, as Mattauch and Hepburn (2016) argue, failing to account for preference evolution and the extent to which preferences can be shaped by policy can significantly overestimate the cost of climate mitigation. The same is also likely to be true for many cost-benefit analyses applied in development contexts, such as large-scale infrastructure investments. The nature of preference formation and its co-evolution with policy choices is an important area for future research that will likely have important implications for understanding transformative change in both development and climate change related areas.

7.4 Methodological tools for modeling and understanding transformational change in climate and development contexts

A wide range of analytical frameworks and decision-making tools are employed within development and climate domains. This section focuses on four key approaches and considers their relative strengths and weaknesses for helping researchers and policy makers better understand and manage the process of transformational change.

7.4.1 Network analysis

Network analysis, which has become a widely embraced analytical framework within a diverse range of disciplines, offers important avenues for understanding

transformational change. While traditional methods have tended to study transformational change in terms of sectoral shifts in resource allocation and within-sector productivity improvements, network analysis allows analysis of the connections *across sectors* and provides a means of estimating the probability of transitioning from one sector to another.

Hidalgo et al. (2007)'s application of network analysis to global trade data is one of the most notable contributions to the development context. Their 'Product Space' network, where traded products are represented as nodes linked to each other if they are more likely to require similar production capabilities (or know-how), has received significant attention from both scholars and policy makers as it provides a new lens to both visualize countries' productive structures and analyse feasible development pathways. Hidalgo et al. (2007) work is based on a theory advanced in the economic complexity and economic geography literature, which proposes that economic development is path dependent due to the underlying knowledge accumulation process. Just as it is easier to become good at making shirts if you already now how to make trousers, Hidalgo et al. (2007) used network analysis to demonstrate that it is easier for countries to become competitive in new products that require similar capabilities (or production know-how) to what they already know how to do.

This style of analysis has also been applied to regional and industry data (O'Clery et al., 2016, Neffke et al., 2011a, Boschma et al., 2013), labour-flows data (Neffke and Henning, 2013, Neffke et al., 2017) and input-output data (Radebach et al., 2016), and offers a number of important insights for developmentally oriented transformational change. First, not all product specialisations (and productive ca-

pabilities) are equal in terms of their knowledge spillover benefits and future diversification opportunities. The Product Space has a distinct core-periphery structure with manufacturing products (such as metals, machinery and chemicals) occupying the densely connected core of the network, and other products (such as agriculture and mining) tending to locate in the sparsely connected periphery. This finding further re-iterates the importance of manufacturing in the developmental process. Not only is manufacturing associated with unconditional convergence in labour productivity (Rodrik, 2012), which means that industries starting at lower labour productivity levels experience more rapid labour productivity growth. Manufacturing sectors also tend to have higher connectivity (in terms of shared capabilities and knowledge spillover opportunities) with other industries. As such, manufacturing has a unique ability to expand the set of 'adjacent possible' industries that are relatively easy to diversify into (Hausmann et al., 2015).

Second, the position of countries' exports in the Product Space can help identify specific new development opportunities that countries could more easily transition towards. Hidalgo et al. (2007) showed that countries were much more likely to diversify into products that were 'nearby' to their existing exports in the Product Space. By developing a measure (known as 'proximity density') that captures how similar a country's current exports are to a given undeveloped product in terms of their requisite production capabilities, they showed that while the probability of moving to exports that were far away (proximity density = 0.1) was almost zero, the probability of transitioning increases to 0.15 if a country's export basket contains closer products (proximity density = 0.8). The proximity density measure is particularly useful for industry players or policy makers seeking to identify

'adjacent possible' development opportunities that are more likely to be successful. However, when it comes to driving radical transformations in countries' productive structures, one of the key challenges is that it invariably involves big, difficult jumps in knowledge accumulation across the Product Space.

Recently, work has also sought to apply this type of analysis to examine transitions towards the green economy. By applying network analysis to a dataset of traded environmental products, Mealy and Teytelboym (2017) examined countries' green productive capabilities and their capacity to develop new green industries in the future. They showed that network-based measures, such as their Green Complexity Potential measure, were significantly predictive of future increases in countries' green trade competitiveness. Moreover, by exploring countries' positions in the 'green product space' and investigating how countries' competitiveness in green products evolved over time, Mealy and Teytelboym (2017) found that green diversification is strongly path-dependent: countries gaining early success in green capabilities tend to have much greater opportunities for diversifying into new product markets.

While recent efforts to incorporate network analysis into development and climate related research is yielding promising results, considerable future work awaits – particularly in better understanding the dynamic nature of networks. Although existing efforts have illuminated cross-country differences in productive structures, we are yet to understand how and why they change in particular ways. Further, very little research has examined the extent to which a nation's productive structure may be shaped by policy.

7.4.2 CGE models

The overarching paradigm currently used to conduct policy analyses in both fields is computable general equilibrium (CGE) modeling. CGE models are large-scale numerical models that aim to simulate how the economy might evolve in response to changes in policy, technology or other exogenous events (such as a drought or flood). Often described as an exercise of 'theory with numbers' (Wing and Balistreri, 2014) they encompass two key components: (i) detailed data on the structure of the economy and (ii) a theoretically derived system of equations dictating how the economy is likely to respond to particular changes. The underlying data used to calibrate CGE models are based on input-output tables or social accounting matrices. Aiming to provide a 'snapshot' of the economy at a particular point in time, these datasets include transaction values between key economic sectors (including government and representative households) and econometrically estimated parameters (or elasticities) capturing how different economic actors respond to changes in relative prices. The data are linked to a system of equations based on general equilibrium theory. These equations are numerically solved to determine how the simulated economy transitions to a new equilibrium by rebalancing supply and demand in different markets (Wing, 2004).

The ability of CGE models to incorporate data on the current state of the economy and provide theoretically grounded projections of its future evolution has made them popular tools within development and climate policy analysis. In development policy, CGE models have been frequently applied to examine the impacts of trade liberalization migration, industrial policy and pro-poor growth strategies (Ackerman and Gallagher, 2008, Devarajan and Robinson, 2013). In

climate policy, CGE models are commonly applied to examine the cost and benefits associated with the introduction of a carbon tax, emission-trading schemes and subsidy-based schemes such as feed-in tariffs (Wing, 2009, Adams and Parmenter, 2013). They form the core component of some Integrated Assessment Models (Bosetti et al., 2006, Scrieciu et al., 2013, Hasegawa et al., 2017), which are frequently used by policy makers to understand inter-relationships between economic, energy and climate systems (Farmer et al., 2015).

When applied to reasonably short-term horizons, CGE models have key strengths. Their capacity to explicitly capture interrelationships between markets for final goods, intermediate goods, government expenditure and households enable policy analysts to quantitatively study how the impacts of a particular change may filter through the economy and directly or indirectly impact different sectors and households (Wing, 2004). Unlike input-output models (discussed below), which are limited in their ability to incorporate actors' behavioral responses to prices and consequently tended to overestimate impacts associated with a given change, CGE models assign a more important role to prices and supply-side constraints (West, 1995). By explicitly specifying household utility functions, CGE models are also able to provide an estimate of the aggregate income and welfare impacts associated with a particular change. The ability to provide a quantitative estimate for impacts on 'winners' and 'losers' of proposed policies have made CGE models significantly influential in many policy debates (Devarajan and Robinson, 2002, Hughes et al., 2016).

However, when CGE models are applied to longer time horizons, or to the context of transformational change, CGE models encounter a number of issues. First,

many CGE models have encompass a relatively static framework, meaning that the modeled producers and consumers make optimizing decisions in a single period (Ahmed and O'Donoghue, 2007). In these models, policy analysis is based on a comparison of two alternative future equilibrium states of the economy – one with the policy change and one without. As these models do not explicitly represent the process of adjustment from one equilibrium to another, any economic impacts associated with transitional dynamics are unable to be accounted for (Scrieciu, 2007, Ackerman and Gallagher, 2008).

An alternative class of *dynamic* CGE models aims to better trace model variables over the projected time path – usually at yearly intervals. These models can be divided into two types: recursive dynamic models and forward looking models (Devarajan and Robinson, 2013). Unfortunately, neither of these options provide convincing frameworks for realistically capturing agent behaviours or structural dynamics associated with transformational change. Recursive dynamic CGE models assume that agents are completely myopic and make optimizing decisions only on the basis of current and past prices and other model variables. Agents' lack of consideration about the future prevents meaningful analysis of any inter-temporal investment and savings decisions (Babiker et al., 2009).

In contrast, forward-looking models do encompass agents that incorporate the future expectations into their decisions (Pratt et al., 2013). However, these models are usually characterized by agents who are unrealistically perfectly rational and omnisciently able to solve inter-temporal optimization problems over all modelled periods (Babiker et al., 2009, Richiardi, 2015). While dynamic stochastic general equilibrium (DSGE) models enable agents to explicitly incorporate uncertainty

about future states of the world, these models are usually so computationally intensive that much of the important economic sectoral detail (which makes these models useful in the first place) needs to be significantly simplified (Devarajan and Robinson, 2013). In addition, projections produced by most forward-looking models are, by construction, smooth, efficient adjustments along a balanced equilibrium path, which are hardly appropriate to characterise the turbulent, out-of equilibrium dynamics often associated with societies experiencing transformational change (Nordhaus and Tobin, 1972, Rezai et al., 2013).

A second issue complicating the application of CGE models to transformational change is their heavy reliance on equilibrium outcomes to characterize dynamics and future projections. Almost all CGE models assume capital and labour markets clear – usually instantaneously at each modeled time period. Not only is this at odds with historical observations in which economic activity was commonly subject to long, unanticipated recessions, persistent unemployment (Blanchard and Summers, 1987), 'secular stagnations' (Hansen, 1939) and institutional-induced capital (Harberger, 1959). It also misses essential characteristics of the development and low-carbon transition process that are important for policy makers to account for (DeCanio, 2003, Lane-Visser, 2015). For example, Chenery (1979) emphasizes that as developing economies are often characterized by the persistence of surplus labour and under-utilized capital, models should be designed to allow for the existence of disequilibrium rather than exclude it by assumption. Similarly, a number of scholars have argued that assuming a first-best world where all resources are optimally employed negates the potential economic benefits that green stimulus industrial policies could potentially provide to stagnant or depressed economies

(Barker et al., 2012, Hasselmann and Kovalevsky, 2013, Wolf et al., 2013). While recent developments in new Keynesian style DSGE models have incorporated frictions in price and wage adjustments and even the possibility of involuntary unemployment (Kemfert, 2003, Smets and Wouters, 2007, Christiano et al., 2010) , few have been applied to climate or development policy (Scrieciu et al., 2014, Fagiolo and Roventini, 2016).

7.4.3 Macro-econometric and input-output simulation models

Macro-econometric and input-output simulation models offer an alternative modeling framework that involves less restrictive assumptions about optimization and equilibrium. These are similar to CGE models in that they are based on detailed economic sectoral data (such as input-output or social accounting matrices) and are able to capture production and consumption flows across different industries. Their dynamics are also built on a system of equations. However, unlike CGE models, they are less strictly tied to neoclassical general equilibrium theory.

Macro-econometric and input-output simulation models instead draw on a more diverse range of economic fields (such as behavioral, ecological and evolutionary economics (Scrieciu et al., 2013)) and alternative economic paradigms (such as the Keynesian or neo-Keynesian framework, where demand is the key driving force of growth and supply adjusts to meet demand, subject to supply constraints (Boulanger and Bréchet, 2005, Rezai et al., 2013)). Instead of making optimization assumptions, they draw on historical data and econometrics to characterise key behavioural parameters and are said to implicitly characterise a form of bounded rationality (Barker et al., 2012). Macro-econometric simulation models also differ

to CGE models in their price formation. Rather than determining prices by imposing market-clearing assumptions, they employ a markup on unit costs, which depend on the level of competition in each sector (Cambridge Econometrics, 2014).

Macro-econometric and input-output simulation models are generally estimated using reasonably long stretches of time-series data and have the advantage that they can account for out-of-equilibrium dynamics that may have characterized historical observations. An important implication is that, unlike CGE models, they can capture the existence of unemployment and underutilized resources, which are particularly relevant considerations when assessing the *benefits* of industrial policies or green stimulus measures (Barker et al., 2016, IRENA, 2016). These types of models are also well suited to capture technological detail underpinning different sectors, as well as direct and indirect effects of endogeneous technological change (Lutz et al., 2007, Wiebe and Lutz, 2016).

On the one hand, macroeconometric models are less bounded by the 'straight-jacket' of strong general equilibrium theoretical impositions. However, on the other hand, they have to be much more intimately tied to historical data. This heavy reliance on past data to estimate future projections can be particularly problematic in the context of transformational change, as by definition, such a process entails a future state that could look significantly different from historical trends (Köhler et al., 2006, Scrieciu, 2007). Moreover, the fairly rigid format of input-output tables makes it challenging to consider how the industrial structure of an economy may evolve in response to climate or development policies.

7.4.4 Agent-based (multi-agent) models

Agent-based (or multi-agent) models offer yet another modeling paradigm. They aim to simulate the behavior of social or economic systems by explicitly representing heterogeneous, networked agents that interact and make decisions through prescribed behavioural rules (Bonabeau, 2002, Farmer and Foley, 2009). Agents do not need to necessarily make optimizing decisions, but can be programmed to make decisions in accordance with a number of behavioural typologies that are usually informed by empirical studies (Valbuena et al., 2008, Smajgl et al., 2011). A key advantage of this modeling paradigm, which is not accessible in CGE or macro-econometric simulation models, is its ability to explicitly model how macro-level dynamics arise from the (often probabilistic) interaction of these heterogeneous entities (Holland and Miller, 1991, Vespignani, 2012, Mercure et al., 2016). This emergence attribute is commonly examined in studies investigating social and collective phenomena, such as traffic jams and stock market crashes (Farmer and Foley, 2009). However, it is also known to be an important characteristic underpinning the 'tipping-point' dynamics associated with technological adoption and product diffusion rates (Dosi, 1982, Silverberg et al., 1988), social opinions (Watts and Dodds, 2007), and climate change awareness (Russill and Nyssa, 2009), which each play a key role in development and low-carbon transition processes.

A further important point is that while agent-based models (ABMs) do not explicitly enforce market clearing, equilibrium prices and quantities often do emerge as the result of agent's buying and selling decisions (Gintis, 2007). This framework consequently allows researchers and policy makers to analyse conditions under which equilibrium outcomes do and don't occur over different time horizons. In

addition, in the presence of multiple equilbria, ABMs are powerful tools to understand the probability and process by which economies end up in different states (such as clean vs dirty energy outcomes) (Arthur, 2006, Farmer et al., 2015).

In comparison to CGE and macro-econometric approaches, agent based modeling is still a relatively young and emerging analytical framework. However, its flexibility and intuitive framework for understanding change processes occurring within complex systems has seen it attract increasing interest, particularly within the climate-related literature. Many studies, which compare the relative merits of different approaches for analysing sustainability and climate policy-making, consistently provide strong appraisals and attest to its potential for offering important complementary insights to existing analytical tools (Boulanger and Bréchet, 2005, Bassi, 2014, Farmer et al., 2015, Balint et al., 2017).

However, there are two key challenges currently hindering faster progress and wider dissemination of ABMs within the policy landscape. First, there is a present lack of commonly accepted modeling standards within the ABM community. As a result, the ABM literature is replete with diverse and bespoke implementations, which often use a variety of different programming languages and design structure (Richiardi et al., 2006, Müller et al., 2013). Further, as models (and their output) are often complicated and poorly documented (Angus and Hassani-Mahmooei, 2015, Lee et al., 2015), replicating results or applying existing models to other contexts can be difficult and time-consuming (Richiardi, 2015). Second, estimation, calibration and validation procedures are often particularly challenging within ABMs (Fagiolo et al., 2007). These procedures aim to ensure the model and its parameters are as scientifically robust and defensible as possible. As ABMs often

involve a large number of parameters and allow macro-level dynamics to emerge from the interactions and behaviors of micro-level agents, their estimation and calibration can be significantly more involved than other modeling approaches discussed in this chapter.[1]

7.5 Conclusion

In light of the pressing need to better understand transformational change processes, particularly as they relate to climate and development contexts, this chapter has sought to draw these fields together, highlighting key commonalities and shared learning opportunities.

It is clear that better integration across climate and development domains is paramount – and there are clear advantages for both research and policy. In relation to research, climate and development economists have traditionally studied the process of transformative change in separate fields and with differing emphases. However, identifying key commonalities in respective change processes may not only improve shared learning outcomes, it could also illuminate a more generalized theory of transformational change. For policy, a lack of integration in climate and development initiatives can lead to outcomes that are at best, myopic and at worst, detrimental to their intended objectives.

In terms of methodological tools for analysing and modeling transformational

[1] That said, calibration and validation of CGE and DSGE models are also subject to criticism. CGE models are often applied to policy analysis without being exposed to any soft of validation process (Beckman et al., 2011, Van Dijk et al., 2016). Calibrating and validating DSGE models is both more challenging and more controversial, as models often require a number of ad-hoc tweaks or 'frictions' to enable them to fit the data (De Grauwe, 2010, Fagiolo and Roventini, 2012)

change, this chapter has reviewed four different approaches that have been used in both climate and development contexts. Network analysis provides a useful framework to investigate relationships across economic sectors, and allows scholars and policy makers to better understand technological diffusion and industrial transition possibilities. However, there has been relatively little work on the dynamism of economic networks or to understand how industrial structure may be shaped by policy. CGE models are the standard policy analysis tool in both climate and development fields, and while these models have key strengths when applied to short-term horizons, our review casts doubt on their ability to appropriately capture dis-equilibrium dynamics and emergent attributes of the transformative change process. While macro-econometric models deal better with out-of-equilibrium settings, their strong reliance on historical data weakens their ability to predict fundamentally new and different economic structures and dynamics.

In contrast, the flexibility of agent-based models make them a promising framework for analysing the formation of new economic arrangements and 'tipping point' dynamics emerging from complex interactions between heterogeneous agents. If appropriate standards and validation procedures can be progressed and adopted, these models could offer researchers and policy makers an important tool for better navigating and driving the process of transformational change to make simultaneous progress on climate and development goals.

8 Conclusion

This thesis has sought to advance a quantitative empirical framework for analysing productive knowledge within socio-economic systems. It was primarily motivated to better understand the *collective* use and distribution of societal knowledge – and to gain new perspectives into the complex, intertwined tapestry of capabilities that are so often obscured in aggregate production functions. It also aimed to demonstrate the value of these alternative analytical methods in the contexts of (i) long run economic development (ii) labour market segregation and job switching and (iii) the transition to the green economy.

This thesis built upon two key analytical lenses that were recently advanced in the economic complexity and economic geography literature to study the *complexity* (or sophistication) and *similarity* (or relatedness) of productive knowledge underpinning nations and regions. By refining and tinkering with these tools, this thesis uncovered important insights about the type of productive capabilities that underpin prosperous places, and how this has changed over time. Moreover, by adjusting the focus and pointing these lenses in new directions, this thesis shed new light on research areas with strong and salient policy relevance, such as the distribution and division of knowledge in the US labour force, and the type of productive capabilities relevant to the green economy. Finally, in taking a wider view of the climate and development disciplines, this thesis identified important parallels and opportunities across these fields for analysing, understanding and driving the process of transformational change in productive knowledge.

In this concluding chapter, I first synthesise and summarise this dissertation's

methodological contributions with respect to each analytical lens. I then discuss its broader contributions to the literature and the study of productive knowledge, highlighting key opportunities for future work.

8.1 Through the ECI-PCI looking glass

Although the ECI and PCI metrics were advanced to capture information about the diversity and sophistication of a countries' productive capabilities, this thesis has shown that, somewhat ironically, they actually capture information about place-based knowledge *similarity*. Comparable to the Dewey Decimal system for arranging books with similar topics close together in the library, the country-based ECI provides an ordering that places countries with similar exports close together in the ranking, and those with dissimilar exports far apart. The corresponding product-based PCI helps identify the type of exports that countries at either end of the ECI spectrum have in common.

As we demonstrated in Chapter 3 and 4, what is remarkable—and relatively unintuitive—is that the pattern of similarities uncovered by the ECI and PCI metrics turns out to consistently provide useful insights into the type of productive capabilities that distinguish richer places from poorer places.

Disentangling the *type* of productive activities places are specialised in from the *number* (or diversity) of productive activities they are specialised in enables a more focused discussion about the important, but different, roles these variables play in the economic development process. The relationship between export diversification and economic development is well-documented in the economics literature. Coun-

tries tend to diversify early in their developmental phase, and begin specialising at higher levels of per capita income (Imbs and Wacziarg, 2003). This U-shaped pattern aligns with other empirical studies that have reported a positive relationship between export diversification and economic growth, which tends to be stronger less developed countries (Al-Marhubi, 2000, Herzer and Nowak-Lehnmann D, 2006, Hesse, 2009). Explanations for the positive relationship between income and diversity are usually made with reference to the dampened impacts of sector-specific shocks (Acemoglu and Zilibotti, 1997) and preferences for diversity, which can be met with local industries where trading costs are high (Murphy et al., 1989). The tendency to specialise later in the development process is cast in terms of the efficiency benefits highlighted by traditional Ricardian trade theory, as well as theories of geographical agglomeration in 'New trade theory' and economic geography (Krugman, 1991d).

The explicit distinction between diversity and the ECI metrics provokes new explanations for the ECI's strong correlation with per capita GDP and future growth rates (Hidalgo and Hausmann, 2009, Hausmann et al., 2014). In Chapter 4, we suggested that these empirical results are, on reflection, both trivial and profound. It is trivial if one considers this notion in a two-region setting: if two places had identical productive capabilities, most economists would expect that these places would have very similar levels of per capita income. Moreover, if they did not, most economists would expect some degree of convergence: the region with lower per capita income should grow relatively more quickly and catch up with the living standards of the other region. However, in an N-region setting, this result becomes more intriguing. It is not particularly obvious that arranging N places in

a linear ordering that minimises the distance between regions with similar concentrations of exports, industries or occupations, would correlate so strongly with per capita GDP and predict future growth. And yet this result holds at the level of both countries and (US and UK) regions, and, unlike the relationship with state industrial and occupational diversity, over the entire 1880-2010 period. This suggests something quite fundamental about prosperity and the development process, which, to the best of my knowledge, theory is yet to explain.

The application of the ECI and PCI metrics to such a long stretch of US regional development history also yielded a further discovery about the analytical framework. The metrics appear to have a unique ability to shed light on the relevance of productive capabilities in changing economic contexts. It is important to stress that the ECI and PCI measures take in no information about income, wages or the classification of industries or occupations regions are specialised in. They only operate on the binary pattern of regional employment concentrations. However, as we demonstrate in Chapter 4, the structure of similarities underpinning this binary matrix is able to signal the increase in relevance (or 'fitness') of high-skill professional service activities and decline in manufacturing activities over the 1970-1980 period. While this finding suggests the ECI and PCI metrics could provide a new analytical lens to study structural change and the dynamics of productive capabilities more generally, further work is required to substantiate this. Having experienced a similar pattern of structural change over the same time period, the UK regional economy could be a particularly interesting avenue for future study (Martin et al., 2018).

More broadly, Chapter 3 showed that the ECI is equivalent to a widely used

spectral clustering algorithm that partitions a similarity graph into two parts. While spectral clustering methods have been applied in a diversity of settings, ranging from image recognition to RNA motif classification, we believe that the work of Hidalgo and Hausmann (2009) and Hausmann et al. (2014) represents the first application of these methods to glean insights about economic development from export data. Although the work in this thesis has emphasised the advantages of applying these metrics and employment data, there are numerous avenues for future work to apply these metrics to other datasets, such as country and regional input-output tables. Moreover, as the ECI and PCI metrics represent just one approach in a family of clustering and dimension reduction methods, further work could explore whether alternative techniques may provide additional insights into the process of economic development.

8.2 "Same-same but different" (and important) applications of knowledge similarity measures

This thesis also advanced novel measures of knowledge similarity (or relatedness) and applied them to new research territories with important policy relevance. Looking at the US labour force, Chapter 5 developed a new measure of occupational similarity on the basis of discrete tasks (or work activities) occupations shared in common. In addition to demonstrating that this measure was more predictive of job-to-job transitions than existing benchmark measures, we also showed that the task-based measure performed substantially better than other occupational descriptors available in the O*NET database (such as descriptors relating to occupational 'knowledge', 'abilities', 'skills', 'generalised work activities'

and 'detailed work activities').

From a methodological perspective, this occupational similarity measure is advantageous on three key fronts. First, it captures the relatedness of productive knowledge at the person-level, rather than the place-level. As existing knowledge-relatedness measures have tended to use export, industry and patent data to infer information about knowledge similarity at the regional (Frenken et al., 2007, Neffke et al., 2011a, Boschma et al., 2013, Kogler et al., 2013, Essletzbichler, 2015) or country level (Hidalgo et al., 2007), the occupational similarity metric developed here provides a useful addition to the analytical tool kit, which could have numerous applications. For example, work is already progressing to apply the occupational similarity measure to undertake a more detailed study of the potential employment impacts associated with automation (Del Rio-Chanona et al., 2018) and the clean energy transition (Mealy et al., 2018).

Second, the use of the occupational similarity measures to construct the Job Space network provide a novel, network-based lens to visualise the US labour force. As we demonstrated in Chapter 5, this can illuminate divisions of labour and job transition possibilities in new ways. Further, as no standard scheme exists for classifying the nature of work undertaken by society, the Work Activity Space network and its associated clusters of related work activities could provide the basis of a useful taxonomy of work.

Third, this analytical approach provides a complementary perspective to the 'task-based' framework (Autor et al., 2003, Autor, 2013), which has been adopted by a growing number of studies to examine labour market phenomena. While these studies tend to use continuous scale task intensity measures to examine the extent

to which occupations are concentrated in abstract, manual and routine tasks, this thesis provides an alternative methodology that (i) exploits data on more detailed discrete tasks, (ii) offers a new networks-based perspective to visualise abstract, manual and routines intense occupations, and (iii) provides additional information on job-to-job transition possibilities.

Chapter 6 demonstrated the policy-relevant advantages of applying knowledge-similarity measures to the context of the green economy. While this paper showed that some countries are better placed than others to succeed in the green economy, we also highlighted how knowledge-relatedness measures can help less well positioned countries take steps to improve their green competitiveness. In particular, identifying countries' *green adjacent possible* (GAP) allows policy makers to strategically assess green diversification opportunities – both in terms of their feasibility (similarity to a countries' current capabilities) and their potential for future knowledge spill-overs and learning opportunities across other industries. Moreover, aggregating countries' GAPs into a single measure – *green complexity potential* – turns out to be significantly predictive of countries' future competitiveness in green products, even after controlling for countries' current levels of per capita GDP, ECI and GCI.

And indeed, this paper's methodological contributions are already being put to good use in policy. The approaches were recently adopted by European Bank of Reconstruction and Development (EBRD) to provide green transition-oriented policy advice to a number of their countries of operation, and to guide detailed case studies for green diversification opportunities in Jordan and Ukraine (EBRD, 2017, Vivid Economics, 2018).

8.3 Broader contributions to the literature and the study of productive knowledge

This thesis has put forth five papers (four empirical papers and a handbook chapter), which each make contributions to key fields of literature. In broad terms, although the first paper (Chapter 3) was primarily a methodological paper, its findings have direct implications for the emerging literature on economic complexity – as well as wider implications for the field of economic development. The second paper (Chapter 4) brought the methodological insights from the first paper to bear on the economic geography literature, and highlighted new insights for structural change, regional income convergence and the decline of the US rustbelt. Focusing on the world of work, the third paper (Chapter 5) made novel contributions to the literature on tasks and the division of labour. The comprehensive dataset of green products and green complexity measures developed in the fourth paper (Chapter 6) yielded important findings relevant to the literature on green diversification and the transition to the green economy. Finally, the fifth paper (Chapter 7) contributed to the literature on transformational change, and highlighted a number of parallels and opportunities for collaboration between the climate and development fields.

Taken together, however, the work in this thesis also highlights some more general findings about productive knowledge. Firstly, and perhaps most fundamentally, this thesis has demonstrated that **what places and people know matters**. In Chapters 3 and 4, we showed that the *type* of productive knowledge concentrated in countries and regions has profound implications for prosperity. Moreover, although the relevance (or fitness) of particular capabilities changes in different contexts and

over time, we demonstrated that the ECI and PCI metrics are able to consistently identify the *type* of capabilities that underpin prosperous places – even in the midst of structural change. In Chapter 5, we highlighted that the *type* of knowledge workers 'learn by doing' key tasks in their day-to-day work activities has important ramifications for their future employment prospects and career path. Finally, in Chapter 6, we demonstrated that the *type* of knowledge embedded across nations is significantly informative of a country's likelihood of thriving in the green economy. It also helps identify country-specific opportunities for green diversification in the future.

Secondly, this thesis has revealed **distinct segregation in societal knowledge across places and people**. These divisions in knowledge could matter more than we realise. At the level of regions, Chapter 4 highlighted the new spatial 'bifurification' in manufacturing vs high-skill professional capabilities that began separating rust-belt states from other prosperous East-coast states (and California) from 1980 onwards. This divergence continues to be a pertinent economic and political issue (Austin et al., 2018), and calls for further research to investigate whether the type of productive capabilities that separate regions could influence a nation's capacity for inclusive growth. At the level of people, Chapter 5 used a networks-based lens to portray distinct divisions in the type of work undertaken by individuals in the US labour force. Here too, the marked pattern of segregation separating workers on the basis of skill, gender and susceptibility to automation motivates the need for further investigation into the potential implications for social mobility and inequality. At the level of nations, Chapter 6 identified significant differences in countries' capacities to benefit from the transition to the green econ-

omy. While some countries like Germany have capabilities that place them in good stead to take advantage of increased demand for environmentally friendly products, other countries (such as oil-rich nations) have competitive advantages that are less aligned to the direction of the green economy. As a successful transition to the green economy requires the participation of all countries, understanding how to minimise such differences is a particularly important area for future work.

Thirdly, the final paper in this thesis suggests that **future improvements in what places and people know will need to become less incremental and more transformational**, if we are to successfully address the 'twin challenges' of poverty alleviation and climate change in the 21st century. The literature on productive knowledge consistently shows that knowledge accumulation favours *small steps* (Hidalgo et al., 2007, Neffke et al., 2011a, Boschma et al., 2013, Essletzbichler, 2015, Neffke et al., 2018). However, given the pressing need to shift growth and development trajectories onto more climate-compatible pathways, this thesis concludes with an urgent call to better understand how to make *large leaps* in productive capabilities. Although impressive efforts are currently underway to better understand the process of transformational change in respective climate and development fields, better integration across these domains would undoubtedly accelerate progress and lead to more coherent and effective policy outcomes.

Finally, let me end where I started. The importance of understanding the collective use and evolution of productive knowledge remains as compelling today as it did when Thorstein Veblen emphasised its pertinence in 1898. As highlighted in this dissertation's literature review, the progress we have made in the last 120 years is considerable. However, today's proliferation of new, detailed datasets and

improved analytical power suggests that we have never been in a better position to make further advances in this valuable area of study. Moreover, the present challenges associated with securing prosperous and sustainable futures in the 21st century underscore the pressing need to do so. While this thesis has taken steps in this direction, much-needed advancements in the study of productive knowledge and its policy-relevant applications still lie ahead.

References

Abramovitz, M. (1993). The search for the sources of growth: areas of ignorance, old and new. *The Journal of Economic History 53*(2), 217–243.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier.

Acemoglu, D. and F. Zilibotti (1997). Was prometheus unbound by chance? risk, diversification, and growth. *Journal of Political Economy 105*(4), 709–751.

Ackerman, F. and K. P. Gallagher (2008). The shrinking gains from global trade liberalization in computable general equilibrium models: a critical assessment. *International Journal of Political Economy 37*(1), 50–77.

Adams, P. D. and B. R. Parmenter (2013). Computable general equilibrium modeling of environmental issues in australia: economic impacts of an emissions trading scheme. In *Handbook of computable general equilibrium modeling*, Volume 1, pp. 553–657. Elsevier.

Aghion, P., J. Boulanger, and E. Cohen (2011). Rethinking industrial policy. Technical report, Bruegel policy brief.

Aghion, P., A. Dechezleprêtre, D. Hemous, R. Martin, and J. Van Reenen (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy 124*(1), 1–51.

Aghion, P., C. Hepburn, A. Teytelboym, and D. Zenghelis (2014). Path dependence, innovation and the economics of climate change. *Centre for Climate Change Economics and Policy/Grantham Research Institute on Climate Change and the Environment Policy Paper & Contributing paper to New Climate Econ-*

omy.

Ahmed, V. and C. O'Donoghue (2007). CGE-microsimulation modelling: a survey.

Al-Marhubi, F. (2000). Export diversification and growth: an empirical investigation. *Applied Economics Letters 7*(9), 559–562.

Albeaik, S., M. Kaltenberg, M. Alsaleh, and C. A. Hidalgo (2017). 729 new measures of economic complexity (addendum to improving the economic complexity index). *arXiv preprint arXiv:1708.04107*.

Albeaik, S., M. Kaltenberg, M. Alsaleh, and C. A. Hidalgo (2017, July). Improving the Economic Complexity Index. *ArXiv e-prints*.

Alder, S., D. Lagakos, and L. Ohanian (2014). The decline of the US rust belt: a macroeconomic analysis.

Allcott, H. and D. Keniston (2017). Dutch disease or agglomeration? the local economic effects of natural resource booms in modern America. *The Review of Economic Studies 85*(2), 695–731.

Allen, M. T., G. Waugh, M. Shaw, S. Tsacoumis, et al. (2012, 09). The Development and Evaluation of a New O*NET Related Occupations Matrix . Technical report, National Center for O*NET Development.

Alstone, P., D. Gershenson, and D. M. Kammen (2015). Decentralized energy systems for clean electricity access. *Nature Climate Change 5*(4), 305.

Althor, G., J. E. Watson, and R. A. Fuller (2016). Global mismatch between greenhouse gas emissions and the burden of climate change. *Scientific Reports 6*, 20281.

Anderson, P., K. Arrow, and D. Pines (1988). The economy as a complex evolving system. *Reading: Addison-Wesley*.

Angus, S. D. and B. Hassani-Mahmooei (2015). "Anarchy" reigns: A quantita-

tive analysis of agent-based modelling publication practices in JASS, 2001-2012. *Journal of Artificial Societies and Social Simulation 18*(4), 16.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2006). Offshoring in a knowledge economy. *The Quarterly Journal of Economics 121*(1), 31–77.

APEC (2012). APEC Leaders Declaration: Annex C. Technical report, Asia-Pacific Economic Cooperation.

Arntz, M., T. Gregory, and U. Zierahn (2016). The risk of automation for jobs in oecd countries: A comparative analysis. *OECD Social, Employment, and Migration Working Papers* (189), 0_1.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity: Economic and social factors*, pp. 609–626. Princeton University Press.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal 99*(394), 116–131.

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. University of Michigan Press.

Arthur, W. B. (2006). Out-of-equilibrium economics and agent-based modeling. *Handbook of Computational Economics 2*, 1551–1564.

Arthur, W. B., S. N. Durlauf, and D. A. Lane (1997). *The economy as an evolving complex system II*, Volume 27. Addison-Wesley Reading, MA.

Attaran, M. (1986). Industrial diversity and economic performance in US areas. *The Annals of Regional Science 20*(2), 44–54.

Austin, B., E. Glaeser, and L. H. Summers (2018). Saving the heartland: Place-based policies in 21st century America. In *Brookings Papers on Economic Activity, Conference Draft. Washington, DC: Brookings Institution. March.*

Autor, D. (2013). The" task approach" to labor markets: an overview. *Journal for Labour Market Research 46*(3), 185–199.

Autor, D. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives 29*(3), 3–30.

Autor, D., L. F. Katz, and M. S. Kearney (2006). The polarization of the US labor market. *American Economic Review 96*(2), 189–194.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–97.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Babiker, M., A. Gurgel, S. Paltsev, and J. Reilly (2009). Forward-looking versus recursive-dynamic modeling in climate policy analysis: A comparison. *Economic Modelling 26*(6), 1341–1354.

Bahar, D., R. Hausmann, and C. A. Hidalgo (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics 92*(1), 111–123.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School 33*(2), 99–123.

Balint, T., F. Lamperti, A. Mandel, M. Napoletano, A. Roventini, and A. Sapio (2017). Complexity and the economics of climate change: A survey and a look

forward. *Ecological Economics 138*, 252–265.

Balland, P.-A., R. Boschma, J. Crespo, and D. L. Rigby (2018). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 1–17.

Baptist, S. and C. Hepburn (2012). Material efficiency, productivity and economic growth. *Philosophical Transactions of the Royal Society A*.

Barbier, E. B. et al. (2010). Green stimulus, green recovery and global imbalances. *World Economics 11*(2), 149–177.

Barker, T., E. Alexandri, J.-F. Mercure, Y. Ogawa, and H. Pollitt (2016). GDP and employment effects of policies to close the 2020 emissions gap. *Climate Policy 16*(4), 393–414.

Barker, T., A. Anger, U. Chewpreecha, and H. Pollitt (2012). A new economics approach to modelling policies to achieve global 2020 targets for climate stabilisation. *International Review of Applied Economics 26*(2), 205–221.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management 17*(1), 99–120.

Barro, R. J. and X. Sala-i-Martin (1992). Convergence. *Journal of Political Economy 100*(2), 223–251.

Bassi, A. (2014). Using simulation models for green economy policy making: A comparative assessment. *Review of Business and Economics Studies 2*(1), 88.

Basso, G., G. Peri, and A. Rahman (2017). Computerization and immigration: Theory and evidence from the United States. Technical report, National Bureau of Economic Research.

Bastian, M., S. Heymann, M. Jacomy, et al. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM 8*, 361–362.

Beaudry, C. and A. Schiffauerova (2009). Who's right, marshall or jacobs? the localization versus urbanization debate. *Research Policy 38*(2), 318–337.

Beckman, J., T. Hertel, and W. Tyner (2011). Validating energy-oriented CGE models. *Energy Economics 33*(5), 799–806.

Bell, M., K. Pavitt, et al. (1995). The development of technological capabilities. *Trade, technology and international competitiveness 22*(4831), 69–101.

Berger, T. and C. B. Frey (2016). Did the computer revolution shift the fortunes of US cities? technology shocks and the geography of new jobs. *Regional Science and Urban Economics 57*, 38–45.

Berman, E. and L. T. Bui (2001). Environmental regulation and labor demand: Evidence from the south coast air basin. *Journal of Public Economics 79*(2), 265–295.

Berry, C. R. and E. L. Glaeser (2005). The divergence of human capital levels across cities. *Papers in regional science 84*(3), 407–444.

Bhattacharya, A., J. Oppenheim, and N. Stern (2015). Driving sustainable development through better infrastructure: Key elements of a transformation program. *Brookings Global Working Paper Series*.

Black, D., T. McKinnish, and S. Sanders (2005). The economic impact of the coal boom and bust. *The Economic Journal 115*(503), 449–476.

Blanchard, O. J. and L. H. Summers (1987). Hysteresis in unemployment. *European Economic Review 31*(1-2), 288–295.

Blau, F. D. and L. M. Kahn (2000). Gender differences in pay. *Journal of Economic Perspectives 14*(4), 75–99.

Blau, F. D., P. Simpson, and D. Anderson (1998). Continuing progress? trends in occupational segregation in the United States over the 1970s and 1980s. *Feminist*

Economics 4(3), 29–71.

Blinder, A. S. et al. (2009). How many US jobs might be offshorable? *World Economics 10*(2), 41.

Blinder, A. S. and A. B. Krueger (2013). Alternative measures of offshorability: a survey approach. *Journal of Labor Economics 31*(S1), S97–S128.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008*(10), P10008.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences 99*(suppl 3), 7280–7287.

Boschma, R. and C. Gianelle (2013). Regional branching and smart specialization policy.

Boschma, R., A. Minondo, and M. Navarro (2013). The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. *Economic Geography 89*(1), 29–51.

Bosetti, V., C. Carraro, M. Galeotti, E. Massetti, and M. Tavoni (2006). WITCH a world induced technical change hybrid model. *The Energy Journal*, 13–37.

Boulanger, P.-M. and T. Bréchet (2005). Models for policy-making in sustainable development: The state of the art and perspectives for research. *Ecological Economics 55*(3), 337–350.

Bowen, A., C. Duffy, and S. Fankhauser (2016). 'Green growth' and the new industrial revolution. *Policy Brief. Grantham Research Institute on Climate Change and the Environment and Global Green Growth Institute*.

Bowen, A. and C. Hepburn (2014). Green growth: an assessment. *Oxford Review*

of Economic Policy 30(3), 407–422.

Boyd, R., F. Green, and N. Stern (2015). The road to paris and beyond. *Centre for Climate Change Economics and Policy and Grantham Research Institute on Climate Change and the Environment*.

Bresnahan, T. F. and V. A. Ramey (1993). Segment shifts and capacity utilization in the US automobile industry. *The American Economic Review 83*(2), 213–218.

Brown, A. (2015). Climate change and Africa. *Nature Climate Change 5*, 811.

Bucher, H., J. Drake-Brockman, A. Kasterine, and M. Sugathan (2014). Trade in environmental goods and services: Opportunities and challenges. Technical paper, International Trade Centre, Geneva.

Burch, S., A. Shaw, A. Dale, and J. Robinson (2014). Triggering transformative change: a development path approach to climate change response in communities. *Climate Policy 14*(4), 467–487.

Caldarelli, G., M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, and A. Tacchella (2012). A network analysis of countries' export flows: firm grounds for the building blocks of the economy. *PloS one 7*(10), e47278.

Cambridge Econometrics (2014). E3ME technical manual, version 6.0. *Cambridge Econometrics, Cambridge, UK*.

Carrillo, J. (2004). Nafta: The process of regional integration of motor vehicle production. In *Cars, Carriers of Regionalism?*, pp. 104–117. Springer.

Castaldi, C., K. Frenken, and B. Los (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of us state-level patenting. *Regional Studies 49*(5), 767–781.

Chenery, H. (1986). Growth and transformation. *Industrialization and growth: A comparative study*.

Chenery, H. B. (1979). Structural change and development policy. Technical report.

Christiano, L. J., M. Trabandt, and K. Walentin (2010). Involuntary unemployment and the business cycle. Technical report, National Bureau of Economic Research.

Clark, G. L. (2018). Learning-by-doing and knowledge management in financial markets. *Journal of Economic Geography 18*(2), 271–292.

Cohen, W. M. and D. A. Levinthal (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly 35*(1), 128–152.

Cole, M. A., A. J. Rayner, and J. M. Bates (1997). The environmental kuznets curve: an empirical analysis. *Environment and development economics 2*(4), 401–416.

Collier, P., G. Conway, and T. Venables (2008). Climate change and Africa. *Oxford Review of Economic Policy 24*(2), 337–353.

Combes, P.-P., T. Magnac, and J.-M. Robin (2004). The dynamics of local employment in france. *Journal of Urban Economics 56*(2), 217–243.

Content, J. and K. Frenken (2016). Related variety and economic development: a literature review. *European Planning Studies 24*(12), 2097–2112.

Cotter, D., J. M. Hermsen, and R. Vanneman (2011). The end of the gender revolution? gender role attitudes from 1977 to 2008. *American Journal of Sociology 117*(1), 259–89.

Cristelli, M., A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero (2013). Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one 8*(8), e70726.

Cristelli, M., A. Tacchella, and L. Pietronero (2015). The heterogeneous dynamics

of economic complexity. *PloS one 10*(2), e0117174.

Cristelli, M. C. A., A. Tacchella, M. Z. Cader, K. I. Roster, and L. Pietronero (2017). On the predictability of growth.

Cypher, J. and J. L. Dietz (2008). *The process of economic development*. Routledge.

David, P. A. (1985). Clio and the economics of QWERTY. *The American Economic Review 75*(2), 332–337.

David, P. A. (1994). Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions. *Structural change and economic dynamics 5*(2), 205–220.

Davis, S. J., K. Caldeira, and H. D. Matthews (2010). Future co2 emissions and climate change from existing energy infrastructure. *Science 329*(5997), 1330–1333.

De Grauwe, P. (2010). The scientific foundation of dynamic stochastic general equilibrium (dsge) models. *Public choice 144*(3-4), 413–443.

de Groot, H. L., J. Poot, and M. J. Smit (2016). Which agglomeration externalities matter most and why? *Journal of Economic Surveys 30*(4), 756–782.

DeCanio, S. (2003). *Economic models of climate change: a critique*. Springer.

Dechezleprêtre, A., R. Martin, and M. Mohnen (2014a). *Knowledge spillovers from clean and dirty technologies*. Centre for Economic Performance, London School of Economics and Political Science.

Dechezleprêtre, A., R. Martin, and M. Mohnen (2014b). Policy brief: Clean innovation and growth.

Del Rio-Chanona, R. M., P. Mealy, F. Lafond, M. Beguerisse, and J. D. Farmer (2018). Occupation mobility networks and the impact of automation on employ-

ment. *Mimeo*.

Delgado, M., M. E. Porter, and S. Stern (2014). Clusters, convergence, and economic performance. *Research Policy 43*(10), 1785–1799.

Denton, F., T. J. Wilbanks, A. C. Abeysinghe, I. Burton, Q. Gao, M. C. Lemos, T. Masui, K. L. O'Brien, and K. Warner (2014). Climate-resilient pathways: adaptation, mitigation, and sustainable development. *Climate change*, 1101–1131.

Devarajan, S. and S. Robinson (2002). The influence of computable general equilibrium models on policy. international food policy institute, trade and macroeconomic division. Technical report, Discussion Paper.

Devarajan, S. and S. Robinson (2013). Contribution of computable general equilibrium modeling to policy formulation in developing countries. In *Handbook of Computable General Equilibrium Modeling*, Volume 1, pp. 277–301. Elsevier.

Dietz, T., G. T. Gardner, J. Gilligan, P. C. Stern, and M. P. Vandenbergh (2009). Household actions can provide a behavioral wedge to rapidly reduce us carbon emissions. *Proceedings of the National Academy of Sciences 106*(44), 18452–18456.

Dinda, S. (2004). Environmental kuznets curve hypothesis: a survey. *Ecological Economics 49*(4), 431–455.

Diodato, D., F. Neffke, and N. O'Clery (2018). Why do industries coagglomerate? how marshallian externalities differ by industry and have evolved over time. Technical report, Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography.

Dopfer, K. (2005). *The evolutionary foundations of economics*. Cambridge University Press.

Dopfer, K., J. Foster, and J. Potts (2004). Micro-meso-macro. *Journal of Evolutionary Economics 14*(3), 263–279.

Dopfer, K. and J. Potts (2004). Evolutionary realism: a new ontology for economics. *Journal of Economic Methodology 11*(2), 195–212.

Dopfer, K., J. Potts, et al. (2008). *The general theory of economic evolution*. Routledge London.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy 11*(3), 147–162.

Dosi, G., C. Freeman, R. Nelson, G. Silverberg, and L. Soete (1988). Technical change and economic theory. Technical report, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

Duranton, G. and D. Puga (2000). Diversity and specialisation in cities: why, where and when does it matter? *Urban Studies 37*(3), 533–555.

Duranton, G. and D. Puga (2005). From sectoral to functional urban specialisation. *Journal of Urban Economics 57*(2), 343–370.

Easterlin, R. (1960). Interregional differences in per capita income, population, and total income, 1840-1950. In *Trends in the American economy in the nineteenth century*, pp. 73–140. Princeton University Press.

EBRD (2017). *Transition Report 2017-18*. European Bank for Reconstruction and Development, 22 November. ISBN: 2017 978-1-898802-46-5.

ECA (2016). *Greening Africa's Industrialisation, Economic Report on Africa*. Economic Commission for Africa, Addis Ababa.

England, P. and S. Li (2006). Desegregation stalled: The changing gender composition of college majors, 1971-2002. *Gender & Society 20*(5), 657–677.

Essletzbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies 49*(5), 752–766.

Fagiolo, G., A. Moneta, and P. Windrum (2007). A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics 30*(3), 195–226.

Fagiolo, G. and A. Roventini (2012). Macroeconomic policy in DSGE and agent-based models. *Revue de l'OFCE* (5), 67–116.

Fagiolo, G. and A. Roventini (2016). Macroeconomic policy in DSGE and agent-based models redux: New developments and challenges ahead.

Fankhauser, S. (2013). A practitioner's guide to a low-carbon economy: lessons from the uk. *Climate Policy 13*(3), 345–362.

Fankhauser, S., A. Bowen, R. Calel, A. Dechezleprêtre, D. Grover, J. Rydge, and M. Sato (2013). Who will win the green race? in search of environmental competitiveness and innovation. *Global Environmental Change 23*(5), 902–913.

Fankhauser, S. and F. Jotzo (2018). Economic growth and development with low-carbon energy. *Wiley Interdisciplinary Reviews: Climate Change 9*(1).

Fankhauser, S. and T. K. McDermott (2016). *The Economics of Climate-Resilient Development*. Edward Elgar Publishing.

Fankhauser, S. and N. Stern (2016). Climate change, development, poverty and economics.

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

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

Farmer, J. D. and F. Lafond (2016). How predictable is technological progress? *Research Policy 45*(3), 647–665.

Felipe, J., U. Kumar, A. Abdon, and M. Bacate (2012). Product complexity and economic development. *Structural Change and Economic Dynamics 23*(1), 36–68.

Felipe, J., A. Mehta, and C. Rhee (2017). *Manufacturing matters... but it's the jobs that count.* ADB Economics Working Paper Series, No. 420.

Fischer, C. and R. G. Newell (2008). Environmental and technology policies for climate mitigation. *Journal of Environmental Economics and Management 55*(2), 142–162.

Floater, G., P. Rode, A. Robert, C. Kennedy, D. Hoornweg, R. Slavcheva, and N. Godfrey (2014). *Cities and the New Climate Economy: the transformative role of global urban growth.* LSE Cities, London School of Economics and Political Science.

Foxon, T. J. (2011). A coevolutionary framework for analysing a transition to a sustainable low carbon economy. *Ecological Economics 70*(12), 2258–2267.

Fraccascia, L., I. Giannoccaro, and V. Albino (2018). Green product development: What does the country product space imply? *Journal of Cleaner Production 170*, 1076–1088.

Frenken, K., F. Van Oort, and T. Verburg (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies 41*(5), 685–697.

Frey, C. B. and M. A. Osborne (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change 114*, 254–280.

Fuller, B. and P. Romer (2014). *Urbanization as opportunity.* World Bank Policy

Research Working Paper: 6874.

Gabrielli, A., M. Cristelli, D. Mazzilli, A. Tacchella, A. Zaccaria, and L. Pietronero (2017, August). Why we like the ECI+ algorithm. *ArXiv e-prints*.

Garnaut, R., C. Fang, and L. Song (2013). China's new strategy for long-term growth and development. *China: A New Model for Growth and Development*, 1–16.

Gaspar, J. and E. L. Glaeser (1998). Information technology and the future of cities. *Journal of Urban Economics 43*(1), 136–156.

Geels, F. W., F. Berkhout, and D. P. van Vuuren (2016). Bridging analytical approaches for low-carbon transitions. *Nature Climate Change 6*(6), 576.

Gertler, M. S. (2005). Tacit knowledge, path dependency and local trajectories of growth. In *Rethinking Regional Innovation and Change*, pp. 23–41. Springer.

Gillingham, K., R. G. Newell, and K. Palmer (2009). Energy efficiency economics and policy. *Annu. Rev. Resour. Econ. 1*(1), 597–620.

Gintis, H. (2007). The dynamics of general equilibrium. *The Economic Journal 117*(523), 1280–1309.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of Political Economy 100*(6), 1126–1152.

Glaeser, E. L. and G. A. Ponzetto (2007). Did the death of distance hurt detroit and help new york? Technical report, National Bureau of Economic Research.

Glaeser, E. L., G. A. Ponzetto, and K. Tobio (2014). Cities, skills and regional change. *Regional Studies 48*(1), 7–43.

Glaeser, E. L. and A. Saiz (2003). The rise of the skilled city. Technical report, National Bureau of Economic Research.

Goos, M. and A. Manning (2007). Lousy and lovely jobs: The rising polarization

of work in britain. *The Review of Economics and Statistics 89*(1), 118–133.

Gowdy, J. M. (2008). Behavioral economics and climate change policy. *Journal of Economic Behavior & Organization 68*(3-4), 632–644.

Gray, W. B., R. J. Shadbegian, C. Wang, and M. Meral (2014). Do EPA regulations affect labor demand? evidence from the pulp and paper industry. *Journal of Environmental Economics and Management 68*(1), 188–202.

Green, F. and N. Stern (2017). China's changing economy: implications for its carbon dioxide emissions. *Climate policy 17*(4), 423–442.

Greenstone, M. (2002). The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. *Journal of Political Economy 110*(6), 1175–1219.

Griggs, D., M. Stafford-Smith, O. Gaffney, J. Rockström, M. C. Öhman, P. Shyamsundar, W. Steffen, G. Glaser, N. Kanie, and I. Noble (2013). Policy: Sustainable development goals for people and planet. *Nature 495*(7441), 305.

Grossman, G. M. and E. Helpman (1993). *Innovation and growth in the global economy*. MIT press.

Grossman, G. M. and E. Rossi-Hansberg (2008). Trading tasks: A simple theory of offshoring. *American Economic Review 98*(5), 1978–97.

Grubb, M., F. Sha, T. Spencer, N. Hughes, Z. Zhang, and P. Agnolucci (2015). A review of chinese co2 emission projections to 2030: the role of economic structure and policy. *Climate Policy 15*(sup1), S7–S39.

Guerrero, O. A. and R. L. Axtell (2013). Employment growth through labor flow networks. *PloS one 8*(5), e60808.

Guivarch, C. and S. Hallegatte (2011). Existing infrastructure and the 2 c target. *Climatic Change 109*(3-4), 801–805.

Hallegatte, S., M. Bangalore, M. Fay, T. Kane, and L. Bonzanigo (2015). *Shock waves: managing the impacts of climate change on poverty.* World Bank Publications.

Hallegatte, S., M. Fay, and A. Vogt-Schilb (2013). Green industrial policies: When and how.

Hallegatte, S., G. Heal, M. Fay, and D. Treguer (2012). From growth to green growth-a framework. Technical report, National Bureau of Economic Research.

Hamwey, R., H. Pacini, and L. Assunção (2013). Mapping green product spaces of nations. *The Journal of Environment & Development 22*(2), 155–168.

Hansen, A. H. (1939). Economic progress and declining population growth. *The American economic review 29*(1), 1–15.

Hansen, M. C., J. J. Norton, C. M. Gregory, A. W. Meade, L. F. Thompson, et al. (2014, 02). O*NET Work Activities Project. Technical report, National Center for O*NET Development.

Harberger, A. C. (1959). Using the resources at hand more effectively. *The American Economic Review 49*(2), 134–146.

Hasegawa, T., S. Fujimori, A. Ito, K. Takahashi, and T. Masui (2017). Global land-use allocation model linked to an integrated assessment model. *Science of The Total Environment 580*, 787–796.

Hasselmann, K. and D. V. Kovalevsky (2013). Simulating animal spirits in actor-based environmental models. *Environmental modelling & software 44*, 10–24.

Hausman, R., C. Hidalgo, et al. (2011). The atlas of economic complexity. *Mapping path to prosperity/R. Hausmann, CA Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, MA Yildirim//Puritan Press.–2011.–364 p.*

Hausmann, R. (2016). Economic development and the accumulation of know-how.

Welsh Economic Review 24, 13–16.

Hausmann, R., J. Chauvin, et al. (2015). Moving to the adjacent possible: Discovering paths for export diversification in rwanda. Technical report, Center for International Development at Harvard University.

Hausmann, R., C. Hidalgo, D. P. Stock, and M. A. Yildirim (2014). Implied comparative advantage.

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, A. Simoes, and M. A. Yildirim (2014). *The Atlas of Economic Complexity: Mapping paths to prosperity*. MIT Press.

Hausmann, R., J. Hwang, and D. Rodrik (2007). What you export matters. *Journal of Economic Growth 12*(1), 1–25.

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review 35*(4), 519–530.

Hémous, D. and M. Olsen (2016). The rise of the machines: Automation, horizontal innovation and income inequality.

Hepburn, C. and A. Bowen (2013). Prosperity with growth: Economic growth, climate change and environmental limits. In R. Fouquet (Ed.), *Handbook on Energy and Climate Change*, pp. 617–638. Cheltenham: Edward Elgar Publishing.

Hepburn, C., A. Pfeiffer, and A. Teytelboym (2018). Green growth. In G. Clark (Ed.), *The New Oxford Handbook of Economic Geography*, Chapter 40, pp. 749–769. Oxford University Press.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2014). Growth and structural transformation. In *Handbook of Economic Growth*, Volume 2, pp. 855–941. Elsevier.

Herzer, D. and F. Nowak-Lehnmann D (2006). What does export diversification do for growth? an econometric analysis. *Applied Economics 38*(15), 1825–1838.

Hesse, H. (2009). *Export diversification and economic growth*.

Hidalgo, C. A. and R. Hausmann (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences 106*(26), 10570–10575.

Hidalgo, C. A., B. Klinger, A. L. Barabási, and R. Hausmann (2007). The product space conditions the development of nations. *Science 317*(5837), 482–487.

Holland, J. H. and J. H. Miller (1991). Artificial adaptive agents in economic theory. *The American Economic Review 81*(2), 365–370.

Huberty, M. and G. Zachmann (2011). Green exports and the global product space: prospects for eu industrial policy. Technical report, Bruegel working paper.

Hughes, A., T. Caetano, and H. Trollip (2016). *Modelling a more equitable future*.

Imbs, J. and R. Wacziarg (2003). Stages of diversification. *The American Economic Review 93*(1), 63–86.

IPCC (2007). Climate Change 2007: Mitigation of Climate Change.Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.

IRENA (2016). *Renewable Energy Benefits: Measuring the Economics*. Abu Dhabi.

Jacobs, J. (1969). *The Life of Cities*. Random House.

Jacobs, M. (2013). Green growth. In R. Falkner (Ed.), *The Handbook of Global Climate and Environment Policy*, Chapter 12, pp. 197–214. John Wiley & Sons Ltd.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly journal*

of Economics 108(3), 577–598.

Jakob, M. and O. Edenhofer (2014). Green growth, degrowth, and the commons. *Oxford Review of Economic Policy 30*(3), 447–468.

Jensen, J. B. and L. G. Kletzer (2010). Measuring tradable services and the task content of offshorable services jobs. In *Labor in the new economy*, pp. 309–335. University of Chicago Press.

Jones, C. I. (2005). Growth and ideas. In *Handbook of Economic Growth*, Volume 1, pp. 1063–1111. Elsevier.

Kaika, D. and E. Zervas (2013). The environmental kuznets curve (ekc) theory—part a: Concept, causes and the co2 emissions case. *Energy Policy 62*, 1392–1402.

Keller, W. (2004). International technology diffusion. *Journal of Economic Literature 42*(3), 752–782.

Kemeny, T. and M. Storper (2015). Is specialization good for regional economic development? *Regional Studies 49*(6), 1003–1018.

Kemfert, C. (2003). Applied economic-environment-energy modeling for quantitative impact assessment. *Puzzle-solving for policy II 1*, 91.

Kemp-Benedict, E. (2014). An interpretation and critique of the method of reflections. Mimeo.

Kim, S. (1998). Economic integration and convergence: US regions, 1840–1987. *The Journal of Economic History 58*(3), 659–683.

Klein, A. (2013). New state-level estimates of personal income in the United States, 1880–1910. In *Research in Economic History*, pp. 191–255. Emerald Group Publishing Limited.

Knopf, B., O. Edenhofer, C. Flachsland, M. T. Kok, H. Lotze-Campen, G. Luderer,

A. Popp, and D. P. Van Vuuren (2010). Managing the low-carbon transition–from model results to policies. *The Energy Journal*, 223–245.

Kocornik-Mina, A. and S. Fankhauser (2015). Climate change adaptation in dynamic economies. *Centre for Climate Change Economics and Policy and Grantham Research Institute on Climate Change and the Environment*.

Kogler, D. F., J. Essletzbichler, and D. L. Rigby (2017). The evolution of specialization in the eu15 knowledge space. *Journal of Economic Geography 17*(2), 345–373.

Kogler, D. F., D. L. Rigby, and I. Tucker (2013). Mapping knowledge space and technological relatedness in US cities. *European Planning Studies 21*(9), 1374–1391.

Kogut, B. and U. Zander (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science 3*(3), 383–397.

Koh, H. and C. L. Magee (2008). A functional approach for studying technological progress: Extension to energy technology. *Technological Forecasting and Social Change 75*(6), 735–758.

Köhler, J., T. Barker, D. Anderson, and H. Pan (2006). Combining energy technology dynamics and macroeconometrics: the E3MG model. *The Energy Journal*, 113–133.

Krugman, P. (1991a). History and industry location: the case of the manufacturing belt. *The American Economic Review 81*(2), 80–83.

Krugman, P. (1991b). History versus expectations. *The Quarterly Journal of Economics 106*(2), 651–667.

Krugman, P. (1991c). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–499.

Krugman, P. R. (1991d). *Geography and trade*. MIT press.

Krugman, P. R. (1993). *Lessons of Massachusetts for EMU*. in Torres F., Giavazzi F. (eds) *Adjustment and growth in the European Monetary Union*, Cambridge University Press.

Kuznets, S. (1973). Modern economic growth: findings and reflections. *The American Economic Review 63*(3), 247–258.

Lall, S. (1992). Technological capabilities and industrialization. *World Development 20*(2), 165–186.

Lall, S. (2000). The technological structure and performance of developing country manufactured exports, 1985-98. *Oxford development studies 28*(3), 337–369.

Lall, S., J. Weiss, and J. Zhang (2006). The "sophistication" of exports: a new trade measure. *World Development 34*(2), 222–237.

Landefeld, J. and R. Marcus (2006). Gross domestic product by state estimation methodology, survey of current business. Technical report, Bureau of Economic Analysis. US Department of Commerce.

Lane-Visser, T. (2015). *State of the art review: development-first modelling frameworks and approaches*.

Leamer, E. E. and M. Storper (2014). The economic geography of the internet age. In *Location of International Business Activities*, pp. 63–93. Springer.

Lee, J.-S., T. Filatova, A. Ligmann-Zielinska, B. Hassani-Mahmooei, F. Stonedahl, I. Lorscheid, A. Voinov, J. G. Polhill, Z. Sun, and D. C. Parker (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation 18*(4), 4.

Lee, K. and S. Ni (2002). On the dynamic effects of oil price shocks: a study using industry level data. *Journal of Monetary Economics 49*(4), 823–852.

Lenzen, M. (2016). Structural analyses of energy use and carbon emissions–an overview. *Economic Systems Research 28*(2), 119–132.

Leten, B., R. Belderbos, and B. Van Looy (2007). Technological diversification, coherence, and performance of firms. *Journal of Product Innovation Management 24*(6), 567–579.

Levin, T. and V. M. Thomas (2016). Can developing countries leapfrog the centralized electrification paradigm? *Energy for Sustainable Development 31*, 97–107.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School 22*(2), 139–191.

Lin, J. (2011). Technological adaptation, cities, and new work. *The Review of Economics and Statistics 93*(2), 554–574.

Liu, M., R. Shadbegian, and B. Zhang (2017). Does environmental regulation affect labor demand in China? evidence from the textile printing and dyeing industry. *Journal of Environmental Economics and Management 86*, 277–294.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3–42.

Lutz, C., B. Meyer, C. Nathani, and J. Schleich (2007). Endogenous innovation, the economy and the environment: Impacts of a technology-based modelling approach for energy-intensive industries in germany. *Energy Studies Review 15*(1).

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics 107*(2), 407–437.

Mariani, M. S., A. Vidmer, M. Medo, and Y.-C. Zhang (2015). Measuring economic complexity of countries and products: which metric to use? *The European Physical Journal B 88*(11), 293.

Marshall, A. (1890). *Principles of economics*. Macmillan And Co., Limited; Lon-

don.

Martin, R., P. Sunley, B. Gardiner, E. Evenhuis, and P. Tyler (2018). The city dimension of the productivity growth puzzle: the relative role of structural change and within-sector slowdown. *Journal of Economic Geography*.

Mattauch, L. and C. Hepburn (2016). Climate policy when preferences are endogenous—and sometimes they are. *Midwest Studies in Philosophy 40*(1), 76–95.

McKinsey (2015). Four fundamentals of workplace automation. *McKinsey Quarterly, November 2015*.

Mealy, P., R. M. del Rio-Chanona, and J. D. Farmer (2018). What you do at work matters: New lenses on labour. *INET Working Paper*.

Mealy, P., J. D. Farmer, and A. Teytelboym (2017). A new interpretation of the economic complexity index. *INET Working Paper*.

Mealy, P., A. Pfeiffer, C. Hepburn, and J. Aldy (2018). Navigating the energy transition: The shifting and stranding of jobs in the us power sector. *Mimeo*.

Mealy, P. and A. Teytelboym (2017). Economic complexity and the green economy. *INET Working Paper*.

Menanteau, P., D. Finon, and M.-L. Lamy (2003). Prices versus quantities: choosing policies for promoting the development of renewable energy. *Energy policy 31*(8), 799–812.

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

Mersmann, F., K. H. Olsen, T. Wehnert, and Z. Boodoo (2014). From theory to practice: Understanding transformational change in NAMAs. *UNEP DTU Partnership*.

Metcalfe, J. S. (1998). *Evolutionary economics and creative destruction*, Volume 1. Psychology Press.

Metcalfe, S. et al. (2018). *Evolutionary theories of economic and technological change: present status and future prospects*, Volume 44. Routledge.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics 90*(4), 683–701.

Michaels, G., A. Natraj, and J. Van Reenen (2014). Has ict polarized skill demand? evidence from eleven countries over twenty-five years. *The Review of Economics and Statistics 96*(1), 60–77.

Michaels, G., F. Rauch, and S. J. Redding (2013). Task specialization in US cities from 1880-2000. *National Bureau of Economic Research*.

Mlachila, M. M., R. Tapsoba, and M. S. J.-A. Tapsoba (2014). *A quality of growth index for developing countries: A proposal*. Number 14-172. International Monetary Fund.

Mokyr, J. (2002). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Morgenstern, R. D., W. A. Pizer, and J.-S. Shih (2002). Jobs versus the environment: an industry-level perspective. *Journal of Environmental Economics and Management 43*(3), 412–436.

Morrison, G., S. V. Buldyrev, M. Imbruno, O. A. Doria Arrieta, A. Rungi, M. Riccaboni, and F. Pammolli (2017). On economic complexity and the fitness of nations. *Scientific Reports 7*(1), 15332.

Müller, B., F. Bohn, G. Dreßler, J. Groeneveld, C. Klassert, R. Martin,

M. Schlüter, J. Schulze, H. Weise, and N. Schwarz (2013). Describing human decisions in agent-based models–odd+ d, an extension of the odd protocol. *Environmental Modelling & Software 48*, 37–48.

Muneepeerakul, R., J. Lobo, S. T. Shutters, A. Goméz-Liévano, and M. R. Qubbaj (2013). Urban economies and occupation space: can they get "there" from "here"? *PloS one 8*(9), e73676.

Murphy, K. M., A. Shleifer, and R. W. Vishny (1989). Industrialization and the big push. *Journal of Political Economy 97*(5), 1003–1026.

NCE (2014). *Better growth, better climate: charting a new path for low-carbon growth and a safer climate*. New Climate Economy Report, London.

NCE (2016). *The sustainable infrastructure imperative*. New Climate Economy Report, London.

Neffke, F. (2017). *Coworker complementarity*. Science Policy Research Unit Working Paper Series.

Neffke, F., M. Hartog, R. Boschma, and M. Henning (2018). Agents of structural change: the role of firms and entrepreneurs in regional diversification. *Economic Geography 94*(1), 23–48.

Neffke, F. and M. Henning (2013). Skill relatedness and firm diversification. *Strategic Management Journal 34*(3), 297–316.

Neffke, F., M. Henning, and R. Boschma (2011a). How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic Geography 87*(3), 237–265.

Neffke, F., M. Henning, R. Boschma, K.-J. Lundquist, and L.-O. Olander (2011). The dynamics of agglomeration externalities along the life cycle of industries. *Regional Studies 45*(1), 49–65.

Neffke, F. and M. S. Henning (2008). Relatedness, revealed and space, mapping industry. Papers in Evolutionary Economic Geography 08.19, Utrecht University.

Neffke, F. M., M. Henning, and R. Boschma (2011b). The impact of aging and technological relatedness on agglomeration externalities: a survival analysis. *Journal of Economic Geography 12*(2), 485–517.

Neffke, F. M., A. Otto, and A. Weyh (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization 142*, 275–292.

Nelson, R. R. and S. G. Winter (1982). *An evolutionary theory of economic change*. Harvard University Press, Cambridge, MA.

Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

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.

Nordhaus, W. D. and J. Tobin (1972). Is growth obsolete? In *Economic Research: Retrospect and prospect, Volume 5, Economic growth*, pp. 1–80. NBER.

O'Clery, N., A. Gomez-Lievano, and E. Lora (2016). The path to labor formality: Urban agglomeration and the emergence of complex industries. Working Paper 78, Center for International Development at Harvard University.

OECD (1999, March). Future liberalisation of trade in environmental goods and services: Ensuring environmental protection as well as economic benefits. Joint Working Party on Trade and Environment COM/TD/ENV(98)37/FINAL, Organization for Economic Cooperation and Development.

OECD (2017, March). Trade in services related to the environment. Joint Working Party on Trade and Environment COM/TAD/ENV/JWPTE(2015)61/FINAL,

Organization for Economic Cooperation and Development.

Ota, T. (2017). Economic growth, income inequality and environment: assessing the applicability of the kuznets hypotheses to asia. *Palgrave Communications 3*, 17069.

Page, J. (2015). Manufacturing, natural resources, and industrialization. *The Oxford Handbook of Africa and Economics: Volume 2: Policies and Practices*.

Patel, P. and K. Pavitt (1997). The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research Policy 26*(2), 141–156.

Penrose, E. (1959). The theory of the firm. *NY: John Wiley & Sons*.

Perez, C. (2015). Capitalism, technology and a green global golden age: The role of history in helping to shape the future. *The Political Quarterly 86*(S1), 191–217.

Peri, G. and C. Sparber (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics 1*(3), 135–69.

Pfeiffer, A., R. Millar, C. Hepburn, and E. Beinhocker (2016). The '2 c capital stock'for electricity generation: Committed cumulative carbon emissions from the electricity generation sector and the transition to a green economy. *Applied Energy 179*, 1395–1408.

Polanyi, M. (1967). The tacit dimension. 1967. *Garden City, NY: Anchor*.

Poncet, S. and F. S. de Waldemar (2013). Export upgrading and growth: the prerequisite of domestic embeddedness. *World Development 51*, 104–118.

Porter, M. (2003). The economic performance of regions. *Regional Studies 37*(6), 549–578.

Pratt, S., A. Blake, and P. Swann (2013). Dynamic general equilibrium model with uncertainty: Uncertainty regarding the future path of the economy. *Economic*

Modelling 32, 429–439.

Preston, A. and G. Whitehouse (204). Gender differences in occupation of employment within australia. *Australian Journal of Labour Economics 7*(3), 309–327.

Quigley, J. M. (1998). Urban diversity and economic growth. *Journal of Economic Perspectives 12*(2), 127–138.

Radebach, A., J. C. Steckel, and H. Ward (2016). Patterns of sectoral structural change–empirical evidence from similarity networks.

Raupach, M. R., G. Marland, P. Ciais, C. Le Quéré, J. G. Canadell, G. Klepper, and C. B. Field (2007). Global and regional drivers of accelerating co2 emissions. *Proceedings of the National Academy of Sciences 104*(24), 10288–10293.

Reid, W. V., D. Chen, L. Goldfarb, H. Hackmann, Y. T. Lee, K. Mokhele, E. Ostrom, K. Raivio, J. Rockström, H. J. Schellnhuber, et al. (2010). Earth system science for global sustainability: grand challenges. *Science 330*(6006), 916–917.

Rezai, A., L. Taylor, and R. Mechler (2013). Ecological macroeconomics: An application to climate change. *Ecological Economics 85*, 69–76.

Richiardi, M. (2015). The future of agent-based modelling. *INET Working Paper*.

Richiardi, M., R. Leombruni, N. Saam, and M. Sonnessa (2006). A common protocol for agent-based social simulation.

Rigby, D. L. (2015). Technological relatedness and knowledge space: entry and exit of US cities from patent classes. *Regional Studies 49*(11), 1922–1937.

Rodrik, D. (2012). Unconditional convergence in manufacturing. *The Quarterly Journal of Economics 128*(1), 165–204.

Rodrik, D. (2013). Structural change, fundamentals, and growth: an overview. *Institute for Advanced Study*.

Rodrik, D. (2014). Green industrial policy. *Oxford Review of Economic Pol-*

icy 30(3), 469–491.

Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth 21*(1), 1–33.

Rogelj, J., M. Den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, and M. Meinshausen (2016). Paris Agreement climate proposals need a boost to keep warming well below 2 c. *Nature 534*(7609), 631.

Romer, P. (1993). Idea gaps and object gaps in economic development. *Journal of Monetary Economics 32*(3), 543–573.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy 94*(5), 1002–1037.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Rondander, C. (1999). The classification of work: Applying 1950 census occupation and industry codes to 1920 responses. *Historical Methods 32*(3), 151–155.

Rosa, E. A. and T. Dietz (2012). Human drivers of national greenhouse-gas emissions. *Nature Climate Change 2*(8), 581.

Rossi-Hansberg, E., P.-D. Sarte, and R. Owens Iii (2009). Firm fragmentation and urban patterns. *International Economic Review 50*(1), 143–186.

Ruggles, S., K. Grenadek, R. Goeken, J. Grover, and M. Sobek (2017). Integrated public use microdata series: Version 7.0.

Russill, C. and Z. Nyssa (2009). The tipping point trend in climate change communication. *Global Environmental Change 19*(3), 336–344.

Sachs, J. D. (2012). From millennium development goals to sustainable development goals. *The Lancet 379*(9832), 2206–2211.

Sala-i-Martin, X. X. (1997). I just ran two million regressions. *The American*

Economic Review 87(2), 178–183.

Saracco, F., R. Di Clemente, A. Gabrielli, and T. Squartini (2015). Randomizing bipartite networks: the case of the world trade web. *Scientific Reports 5*, 10595.

Sauvage, J. (2014). The stringency of environmental regulations and trade in environmental goods. OECD Trade and Environment Working Paper 2014/03, Organization for Economic Cooperation and Development, Annex 1.

Schilling, M. A. and M. Esmundo (2009). Technology s-curves in renewable energy alternatives: Analysis and implications for industry and government. *Energy Policy 37*(5), 1767–1781.

Schumpeter, J. A. (1994). *Capitalism, Socialism and Democracy*. Routledge London and New York.

Scrieciu, S., A. Rezai, and R. Mechler (2013). On the economic foundations of green growth discourses: the case of climate change mitigation and macroeconomic dynamics in economic modeling. *Wiley Interdisciplinary Reviews: Energy and Environment 2*(3), 251–268.

Scrieciu, S. S. (2007). The inherent dangers of using computable general equilibrium models as a single integrated modelling framework for sustainability impact assessment. a critical note on böhringer and löschel (2006). *Ecological Economics 60*(4), 678–684.

Scrieciu, S. Ş., V. Belton, Z. Chalabi, R. Mechler, and D. Puig (2014). Advancing methodological thinking and practice for development-compatible climate policy planning. *Mitigation and Adaptation Strategies for Global Change 19*(3), 261–288.

Selden, T. M. and D. Song (1994). Environmental quality and development: is there a kuznets curve for air pollution emissions? *Journal of Environmental*

Economics and Management 27(2), 147–162.

Seron, C., S. S. Silbey, E. Cech, and B. Rubineau (2016). Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations 43*(2), 178–214.

Shi, J. and J. Malik (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence 22*(8), 888–905.

Shutters, S. T., R. Muneepeerakul, and J. Lobo (2016a). Constrained pathways to a creative urban economy. *Urban Studies 53*(16), 3439–3454.

Shutters, S. T., R. Muneepeerakul, and J. Lobo (2016b). How hard is it for urban economies to become 'green'? *Environment and Planning B: Planning and Design 43*(1), 198–209.

Silva, E. G. and A. A. Teixeira (2008). Surveying structural change: Seminal contributions and a bibliometric account. *Structural Change and Economic Dynamics 19*(4), 273–300.

Silverberg, G., G. Dosi, and L. Orsenigo (1988). Innovation, diversity and diffusion: a self-organisation model. *The Economic Journal 98*(393), 1032–1054.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization 1*(1), 161–176.

Smajgl, A., D. G. Brown, D. Valbuena, and M. G. Huigen (2011). Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling & Software 26*(7), 837–844.

Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A bayesian DSGE approach. *American economic review 97*(3), 586–606.

Smith, A. (1776). The wealth of nations. *Oxford University Press, Oxford*.

Smulders, S., M. Toman, and C. Withagen (2014). Growth theory and 'green

growth'. *Oxford Review of Economic Policy 30*(3), 423–446.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics 70*(1), 65–94.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 312–320.

Spence, M. and S. Hlatshwayo (2012). The evolving structure of the American economy and the employment challenge. *Comparative Economic Studies 54*(4), 703–738.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics 24*(2), 235–270.

Steg, L. (2008). Promoting household energy conservation. *Energy policy 36*(12), 4449–4453.

Stern, D. I. and M. S. Common (2001). Is there an environmental kuznets curve for sulfur? *Journal of Environmental Economics and Management 41*(2), 162–178.

Stern, D. I., M. S. Common, and E. B. Barbier (1996). Economic growth and environmental degradation: the environmental kuznets curve and sustainable development. *World Development 24*(7), 1151–1160.

Stern, N. (2008). The economics of climate change. *American Economic Review 98*(2), 1–37.

Stern, N. (2016). Economics: current climate models are grossly misleading. *Nature Climate Change 530*(7591), 407.

Stojkoski, V., Z. Utkovski, and L. Kocarev (2016, 08). The impact of services on economic complexity: Service sophistication as route for economic growth. *PLOS ONE 11*(8), 1–29.

Storper, M. (2018). Separate worlds? explaining the current wave of regional economic polarization. *Journal of Economic Geography 18*(2), 247–270.

Straka, M. J., G. Caldarelli, and F. Saracco (2017). Grand canonical validation of the bipartite international trade network. *Physical Review E 96*(2), 022306.

Su, B. and B. W. Ang (2012). Structural decomposition analysis applied to energy and emissions: some methodological developments. *Energy Economics 34*(1), 177–188.

Sutton, J. and D. Trefler (2016). Capabilities, wealth, and trade. *Journal of Political Economy 124*(3), 826–878.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports 2*, 723.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2013). Economic complexity: conceptual grounding of a new metrics for global competitiveness. *Journal of Economic Dynamics and Control 37*(8), 1683–1691.

Taylor, L. (2009). Energy productivity, labor productivity, and global warming. *Twenty-First Century Macroeconomics: Responding to the Climate Challenge, Northhampton MA: Edward Elgar*.

Teece, D. and G. Pisano (1994). The dynamic capabilities of firms: an introduction. *Industrial and corporate change 3*(3), 537–556.

Teece, D. J., G. Pisano, and A. Shuen (1997). Dynamic capabilities and strategic management. *Strategic Management Journal 18*(7), 509–533.

Thissen, M., F. Van Oort, D. Diodato, and A. Ruijs (2013). *Regional competitiveness and smart specialization in Europe: Place-based development in international economic networks*. Edward Elgar Publishing.

Trancik, J. E. (2014). Renewable energy: Back the renewables boom. *Nature 507*(7492), 300–302.

Turnheim, B., F. Berkhout, F. Geels, A. Hof, A. McMeekin, B. Nykvist, and D. van Vuuren (2015). Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. *Global Environmental Change 35*, 239–253.

UN (2014). *World Urbanization Prospects: The 2014 Revision-Highlights*. United Nations, Department of Economic and Social Affairs, Population Division.

UN (2015). Transforming Our World: the 2030 Agenda for Sustainable Development: Resolution adopted by the General Assembly on 25 September 2015. *A/RES/70/1*.

UNEP (2011). Towards a green economy: Pathways to sustainable development and poverty eradication. Technical report, United Nations Environment Programme.

UNFCCC (2015). Adoption of the Paris Agreement. *Report No. FCCC/CP/2015/L.9/ Rev.1*.

UNIDO (2016). *Industrial Development Report 2016. The Role of Technology and Innovation in Inclusive and Sustainable Industrial Development*. Vienna.

US Department of Commerce (1962). Personal income by states, a supplement to the survey of current business and survey of current business. Technical report, Office of Business Economics.

Valbuena, D., P. H. Verburg, and A. K. Bregt (2008). A method to define a typology for agent-based analysis in regional land-use research. *Agriculture, Ecosystems & Environment 128*(1-2), 27–36.

Van Dijk, M., G. Philippidis, and G. Woltjer (2016). Catching up with history: A

methodology to validate global CGE models.

Veblen, T. (1898). Why is economics not an evolutionary science? *The Quarterly Journal of Economics 12*(4), 373–397.

Verspagen, B. and M. Kaltenberg (2015). Catching-up in a globalised context: Technological change as a driver of growth. Working Paper Series 49, United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).

Vespignani, A. (2012). Modelling dynamical processes in complex socio-technical systems. *Nature physics 8*(1), 32.

Vivid Economics (2018). *Liberalisation of trade in environmental goods and services: Opportunities for EBRD countries of operation*. Report prepared for the European Bank for Reconstruction and Development, February 2018.

Vona, F., G. Marin, D. Consoli, and D. Popp (2017). Environmental regulation and green skills: an empirical exploration. *Journal of the Association of Environmental and Reseource Economists forthcoming*.

Walker, W. R. (2011). Environmental regulation and labor reallocation: Evidence from the clean air act. *American Economic Review 101*(3), 442–47.

Watts, D. J. and P. S. Dodds (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research 34*(4), 441–458.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics 113*(2), 331–360.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal 5*(2), 171–180.

West, G. R. (1995). Comparison of input–output, input–output+ econometric and computable general equilibrium impact models at the regional level. *Economic*

Systems Research 7(2), 209–227.

Wiebe, K. S. and C. Lutz (2016). Endogenous technological change and the policy mix in renewable power generation. *Renewable and Sustainable Energy Reviews 60*, 739–751.

Wing, I. S. (2004). Computable general equilibrium models and their use in economy-wide policy analysis. *Technical Note, Joint Program on the Science and Policy of Global Change, MIT*.

Wing, I. S. (2009). Computable general equilibrium models for the analysis of energy and climate policies. In *International Handbook on the Economics of Energy*. Edward Elgar Publishing.

Wing, I. S. and E. J. Balistreri (2014). Computable general equilibrium models for policy evaluation and economic consequence analysis.

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

World Bank (2012). Inclusive green growth: The pathway to sustainable development. Technical report, World Bank, Washington, DC: World Bank Publications.

WTO (2001). The Doha Mandate. Technical report, World Trade Organisation.

WTO (2010, March). Report by the Chairman to the Trade Negotiations Committee on the Committee and Trade and Environment in Special Session. Technical Report TN/TE/19, World Trade Organisation.

WTO (2011, April). Report by the Chairman to the Trade Negotiations Committee on the Committee and Trade and Environment in Special Session. Technical Report TN/TE/19, World Trade Organisation.

Yao, C., K. Feng, and K. Hubacek (2015). Driving forces of co2 emissions in the g20 countries: An index decomposition analysis from 1971 to 2010. *Ecological informatics 26*, 93–100.

Youn, H., L. M. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, and G. B. West (2016). Scaling and universality in urban economic diversification. *Journal of The Royal Society Interface 13*(114), 1–7.

Zaccaria, A., M. Cristelli, A. Tacchella, and L. Pietronero (2014). How the taxonomy of products drives the economic development of countries. *PloS one 9*(12), e113770.

Zenghelis, D. (2015). Decarbonisation: Innovation and the economics of climate change. *The Political Quarterly 86*(S1), 172–190.