Input-Output Analysis and Growth Theory

Charles Savoie

Lincoln College

University of Oxford

A Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy in Mathematics

Trinity 2017

Acknowledgements

First, I am really grateful to my parents for their love and support during my childhood up to the most recent years during my PhD. You teach me *Labor Omnia Vincit* and it remains a leitmotiv. I would like to thank my brother Antoine and my sister Hélène-Astrid for their love.
I would like to thank my supervisor Doyne Farmer for his guidance and trust through these three years of doctorate. I have learnt a lot working with you and became a better scientist. I am thankful to Sam Howison and Sam Johnson for their useful feedbacks during the different stages of this thesis.
I also acknowledge INET, the Physics Department of the École Normale Supérieure of Lyon, and my parents for their generous financial support along my studies.
I am thankful to my co-authors James McNerney, Francesco Caravelli, and Daniel Fricke who have contributed to the work presented in this thesis. I have also benefited from inspiring conversations on my research with colleagues and friends Penny Mealy, François Lafond, Dan Tang, David Pugh, Christoph Aymanns, Marco Pangallo and to the rest of the INET group.
I would also like to thank my friends in Oxford François Lavergne, Adrien Hitz, Déborah Thebault, and all the Lincoln gang. Thank you to all the folks from the Lincoln College Boat Club for these amazing three years of rowing. A special thank you to Benjamin de Thoré and to my other friends from France for their patience and support.

Abstract

This thesis studies a theory for the amplification of technological improvement by the production network structure of the economy.

The theory is motivated by the idea that, to the extent that inputs and outputs of industries form a chain, improvements are passed down the chain and accumulate multiplicatively. Under a simple model for technological improvement it is possible to compute the overall improvement factor for the general case where the production network has a complicated structure containing cycles. We call this the trophic depth by analogy with ecology.

This leads to testable predictions about GDP growth and its variance. We analyse data for 40 countries and 35 industries from 1995 to 2009 and demonstrate that trophic depths are strongly correlated with economic growth. A regression of GDP growth of countries against their trophic depths has a highly statistically significant R-squared equal to 0.38, and when other standard explanatory variables are added to the regression, the trophic depth remains a robust and statistically significant contributor.

We perform some statistical analysis to understand the evolution of trophic depths at different stage of the economic development. We identify two growth paths. Along the first growth path, countries are catching up frontier economies while along the second growth path countries are falling behind. This approach allows us to make some forecasts about the evolution of trophic depths and of the wealth of countries. This provides a comprehensive framework to understand the acceleration and deceleration of economic growth.

Then, we study another type of technological progress that corresponds to the adoption of new goods in the production chain. This mechanism is related to the dynamic of the production network and for this purpose we perform a link prediction analysis to determine some key factors for new adoptions.

Finally, we analyse the relation between stock return comovement and institutional preferences across stocks of various size. A growing literature highlights the role of investors' common asset holdings on market dynamics. While previous studies focused on large stocks we also include small stocks in the sample in order to acknowledge the shift in institutional preferences towards small stocks over the last decades. Moreover, we add the input-output linkages between firms from different industries to our set of explanatory variables.

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.
I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Economic growth is a recent phenomenon and varies largely across regions. Since the industrial revolution, economic growth has brought material well-being and people are living longer and in better health. Beyond this global statement some regions are still in extreme poverty while others are wealthy. To understand why in average the world becomes wealthier and why this wealth is inequitably distributed we need to understand what drives economic growth.

Technological change was first stressed as a key factor for long-run growth in the work of Solow (1956). This breakthrough contrasted with other attempts that instead considered capital accumulation as a key ingredient for economic growth. While it was first determined exogenously, a second generation of models endogenized technological progress to make macroeconomic models build on microeconomic foundations. In all those approaches, growth theory assumes an aggregate production function. The aggregation process makes the bold assumption there is no heterogeneity across sectors. For a first approximation, it can be satisfying but it cannot be true that agriculture, manufacturing, and services have the same properties when studying growth theory. When building more realistic models, preserving the heterogeneity across sectors is crucial as the dynamic is likely to be richer.

Once heterogeneity matters, disaggregation becomes a necessity. This requires the use of networks to map the interaction between these entities. In this thesis, the economy is described as a network in which nodes are producers that purchase goods, convert them into different goods, and sell them to households or other producers. Links in the network can either be goods flows or money flows. A typical good is a composition of many other goods, each of which is in turn composed of many other goods. This makes the economy a highly distributed process with a complicated network structure. Technological shocks will propagate and get amplified by the network. However, different networks will have different responses to shocks, thus we need growth theory to account for network effects.

In Chapter 2 I provide a literature review on growth theory. I will start by describing the Malthusian view to highlight the importance of technological change in long-run growth theory introduced by Solow. Then, I discuss the emergence of endogenous growth theory that will cover Romer's model and the Schumpeterian model. I also review recent works that have been done on network economics.

In Chapter 3 I describe a theory for the amplification of technological improvement by the production network structure of the economy. This chapter and the following one are the result of a collaboration with J. McNerney, F. Caravelli, and J. D. Farmer. Using a simple model for technological improvement it is possible to compute the overall improvement factor for the general case where the production network has a complicated structure containing cycles. A generalized output multiplier quantifies this amplification and is equal to the ratio of gross output to net output, which for a closed economy is invariant under aggregation. This leads to testable predictions about GDP growth and its variance.

In Chapter 4 I analyze data for 40 countries and 35 industries from 1995 to 2009 from the World Input-Output Database and demonstrate that output multipliers are

strongly correlated with economic growth. A regression of GDP growth of countries against their output multipliers has a highly statistically significant R-squared equal to 0.38, and when other standard explanatory variables are added to the regression, the output multiplier remains a robust and statistically significant contributor. We show that output multipliers of a country are determined by a combination of labour share and industry composition.

In Chapter 5 I analyse the evolution of the trophic depth at different stages of development. In the previous chapters, the input-output network was given and assumed constant. In fact, its dynamic is more complex as both the weights and the links change over time. Using different snapshots of input-output networks and a model selection algorithm, I show that there exist two growth paths. One allows countries to catch up the most wealthy economies while along the second growth path countries are falling behind. This method allows us to make some long-run growth forecasts.

In Chapter 6 I relax the assumption that the input-output network is static and I focus on one aspect of its dynamics: the linkage dynamic. Link prediction is an important part of the network dynamic as it captures the evolution of technologies. I use input-output networks provided by the Bureau of Economic Analysis for the USA over 25 years to highlight the appropriate parameters. This work is based on a working paper by Carvalho and Voigtländer (2014). My contribution consists in adding additional parameters that turn out to be more significant than the ones previously introduced by Carvalho and Voigtländer.

In Chapter 7 I analyse the relation between stock return comovement and institutional preferences across stocks of various size. A growing literature highlights the role of investors' common asset holdings on market dynamics. For example, focusing on the subset of large stocks, Antón and Polk (2014) showed that stocks sharing many

common investors tend to co-move more strongly with each other in the future than otherwise similar stocks. This chapter is independent of the previous chapters and is the result of a collaboration with D. Fricke; nonetheless, we use the input-output network to capture the complex interaction between firms. We perform a similar analysis as in Antón and Polk (2014), but also include small stocks in the sample in order to acknowledge the shift in institutional preferences towards small stocks over the last decades. First, we confirm that common ownership is significantly related to future return correlations for relatively large stocks, but the strength of the relationship tends to decrease over time. Moreover, we add the input-output linkages between firms from different industries to our set of explanatory variables. This effect is largely significant, robust over time and it does not depend of the size of the stocks. For relatively small stocks we find that the relationship is often insignificant and we find that institutional ownership does not Granger-cause return correlations for the subset of small stocks.

Finally, in Chapter 8 I summarize the work of my thesis and I provide some ideas on how to use my results in future research.

Chapter 2

Literature review

First, I review the evolution of growth theory from the Malthusian growth model to the Schumpeterian model. In the second part, I present the recent focus on network properties of the economy and its ability to amplify shocks.

2.1 Economic growth theory

The first economic growth model is the Malthusian model presented by Malthus (Malthus, 1798). The population grows exponentially at a constant rate such that it doubles every 25 years but in the meantime food production grows arithmetically. It leads to booms and clashes when the population over-reaches the land capacity. The limited factors are the food supply and the extraction of other resources. Under this model, the economy is stagnant and economic growth equals population growth.

It fails to explain the shift from the Malthusian trap into sustained economic growth. Recent growth modelling started with Neumann (1945) and neoclassical models emerged later in the 50's and 60's. Solow (1956) and Swan (1956) independently derived an exogenous growth model with a balanced growth path. The Solow-Swan model is an extension of the Harrod-Domar model (Harrod, 1939; Domar, 1946) with an additional term capturing productivity growth. Given the importance of this

model in the economic growth theory, we re-derive the equations that lead to the growth rate of the economy.

The aggregate production function of the economy is

$$Y(t) = F[K(t), L(t), A(t)], \tag{2.1}$$

where Y is the total output of the economy, K is the stock of capital, L is the stock of labour, and A is the stock of knowledge. To guarantee the stability of the solution F has to satisfy the Inada conditions (Inada, 1963).[1] Differentiating the aggregate production function with respect to time leads to

$$\frac{\dot{Y}}{Y} = \frac{F_A A}{Y}\frac{\dot{A}}{A} + \frac{F_K K}{Y}\frac{\dot{K}}{K} + \frac{F_L L}{Y}\frac{\dot{L}}{L}, \tag{2.2}$$

where F_i is the first derivative of F with respect to the variable i. The contribution of technology to growth, the Total Factor Productivity (TFP), is defined as

$$TFP = \frac{F_A A}{Y}\frac{\dot{A}}{A}. \tag{2.3}$$

The growth rates of output, capital stock, and labour are $g = \dot{Y}/Y$, $g_K = \dot{K}/K$, and $h = \dot{L}/L$ respectively. Defining the share of capital as $\alpha_K = \frac{F_K K}{Y}$ and the share of labour as $\alpha_L = \frac{F_L L}{Y}$, Eq. 2.2 is rewritten

$$TFP = g - \alpha_K g_K - \alpha_L h. \tag{2.4}$$

We now consider a Cobb-Douglas production function (Cobb and Douglas, 1928)

$$Y = (AL)^{1-\alpha} K^{\alpha}. \tag{2.5}$$

[1] There are 6 conditions: F is 0 at 0, F is strictly increasing, F is concave, the limit of the first partial derivatives is positive infinity as the variable approaches 0, and the limit of the first partial derivatives is zero as the variable approaches positive infinity.

In this form, technological progress is the same as an improvement of the "effective" supply of labour. When previous equation is renormalized using the output per person $y = Y/L$ and the stock of capital per efficiency unit of labour $\kappa = K/(AL)$, it becomes

$$y = A\kappa^{\alpha}. \tag{2.6}$$

The net investment, which is the net rate of increase of the capital stock per unit of time, is

$$I = sY - \delta K, \tag{2.7}$$

where s is the saving rate and δ is the depreciation rate. By definition, the net investment is equal to $\dot{K}$, thus

$$\dot{K} = sY - \delta K. \tag{2.8}$$

So the rate of increase of the stock of capital per efficiency unit of labour is

$$\frac{\dot{\kappa}}{\kappa} = s\kappa^{\alpha-1} - (h + a + \delta), \tag{2.9}$$

where a is the growth rate of technological progress. In the long run, under the assumption of constant returns to scale, the economy will approach a steady-state level of output per capita, denoted $y^{\star}$, defined by

$$y^{\star} = A\kappa^{\star\alpha} \tag{2.10}$$

where $\kappa^{\star}$ is the stock of capital per efficiency unit of labour at the steady-state. A first order Taylor expansion of $\log(y)$ with respect to $\log(\kappa)$ around $\log(\kappa^{\star})$ leads to

$$\log(y) - \log(y^{\star}) \simeq \alpha(\log(\kappa) - \log(\kappa^{\star})). \tag{2.11}$$

Differentiating Eq. 2.6 with respect to time leads to

$$g - h \simeq a - (1 - \alpha)(\delta + a + h)(\log(y) - \log(y^\star)). \tag{2.12}$$

By definition Eq. 2.9 is equal to 0 at the steady state level and

$$\kappa^{\star\alpha} = \frac{\delta + a + h}{s}\kappa^\star. \tag{2.13}$$

Finally, the growth rate of the output per capita is given by

$$g - h \simeq a - \alpha_L(\delta + a + h)\left(\log(y) - \log(A) - \log\left(\kappa^\star \frac{\delta + a + h}{s}\right)\right), \tag{2.14}$$

where $y^\star$ is replaced by its expression and $\alpha_L = 1 - \alpha$ is the labour share. In a Solow-Swan model regardless the starting point, the economy will converge toward a balanced growth path and then the growth rate of output per worker is only determined by the rate of technological change. While the rate of technological change is a core component of the Solow-Swan model, it is an exogenous parameter.

This is not satisfactory because it does not relate the technological change to an economic mechanism. Frankel (1962) developed the first endogenous growth model. It is an AK type model meaning that there are no diminishing returns to capital but instead there is a linear relation between the gross output and the stock of capital. This approach is a direct consequence of the work of Arrow on learning by doing that compensates the diminishing returns of capital (Arrow, 1962). In an AK model, no distinction is made between capital accumulation and technological progress but instead the new exogenous variables are the saving rate and the level of technology.

Lucas (1988) differentiates two kinds of capital. Physical capital refers to technology and human capital is associated to labour productivity. Lucas' model is focused on the accumulation of knowledge at the individual level. Hence, he introduced the

concepts of learning-by-doing and on-the-job training as mechanisms for human capital accumulation. Moreover, the educational system has a central role in raising human capital. It creates an opportunity to increase the skills of agents and has consequences for productivity improvements. From the point of view of Lucas, accumulated knowledge is a long and difficult process.

Another approach considers knowledge as easy to learn and leads to technological spillovers and competition. Romer (1990) views technology as a non-rival and partially excludable good. In his model, he adds a third sector to the existing intermediate-goods sector and final-goods sector which is the research sector. It uses labour and existing knowledge to produce new knowledge. In Romer's model, the stock of knowledge evolves proportionally with the number of people employed in the research sector. Assuming that the share of people involved in research remains constant, the model explains super-exponential economic growth with the super-exponential growth of the population; however, a large population is not a sufficient condition for a country to have a long-term growth.

A model based on the same framework as Romer's model provides a key contribution and connects economic growth to the dynamics of firm behavior. This model called the Schumpeterian model, named after the process of creative destruction discussed in Schumpeter's book (Schumpeter, 1942), was introduced by Aghion and Howitt (1992). The creative destruction process refers to the mechanism whereby as new industries enter the market some are destroyed. In the long-run, both models have identical results but the differences are in the nature of innovation and in the use of intermediate-goods.

Romer's model exhibits a scaling effect that is not supported by data. The scaling effect is that small countries should grow at a lower rate compared to very large countries; however, Jones' work (Jones, 1995) shows no evidence of this effect. Thus, aggregation and scale effects are crucial issues. An alternative to the neoclassic ap-

proach is provided by neo-Ricardians. This economic school re-assesses the importance of income distribution within the economic growth framework. It is closer to the approach of post-Keynesian and neo-Marxian economics.

The neo-Ricardianism is more focused on the balance of growth, the rate of profits for the capitalists and savings for workers, and wages. The fact that technologies are highly decomposable (Arthur, 2009) makes it easy for producers to specialize and to coordinate their activity (Stein, 1982). Our model is in the spirit of earlier work by the Neo-Ricardians, including Sraffa (1960), Pasinetti (1981, 1993) and Kurtz and Salvadori (1995). They were interested in the relationship between wages, profits, growth and consumption, which they studied in the context of input-output networks. In particular Pasinetti considered the relationship between price changes and technological improvement. He did not recognize the connection between network properties and growth that we identify here, but his work can nonetheless be viewed as a precursor.

2.2 Network economics

The idea of network production in the economy was first introduced by Quesnay (1758) who worked on describing the economy from an analytical point of view. Based on a general system of equations proposed by Walras (1874) that relates prices and goods flows, Marshall (1890) proposed a model of aggregated increasing returns of firms. Potron (1912) was the first to propose a full theoretical model of economic interdependence. His work remained relatively unknown and Leontief (1936) independently began the field of input-output analysis. He conducted many empirical studies on the production network of the United States (Leontief, 1941).

Lots of work has been done on understanding the impact of government intervention and of price shocks on input-output economies (Miller and Blair, 2009; ten Raa,

2010). However, less has been done on using the input-output properties to study the network effects on technological evolution. Hulten (1978) introduced the idea that the economy can amplify technolgical shocks. However, he did not provide any explanation of the underlying mechanism and the multiplier is set arbitrarily. Recently, networks became a field of interest for economists like Acemoglu et al. (2012) with their work on aggregated fluctuations.

The network properties of the economy have been the subject of much analysis, both empirical (Chenery and Watanabe, 1958; Aroche-Reyes, 2003; Carvalho, 2008, 2014; Blöchl et al., 2011; McNerney et al., 2013; Fadinger et al., 2015; Bartelme and Gorodnichenko, 2015) and theoretical (Miller and Blair, 2009; Carvalho, 2008; Atalay et al., 2011). There has been a focus on the ability of the network structure of the economy to selectively amplify shocks. Gabaix (2011) showed that shocks hitting firms of different sizes are not balanced out meaning that the law of large numbers cannot be applied. Following Long and Plosser (1983); Carvalho (2008), Acemoglu et al. (2012), Acemoglu et al. (2015), and Baqaee (2015) show that shocks on the input-output linkages lead to aggregate fluctuations. Several mechanisms have been proposed to evaluate the importance of the amplification and propagation of shocks (Horvath, 2000; Carvalho, 2008; Foerster et al., 2011). Recently, some papers focus on the propagation of large shocks in the input-output network, for instance the competition between China and the US (Acemoglu et al., 2016) and the 2011 Japanese earthquake (Boehm et al., 2014; Barrot and Sauvagnat, 2014; Carvalho et al., 2014).

Chapter 3

Model of economic growth

3.1 Introduction

In this chapter we show how the network structure of the economy amplifies long-run growth.[1] To see the basic ideas compare the two schematic economies shown in Figure 3.1. The economy on the left is flat, meaning that each of the two industries S_1 and S_2 produce a final good, which is consumed by households. Labour is the only input and for simplicity we assume a single final good.

The economy on the right, in contrast, has a chain structure, in which industry S_1 produces an intermediate good that is an input to industry S_2, and only S_2 produces the final good. Suppose that industries S_1 and S_2 each locally improve at rate γ. (The precise meaning of this will be made clear later.) In the flat economy the improvements made by the two industries are directly reflected in the final good, and the economy grows at rate γ. In contrast, in the chain economy improvements are amplified multiplicatively going downstream: Industry S_1 improves the intermediate input good at rate γ, then S_2 further improves it, so that the economy grows at a

[1] An early version of the model presented here was developed in J. McNerney's PhD thesis (McNerney, 2012), for which D. Farmer was an advisor. Most aspects of the model that is presented in this chapter were developed by McNerney, Caravelli and Farmer before I joined the collaboration. An exception is section 3.3.2.7, on aggregate output multipliers for open economies, which is my original work. An older version of Figure 4.5 was prepared by McNerney.

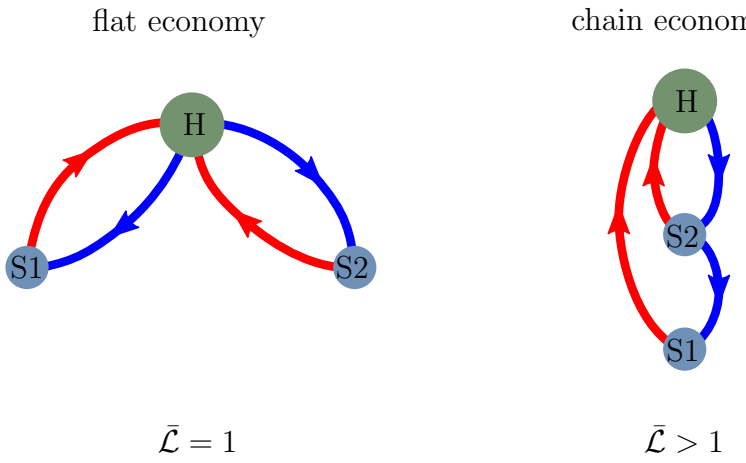


Figure 3.1: *An illustration of how network structure can amplify growth.* We compare two economies with industries S_1 and S_2 and one final good, which is consumed by households (H). Red links represent labour and blue links represent physical goods; arrows indicate the flow of money. In the flat economy (left) both industries make the final good, while in the chain economy (right) industry S_2 produces the final good and industry S_1 produces an intermediate good. If each industry improves locally at the same rate, the multiplicative amplification of improvements, denoted $\bar{\mathcal{L}}$, in the chain economy amplifies growth and causes it to grow faster than the flat economy.

rate greater than γ.

To understand this quantitatively we make a simple model for technological improvement based on the tuning of physical inputs. The model makes it possible to relate technological improvement rates to economic growth at the level of industries or countries for the realistic case in which the input-output matrix has a complicated network structure containing cycles. This makes it possible to distinguish the improvement made by a given industry on its own from the improvements that it inherits from other industries.

We introduce a notion of generalized output multipliers that quantify the amplification of technological improvement for a given industry or country. In general these depend both on improvement rates and on the production network, but we show that for a closed economy the output multiplier is equal to the ratio of gross output to net output. This implies that it is independent of improvement rates and that it remains invariant under aggregation as long as accounting is done properly.

First, I explain the simplification that we make and then I describe some general results from the input-output theory. In the second part, we derive the main equations of the model and study how it affects prices and consumptions. Finally, we will analyse a simple example to give some intuition how the model works.

3.2 Input-output network

This section is divided into a general approach to input-output economics with a description of some hypotheses and a presentation of key results from the input-output theory that are prerequisites for the model.

3.2.1 Fundamental approach

The input-output analysis named by Leontief (1936, 1941) introduces a framework for the analysis of industry interdependence that describes how a product is distributed between economic agents. The condition for the viability of an economy is defined by the Hawkins-Simon theorem (Hawkins and Simon, 1949). Here, I provide a general description that is sufficient for the derivation of the model. An extensive description can be found in Miller and Blair (2009).

3.2.1.1 Simplifications

The economy considered here is represented by a network of exchanges between two types of agents: households and firms. Households are aggregated into a single representative node while each firm or sector is an individual node. Nodes refer to both industries and the good they produce. This can be done at an arbitrarily fine-grained level, i.e. the nodes could also correspond to individual technologies and the process that manufactures them. Thus, other economic agents such as government, bank, and central bank are left apart and import and export exchanges are discarded. The

economy is a closed system where interactions occur only between agents that are explicitly defined.

One can either think in terms of the physical production network, in which case the edges correspond to the flow of goods between industries, or the monetary network, in which cases the edges represent the flow of money. Goods and money flow in opposite directions. We ignore goods flow from and towards the environment.

3.2.1.2 General framework

An input-output economy composed of n firms and 1 representative household is mathematically described by a set of $n + 1$ linear equations with $n + 1$ unknown variables. The set of equations can be easily written in a compact way using matrix notation:

$$\mathbf{Y} = M\mathbf{1} + \mathbf{f}, \tag{3.1}$$

where $\mathbf{Y}$ is a column vector of total outputs, M is the inter-industry transactions matrix, $\mathbf{1}$ a column vector of 1s, and $\mathbf{f}$ the final demand of the production. The transaction between sector i and sector j is given by the coefficient M_{ij} of the inter-industry sales matrix.[2] This system of equations is summarized in an input-output table illustrated by Table 3.1 and it represents the raw data available for this work.

The final demand column summarizes the household consumption and government expenditure. Moreover, the input-output table includes an extra row that describes value added. The value added is the sum of households' compensation, taxes on production and imports, and gross operating surplus less subsidies. In a closed economy, the sum of each row has to be equal to the total output and it is also equal to the sum of each column.

From a network perspective, industries are the nodes and money flows are the edges. Thus, an input-output economy forms a directed network. As mentioned ear-

[2] All the entries in the table are in dollars.

	Buyer Industry			
Seller Industry	Industry 1	Industry 2	Final Demand	Total Output
Industry 1	M_{11}	M_{12}	f_1	Y_1
Industry 2	M_{21}	M_{22}	f_2	Y_2
Value Added	v_1	v_2		
Total Output	Y_1	Y_2		

Table 3.1: *Input-output table illustration of a closed economy with 2 industries and households.* This table representation is the same as in Figure 3.1 but more general. The main property of an input-output table is that the row sum is equal to the column sum.

lier, two types of flows exist. These are goods flows and money flows and both are related through prices of goods. In general, the input-output table is expressed in terms of money flows.

3.2.2 Main results derived from input-output theory

This section summarizes the derivation of basic results from input-output theory that are used in the model. The flow rate of goods from j to i is denoted by X_{ij} and the the flow rate of money from j to i by M_{ij}. If p_i is the price of good i then both flow rates are related by:

$$M_{ij} = X_{ji} p_i, \tag{3.2}$$

we note that while the money flow is directed from j to i, the goods flow is from i to j and is multiplied by the price of one unit of good i. By definition, because there is no saving or investment in the model, the system is at equilibrium and so the flow of money towards a node is equal to all its expenses:

$$\sum_j M_{ij} = \sum_j M_{ji}, \tag{3.3}$$

where the summation occurs over the whole set of nodes. This assumption is essential to the model and is the only one required from input-output theory. Using Eq. (3.2), it becomes:

$$\sum_j X_{ji} p_i = \sum_j X_{ij} p_j. \tag{3.4}$$

The matrix of usage ratios, denoted $\bar{\Phi}$, gives the amount of input needed to produce one unit of output and its elements are:

$$\bar{\phi}_{ij} = \frac{X_{ij}}{\sum_j X_{ji}}, \tag{3.5}$$

where $\sum_j X_{ji}$ is the total output production of the industry i. With this normalization, the matrix $\bar{\Phi}$ can be viewed as a stochastic matrix describing a Markov process. Eq. (3.4) can be rewritten in a compact vector notation:

$$\mathbf{p} = \bar{\Phi}\mathbf{p}. \tag{3.6}$$

This equation is an eigenvector equation for the prices of goods in the input-output economy. It sets the equilibrium price of goods such that it equilibrates physical inputs. Because, available data are expressed in terms of money flow, using Eqs. (3.3), (3.2), and (3.5) the matrix of money transition rates, $\bar{A}$, is defined by:

$$\bar{a}_{ij} = \frac{M_{ij}}{\sum_i M_{ij}} = \frac{M_{ij}}{\sum_i M_{ji}} = \frac{X_{ji} p_i}{\sum_i X_{ij} p_j} = \bar{\phi}_{ji} \frac{p_i}{p_j}. \tag{3.7}$$

$\bar{A}$ is also a stochastic matrix because of the normalization and for any j, $\sum_i \bar{a}_{ij} = 1$. Thus, $\bar{A}$ and $\bar{\Phi}$ are related to each other by the relation:

$$\bar{A} = P \bar{\Phi}^T P^{-1}, \tag{3.8}$$

where P is a matrix with the prices p_i along the diagonal. The element $\bar{a}_{ij}$ could have been formally defined as

$$\bar{a}_{ij} = \frac{M_{ij}}{\sum_i M_{ji}}, \tag{3.9}$$

which is the expenditure on good i per dollar of production by j. In input-output economics these quantities are often called input coefficients or technical coefficients. The interpretation of these quantities as transition probabilities is preferred as it returns to the intuition of the network as a Markov chain of money flows.

Now the economy is split into households and firms. Assuming there are n firms and one household representative agent, the amount of labour provided from the household to the firm i is $L_i = X_{i,n+1}$. Similarly, the household's consumption of good i is $C_i = X_{n+1,i}$ and with no loss of generality because it plays no role in the model so the use of labour by households, ϕ_{HH}, is set to 0. Then, Eq. (3.5) can be rewritten

$$\bar{\Phi} = \begin{pmatrix} \Phi & \boldsymbol{\ell} \\ \mathbf{c} & \phi_{HH} \end{pmatrix}, \tag{3.10}$$

where Φ is a $n \times n$ matrix with components

$$\phi_{ij} = \frac{X_{ij}}{\sum_{j=1}^{n} X_{ji} + C_i}. \tag{3.11}$$

here $\boldsymbol{\ell}$ is the labour column vector of size n and its elements are

$$\ell_i = \frac{L_i}{\sum_{j=1}^{n} X_{ji} + C_i}, \tag{3.12}$$

and $\mathbf{c}$ is the row vector of consumption, the n components are given by

$$c_i = \frac{C_i}{\sum_{j=1}^{n} L_j} = \frac{C_i}{L}, \tag{3.13}$$

with $L = \sum_{j=1}^{n} L_j$. Finally, using Eq. (3.8) the equation for money transition rates of the network is

$$\bar{A} = \begin{pmatrix} A & \tilde{\mathbf{c}} \\ \tilde{\boldsymbol{\ell}} & 0 \end{pmatrix}, \tag{3.14}$$

where A is the $n \times n$ matrix of industrial money transitions and its elements are defined by Eq. (3.7). Here $\tilde{\boldsymbol{\ell}}$ and $\tilde{\mathbf{c}}$ are respectively a row and a column vector of size n and their elements are

$$\tilde{\ell}_i = \frac{\omega \ell_i}{p_i}, \tag{3.15}$$

$$\tilde{c}_i = \frac{p_i c_i}{\omega}, \tag{3.16}$$

where ω is the wage per unit of labour provided by households.

3.3 Description of the model

Let us describe a model of technological change using input-output theory and derive equations that describe how network effects influence the price and the GDP dynamics.

3.3.1 Model assumptions

We consider a very simple model of technological improvement in which the economy becomes more efficient by producing the same quantity of goods with fewer inputs.

This can be represented as a transformation of the input coefficients of the form

$$\phi_{ij} \to \alpha_{ij}\phi_{ij}. \tag{3.17}$$

The factor α_{ij} represents a change in the quantity of good j needed to produce good i. If $\alpha_{ij} < 1$ then the same quantity of i is produced with less of input j and the economy has become more efficient.

This model attempts to capture the essence of a complicated process. In general all the inputs of a given good change differently according to factors α_{ij} that differ for each possible pair of inputs and outputs. Changes in Φ may also come about for other reasons, such as changes in the availability of raw materials, creation and destruction of goods or changes in quality. A more realistic model would incorporate the creation of new goods and their substitution for old ones. For tractability we neglect all such effects.

We simplify things further by assuming that improvement happens at the level of industries, so that we can assign a time-varying improvement factor $\alpha_i(t)$ to industry i. This also includes the labour inputs. Since we are interested in long-run growth we assume that $\alpha_i(t)$ and all the other quantities in the model are differentiable. Letting a dot over a variable indicate a time derivative, we define the local rate of technological improvement as

$$\gamma_i(t) = -\frac{\dot{\alpha}_i(t)}{\alpha_i(t)}, \tag{3.18}$$

and the core assumptions of the model can be written

$$\dot{\phi}_{ij} = -\gamma_i\phi_{ij}, \tag{3.19}$$

$$\dot{\ell}_i = -\gamma_i\ell_i. \tag{3.20}$$

We use the word "local" to refer to γ_i because it describes the improvement in a given industry's production process in isolation from other industries, i.e. the improvement rate of a single industry if none of the other industries made any improvements. Improvement corresponds to $\gamma_i > 0$; nothing prevents the improvement rate from varying in time.

It is convenient to define the local improvement vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)^T$, where T denotes the transpose of a vector, and to let Γ be the diagonal matrix

$$\Gamma \equiv \begin{pmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_N \end{pmatrix}. \tag{3.21}$$

The local improvement rate γ_i can be understood as the rate of total factor productivity (TFP) improvement of industry i. Several measures of TFP change exist, that quantify changes in output quantity not attributable to changes in input quantity (Coelli et al., 2005). From Eq. (3.20) and Eq. (3.5) one can show that γ_i equals the difference in growth rates of input use and production of outputs for industry i.

3.3.2 How growth depends on network structure

In this section we compute the rate of change of prices and the rate of growth of consumption. This allows us to compute the relationship between the rate of growth of GDP and the local improvement vector $\boldsymbol{\gamma}$.

With the notation introduced in the previous section we can write Eq. (3.6) in the form

$$\mathbf{p}(t) = \Phi(t)\,\mathbf{p}(t) + \boldsymbol{\ell}(t) w(t), \tag{3.22}$$

$$w(t) = \mathbf{c}(t)^T \mathbf{p}(t). \tag{3.23}$$

All variables are time-dependent, so for simplicity we do not write this explicitly unless needed for emphasis.

3.3.2.1 How improvement affects prices

We first compute how prices change as a result of technological improvement. Differentiating Eq. (3.22) gives

$$\dot{\mathbf{p}} = \dot{\Phi}\mathbf{p} + \Phi\dot{\mathbf{p}} + \dot{\boldsymbol{\ell}}w + \boldsymbol{\ell}\dot{w}. \tag{3.24}$$

Let $\rho \equiv \dot{w}/w$ denote the rate of change of nominal wages. Using Eq. (3.20) this becomes

$$\dot{\mathbf{p}} = -\Gamma\Phi\mathbf{p} + \Phi\dot{\mathbf{p}} - \Gamma\boldsymbol{\ell}w + \rho\boldsymbol{\ell}w. \tag{3.25}$$

Solving for $\dot{\mathbf{p}}$ gives

$$\dot{\mathbf{p}} = (I - \Phi)^{-1}\left[-\Gamma\mathbf{p} + \rho\boldsymbol{\ell}w\right]. \tag{3.26}$$

Working directly with goods flows has the problem that different types of goods are not directly comparable because different goods have different units. To solve this problem we convert to money flows using Eqs. (3.8), (3.15), and (3.16). Defining the vector giving the rate of change of nominal prices as $\mathbf{r}' \equiv \dot{\mathbf{p}}/\mathbf{p}$, the result is

$$\mathbf{r}' = (I - A^T)^{-1}[-\boldsymbol{\gamma} + \rho\tilde{\boldsymbol{\ell}}]. \tag{3.27}$$

The quantity $(I - A^T)^{-1}$ is the Leontief inverse, which appears ubiquitously in input-output economics. Letting $\mathbf{r} = \mathbf{r}' - \rho\mathbf{1}$ be the rate of change of real prices, and making

use of the relationship

$$(I - A^T)^{-1}\tilde{\boldsymbol{\ell}} = \mathbf{1}, \tag{3.28}$$

as shown in Appendix 3.A, Eq. (3.27) becomes

$$\mathbf{r} = -(I - A^T)^{-1}\boldsymbol{\gamma}. \tag{3.29}$$

3.3.2.2 How improvement affect consumption

To calculate the rate of consumption we take the time derivative of Eq. (3.23), which is

$$\dot{w} = \dot{\mathbf{c}}^T\mathbf{p} + \mathbf{c}^T\dot{\mathbf{p}}. \tag{3.30}$$

Using Eq. (3.13) this can be written

$$\dot{w} = \sum_i \frac{C_i}{L}\left(\frac{\dot{C}_i}{C_i} - \frac{\dot{L}}{L}\right)p_i + \sum_i \frac{C_i}{L}\dot{p}_i. \tag{3.31}$$

To simplify the notation and interpret this result we make the following definitions:

Total labour growth rate :

$$h \equiv \dot{L}/L,$$

GDP :

$$\Pi \equiv \sum_i p_i C_i,$$

Fractional share of GDP for good i :

$$\theta_i \equiv p_i\, C_i/\Pi,$$

Growth rate of consumption of good i :

$$g_i' \equiv \dot{C}_i/C_i.$$

By definition $\sum_i \theta_i(t) = 1$. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)^T$ be the vector of GDP shares and $\mathbf{g}' = (g'_1, g'_2, \ldots, g'_N)^T$ be the vector of GDP growth rates. Substituting these definitions into Eq. (3.31), this becomes

$$\dot{w} = \Big(\boldsymbol{\theta}^T \mathbf{r}' + \boldsymbol{\theta}^T \mathbf{g}' - h\Big) w. \tag{3.32}$$

Letting the overall (real) growth rate of GDP be $g' \equiv \boldsymbol{\theta}^T \mathbf{g}'$, the inflation rate be $r' \equiv \boldsymbol{\theta}^T \mathbf{r}'$, and letting the real growth rate of GDP per unit of labour be $g = g' - h$, this can be written in the simple form

$$g = -r. \tag{3.33}$$

Eq. (3.33) says that the real growth of GDP per unit of labour is equal to the average rate at which real prices decrease.

We can now compute the relationship to growth by multiplying both sides of Eq. (3.29) by $\boldsymbol{\theta}$, which gives the simple relation

$$g = \boldsymbol{\theta}^T (I - A^T)^{-1} \boldsymbol{\gamma}. \tag{3.34}$$

If we make a homogeneity assumption that the local improvement rates are the same for all industries, i.e. $\boldsymbol{\gamma} = \bar{\gamma}\mathbf{1}$, then Eq. (3.34) becomes

$$g = \bar{\gamma} \boldsymbol{\theta}^T (I - A^T)^{-1} \mathbf{1}. \tag{3.35}$$

We define the vector of industry output multipliers as

$$\boldsymbol{\mathcal{L}} = (I - A^T)^{-1} \mathbf{1}. \tag{3.36}$$

Defining the aggregate output multiplier as $\bar{\mathcal{L}} \equiv \boldsymbol{\theta}^T \boldsymbol{\mathcal{L}}$, the previous equation becomes

$$g = \bar{\gamma}\bar{\mathcal{L}}. \tag{3.37}$$

The right hand side is the product of a term $\bar{\gamma}$ that is the local improvement rate and a multiplier that depends only on the flows of money through the production network at a given time.

In order to get a similar expression in the more general case of heterogeneous rates of improvement, we define $\tilde{\gamma}$ as a weighted average of the local improvement rates of each industry, with weights proportional to gross output. Letting $M_j = \sum_i M_{ij}$ be the total money flow from industry j,

$$\tilde{\gamma} = \sum_j \frac{M_j \gamma_j}{\sum_k M_k} = \boldsymbol{\Theta}^T \boldsymbol{\gamma}, \tag{3.38}$$

where the components of $\boldsymbol{\Theta}$ for industry j are the gross output weights $M_j / \sum_k M_k$. We will see in a moment that this choice of $\tilde{\gamma}$ is not arbitrary. We define the vector of generalized industry output multipliers as

$$\tilde{\boldsymbol{\mathcal{L}}} = (I - A^T)^{-1} \boldsymbol{\gamma} / \tilde{\gamma}. \tag{3.39}$$

With this change of notation, Eq. (3.34) becomes

$$g = \tilde{\gamma} \boldsymbol{\theta}^T \tilde{\boldsymbol{\mathcal{L}}}. \tag{3.40}$$

The factors $\gamma_i / \tilde{\gamma}$ can be viewed as distorting the Leontief inverse based on the heterogeneity in the local improvement rates. A surprising result is that even though $\tilde{\boldsymbol{\mathcal{L}}}$ depends on the local improvement rates $\boldsymbol{\gamma}$, for a closed economy the aggregate output multiplier is independent of improvement rates. However the factor $\boldsymbol{\gamma}$ depends

on both improvement rates and network flows. We explain this in Appendix 3.B.

Using arguments that parallel those above, under the homogeneity assumption Eq. (3.29) becomes

$$\mathbf{r} = -\bar{\gamma}\boldsymbol{\mathcal{L}}, \tag{3.41}$$

which in the general heterogeneous case is

$$\mathbf{r} = -\tilde{\gamma}\tilde{\boldsymbol{\mathcal{L}}}. \tag{3.42}$$

3.3.2.3 Relation between variance in growth and aggregation output multiplier

We now derive a relationship between the time variance in growth and the aggregate output multiplier that is closely related to a result of Acemoglu et al. (2012). So far we have assumed differentiability for convenience, but we can alternatively consider a series of homogeneous technological shocks γ_t that influence the corresponding series of GDP growth g_t. Assume that $\bar{\mathcal{L}}$ varies sufficiently slowly that it can be regarded as constant during a single period. Letting $\mathrm{Var}(g_t)$ be the variance of GDP growth per unit of labour and letting $\mathrm{Var}(\gamma_t)$ be the variance of technological shocks, arguments paralleling those leading to Eq. (3.37) give

$$\mathrm{Var}(g_t) = \bar{\mathcal{L}}^2 \mathrm{Var}(\gamma_t), \tag{3.43}$$

or

$$\sigma(g_t) = \bar{\mathcal{L}}\,\sigma(\gamma_t), \tag{3.44}$$

where σ denotes the standard deviation. This is closely related to the main result of Acemoglu et al. (2012).[3]

[3]The result of Acemoglu et al. (2012) is an inequality based on the assumption of an economy

We have a similar relation for the price return of industry i

$$\sigma(r_i) = \mathcal{L}_i \, \sigma(\bar{\gamma}). \tag{3.45}$$

3.3.2.4 Invariance under aggregation

We now address the important question of whether the aggregate output multiplier $\bar{\mathcal{L}}$ is well defined. There are at least two reasons to worry. The first is that there are many ways to partition the economy into industries, and it might seem that the output multiplier could depend on the way in which the partitioning is done, and in particular on the level of granularity. The second is that the local improvement rate depends on the partitioning.[4]

Surprisingly, despite these problems, for a closed economy the aggregate output multiplier $\bar{\mathcal{L}}$ is well-defined, is independent of the level of aggregation, and is independent of growth rates. Defining the gross output of the economy as $\mathcal{O} = \sum_i M_i$, where M_i is the revenue of each industry, we show in Appendix 3.B that

$$\bar{\mathcal{L}} = \boldsymbol{\theta}^T \boldsymbol{\mathcal{L}} = \boldsymbol{\theta}^T \tilde{\boldsymbol{\mathcal{L}}} = \frac{\mathcal{O}}{\Pi}. \tag{3.46}$$

This says that the aggregate output multiplier $\bar{\mathcal{L}}$ is simply the ratio of gross output to net output, and proves the relationship proposed on empirical grounds by Jones (2013).[5]

The aggregate output multiplier only depends on the relative amount of money spent on intermediate goods versus final goods – a high fraction spent on intermediate

with only two levels and no cycles, whereas ours is an equality with no such assumptions. In their case the shocks are on the money flows while here shocks happen on the goods flows.

[4]To see why, suppose their exists a partitioning in which the local improvement rates are homogeneous, and suppose that two of the industries in this partitioning form a chain economy, as shown in Figure 3.1. If these two industries are lumped together in a new partition the improvement rate of the new industry will be larger than that of either industry alone.

[5]Based on empirical observations Jones noted that one over one minus the fraction of intermediate goods, i.e. $\frac{1}{1-\frac{\mathcal{O}-\Pi}{\mathcal{O}}} = \frac{\mathcal{O}}{\Pi}$, approximates the aggregate output multiplier.

goods implies a higher multiplier. This ratio does not depend on the way industries are partitioned, and so the output multiplier is well-defined. In fact under sensible restrictions on growth rates it is possible to show that it is very robust under distortions in accounting.[6]

This is also good news for empirical applications because it means that the output multiplier can be accurately computed even at a high level of aggregation.[7]

This result also implies that, even though the generalized output multipliers of individual industries depend on local improvement rates, the aggregate output multiplier does not. This is clear from the fact that $\boldsymbol{\theta}^T \tilde{\boldsymbol{\mathcal{L}}} = \mathcal{O}/\Pi$, which is independent of improvement rates. As already mentioned, this implies that

$$\boldsymbol{\theta}^T \tilde{\boldsymbol{\mathcal{L}}} = \boldsymbol{\theta}^T \boldsymbol{\mathcal{L}} = \bar{\mathcal{L}}. \tag{3.47}$$

While the generalized industry multipliers change with improvement rates, their GDP weighted sum does not.

It is important to emphasize that, despite the nice invariance of the aggregate output multiplier, the factor $\tilde{\gamma}$ in Eq. (3.40) depends on both improvement rates and network flows. For example, suppose that there are no goods that are both final goods and intermediate goods at the same time, and consider the special case where $\gamma_i = \gamma_F$ for final goods and $\gamma_i = \gamma_I$ for intermediate goods. It is then easy to show that

$$g = \gamma_I \bar{\mathcal{L}} + \gamma_F - \gamma_I. \tag{3.48}$$

[6]To see the potential problem of accounting distortions consider a dummy industry that simply takes in dollars and passes them along to another industry. The presence of this industry cannot affect net output (since it produces no final goods of value), but it could potentially inflate gross output. It is easy to show that providing two conditions are met it does not affect the aggregate output multiplier. The two conditions are that (1) the industry is assigned a zero improvement rate, corresponding to the fact that since it does nothing it cannot improve, and (2) it does not pay wages.

[7]This is true as long as the production of intermediate goods within an industry is properly accounted for. If internal intermediate goods are registered as a self-loop then this is not a problem – the multiplier will be the same as if they were provided by another industry. But if these goods are simply ignored then the multiplier will be underestimated.

In the special case where $\gamma_I = 0$, g is a constant, i.e. growth is independent of the aggregate output multiplier $\bar{\mathcal{L}}$. This is expected since if intermediate goods are not improving then there is no amplification due to network effects. From a mathematical point of view this happens because in this case it is possible to show that $\tilde{\gamma} = \gamma_F/\bar{\mathcal{L}}$. Thus one must be careful in interpreting Eq. (3.40) as a purely multiplicative relationship.

3.3.2.5 Output multipliers and trophic levels

From the point of view of network science the output multiplier is a measure of node centrality, and appears in different guises in many other fields, including ecology, physics, and computer science (Newman, 2010). The formula for the output multiplier of an industry, Eq. (3.28), is similar to Katz centrality (Katz, 1953). It is also similar to PageRank, which is the basis of the algorithm used by Google to rank webpages (Page et al., 1999).

There is a direct analogy between input-output networks and food webs. Food webs characterize the energy flows in an ecosystem, and are one of the most important concepts in ecology. A food web is a directed network in which the nodes are species, analogous to industries (Levine, 1980; Adams et al., 1983; Williams and Martinez, 2004). The link from species i to species j is the fraction of species j in the diet of species i, which is analogous to the input coefficient a_{ji}. In a food web, species are arranged into trophic levels corresponding to their distance from the primary producers that bring energy into the ecology. The trophic level of a species corresponds to the output multiplier of an industry and the trophic depth of the entire food web corresponds to the aggregate output multiplier.[8] Like input-output networks, food webs can be quite complicated and typically contain cycles. By analogy to ecology

[8]In order to preserve the correspondence to output multipliers in economics, we take the convention that the trophic level of households is zero. In contrast, in ecology the trophic level of photosynthesizers is taken to be one.

we will refer to the properties of the input-output network that determine the output multipliers as the trophic depth of the economy. In the following, we use aggregate output multiplier or trophic depth indifferently.

To view industries in terms of their trophic level it is useful to rewrite the definition of the output multiplier, Eq. (3.36), in the form

$$\mathcal{L}_i = 1 + \sum_j \mathcal{L}_j a_{ji}. \tag{3.49}$$

As we will explicitly illustrate in the next section, for a chain economy this becomes a recursive equation in which the output multiplier of each industry can be computed in terms of the industry below it in the chain.

The output multiplier can also be interpreted as the network distance to a given reference point, in this case the mean path length for a dollar to arrive at the household sector. To state this more precisely recall that $\bar{A}$ is a stochastic matrix, with elements a_{ij} that give the probability of a dollar from industry j being paid to industry i. In Appendix 3.B we show that Eq. (3.36) can be rewritten in the form

$$\mathcal{L}_i = \sum_{k=1}^{\infty} k \sum_{j=1}^{N} \tilde{\ell}_j (A^{k-1})_{ji}. \tag{3.50}$$

This is the mean number of times that a dollar starting at industry i changes hands before it is paid to a worker from the household sector.[9] Bearing in mind that goods flows and money flows travel in opposite directions, this path mirrors that taken by the physical flow of goods travelling the other way, and $\mathcal{L}_i$ effectively measures the mean path length of production flows. Averaging over industries, the aggregate output multiplier $\bar{\mathcal{L}}$ is simultaneously the average path length for a dollar received in payment for a good to be paid for labour and the average path length for goods to

[9]The factor $\tilde{\ell}_j (A^{k-1})_{ji}$ is the probability that a dollar from industry i is paid to labour by industry j after k steps; the sums are over all possible industries and all possible numbers of steps.

arrive at and be consumed by the household sector.

3.3.2.6 A simple example

We are now ready to solve the simple example of a chain economy discussed in the introduction and shown in Figure 3.1. For convenience assume the labour supply remains constant so $h = 0$. Industry 1 supplies industry 2 with an intermediate good, and industry 2 produces a final good that it sells to the household sector. The household sector supplies labour to both industries. The input coefficients are

$$\bar{A} \equiv \begin{pmatrix} 0 & a_{12} & 0 \\ 0 & 0 & \tilde{c} \\ \tilde{\ell}_1 & \tilde{\ell}_2 & 0 \end{pmatrix}. \tag{3.51}$$

The normalization condition for $\bar{A}$ requires that $\tilde{\ell}_1 = 1$, $a_{12} + \tilde{\ell}_2 = 1$, and $\tilde{c} = 1$. Because the household sector is the reference node its output multiplier is equal to 0; thus it is clear that since the only input of industry 1 is labour, it has output multiplier $\mathcal{L}_1 = 1$. Using Eq. (3.49) industry 2 has output multiplier

$$\mathcal{L}_2 = \mathcal{L}_1 a_{12} + 1 = 2 - \tilde{\ell}_2. \tag{3.52}$$

Thus depending on its labour share, the output multiplier of industry 2 lies between 1 and 2, reaching its upper limit $\mathcal{L}_2 = 2$ when the labour share $\tilde{\ell}_2 = 0$. Since industry 2 produces the only consumption good its GDP share is one and the aggregate output multiplier is $\bar{\mathcal{L}} = \mathcal{L}_2 = 2 - \tilde{\ell}_2$.

The same result can be derived using Eq. (3.50) by thinking in terms of the mean path length for a dollar spent by industry 2 to reach the household sector. With probability $\tilde{\ell}_2$ it is spent on labour and reaches the household sector in one step, and with probability $(1 - \tilde{\ell}_2)$ it buys the intermediate good produced by industry 1, in

which case it reaches the household sector in two steps. This gives $\mathcal{L}_2 = \tilde{\ell}_2 + 2(1-\tilde{\ell}_2) = 2 - \tilde{\ell}_2$ as before.

This example illustrates that the output multiplier of a given industry depends on two factors: The first is the labour share of the industry and the second is the average output multiplier of the industries that supply it with inputs. This can be seen by the fact that a dollar spent on an intermediate input produced by an industry with a high output multiplier will take more steps to reach the household sector than a dollar spent on a good with a low output multiplier. This mode of thinking will be useful later when we try to understand the trophic depth of real economies.

To see how the output multipliers affect growth, suppose both industries and labour improve locally at rate γ. The relationships derived in the previous section imply that nominal wages remain constant ($w = 0$), while the price of the intermediate good drops at a rate $r_1 = -\gamma$ and the price of the final good drops at the faster rate $r_2 = -\gamma(2 - \tilde{\ell}_2)$. Because the final good gets cheaper while wages remain constant, consumption rises exponentially at the rate $g = \gamma(2 - \tilde{\ell}_2)$. The amplification is evident from the fact that the output multiplier $(2-\tilde{\ell}_2)$ of the industry producing the final good is greater than one. Intuitively this is because improvements in upstream goods are passed on to downstream goods, which benefit both from their own local improvements as well as those of the upstream goods.

3.3.2.7 Aggregate output multipliers for open economies

So far we assumed that we have one closed economy however in reality several countries are trading with each other. Here we show conceptually how the two cases are related, i.e. how trade causes the open-economy multipliers to differ.

We consider a simple example where the world contains just two economies, labelled country 1 and country 2, each with n industries. If the economies are closed

to trade the world input matrix A takes the form

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \tag{3.53}$$

where A_1 and A_2 are the input matrices of the two countries and 0 is an $n \times n$ matrix of zeros. The Leontief inverse is given by

$$(I - A^T)^{-1} = \begin{pmatrix} (I_n - A_1^T)^{-1} & 0 \\ 0 & (I_n - A_2^T)^{-1} \end{pmatrix}, \tag{3.54}$$

where I_n is the $n \times n$ identity matrix, and as shown in Appendix 3.B the aggregate output multiplier of the closed cth economy $\bar{\mathcal{L}}_c^C = \boldsymbol{\theta}_c^T (I - A_c^T)^{-1} \mathbf{1}$ is

$$\bar{\mathcal{L}}_c^C = \frac{\mathcal{O}_c}{\Pi_c}. \tag{3.55}$$

Here $\boldsymbol{\theta}_c$ is the vector of GDP shares for country c, $\mathcal{O}_c$ is gross output, and Π_c is net output. The superscript C denotes a closed economy.

To see how trade causes the aggregate output multipliers to change we consider a first-order perturbation, in which each country uses some imported goods from the other country and correspondingly less domestic goods. We make a small increase to the input coefficients of country 1's imports by an amount $\epsilon_{21} A_1$, while simultaneously reducing country 1's domestic input coefficients by the same amount. Thus total input requirements remain the same (in value terms) for each good. Multiplying all elements of A_1 by the same factor $1 - \epsilon_{21}$ means that for every good a fraction $1 - \epsilon_{21}$ is spent on the domestic version and ϵ_{21} is spent on the foreign version. Making a similar change to country 2's input coefficients, the matrix of input coefficients for

the world becomes

$$A = \begin{pmatrix} (1-\epsilon_{21})A_1 & \epsilon_{12}A_2 \\ \epsilon_{21}A_1 & (1-\epsilon_{12})A_2 \end{pmatrix}. \tag{3.56}$$

It is convenient to define $H_c \equiv (I_n - A_c^T)^{-1}$. The Leontief inverse for the world is then

$$(I - A^T)^{-1} = \begin{pmatrix} B & C \\ D & E \end{pmatrix} + O(\epsilon^2), \tag{3.57}$$

where

$$B = H_1 - \epsilon_{21} H_1 A_1^T H_1, \tag{3.58}$$

$$C = \epsilon_{21} H_1 A_1^T H_2, \tag{3.59}$$

$$D = \epsilon_{12} H_2 A_2^T H_1, \tag{3.60}$$

$$E = H_2 - \epsilon_{12} H_2 A_2^T H_2. \tag{3.61}$$

The aggregate output multiplier of economy 1 becomes

$$\bar{\mathcal{L}}_1^O = \bar{\mathcal{L}}_1^C + \epsilon_{21} \boldsymbol{\theta}_1^T H_1 A_1^T \left(\boldsymbol{\mathcal{L}}_2^C - \boldsymbol{\mathcal{L}}_1^C \right) + O(\epsilon^2). \tag{3.62}$$

Suppose the industry output multipliers $\boldsymbol{\mathcal{L}}_1^C$ of country 1 start out generally smaller than those in country 2. Noting that the elements of $\boldsymbol{\theta}_1$ and $H_1 A_1^T$ are positive, the second term above will lead to a higher aggregate multiplier than if the economy were closed. The opposite effect occurs in country 2, whose aggregate multiplier falls. Thus trade pulls the aggregate output multipliers of the two countries closer together – it raises those of the country with lower multipliers and vice versa. The calculation

can be generalized to N countries, with the result

$$\bar{\mathcal{L}}_c^O = \bar{\mathcal{L}}_c^C + \sum_{\substack{b=1 \\ b\neq c}}^{N} \epsilon_{bc} \boldsymbol{\theta}_c^T H_c A_c^T \left(\boldsymbol{\mathcal{L}}_b^C - \boldsymbol{\mathcal{L}}_c^C \right) + O(\epsilon^2). \tag{3.63}$$

3.4 Conclusion

In this chapter, we relate the economic growth to one property of the input-output network that we call the trophic depth. Under certain assumptions, the relation is simply linear which makes it easy to validate. Economies with high trophic depth are expected to growth at a faster rate as technological improvement will be more amplified by the network structure. We show that this quantity is invariant under aggregation of the network. The result about the invariance is important as it will allow us to test the model even without data at the firm level. This was not expected as in general network properties are sensitive to the level of aggregation. Additionally, we are predicting that the price return of an industry is directly related to the position of the industry in the production chain.

We have highlighted the importance of the connection between input-output networks in economy and food webs in ecology. In future research, we could take advantages of this parallel in more depth. For instance, when modelling different input-output networks we can use results from Johnson et al. (2014) to obtain networks for which key features related to the stability of the dynamic are preserved.

In the following chapter, we will empirically test the predictions of our model to validate our theory. We will demonstrate the importance of the production chain and of the depth of the economy while studying economic growth.

Appendix

3.A A useful identity for computing output multipliers

In this appendix we derive two simple formulas for computing output multipliers that are useful both for calculations and for interpreting results. We begin by showing Eq. (3.28), which states that $(I - A^T)^{-1}\tilde{\boldsymbol{\ell}} = \mathbf{1}$.

$$\begin{aligned}\left[\left(I - A^T\right)^{-1}\tilde{\boldsymbol{\ell}}\right]_i &= \sum_j \tilde{\ell}_j \left[\left(I - A\right)^{-1}\right]_{ji} \\ &= \sum_{k=0}^{\infty}\sum_j \tilde{\ell}_j \left(A^k\right)_{ji} \\ &= 1. \end{aligned} \tag{3.64}$$

To obtain the last line, note that $\sum_j \tilde{\ell}_j \left(A^k\right)_{ji}$ is the probability that a unit of currency starting at industry node i arrives at the household node in exactly $k+1$ steps, with k steps between industries plus a final step to households. With the sum over k the second line gives the probability of ever reaching households, which is 1.

The definition of the output multiplier, $\boldsymbol{\mathcal{L}} = (I - A^T)^{-1}\mathbf{1}$, can be written in two other useful ways. Eq. (3.50) emphasizes its interpretation as a mean path length in

a Markov chain:

$$\mathcal{L}_i = \sum_{k=1}^{\infty} k \sum_{j=1}^{n} \tilde{\ell}_j (A^{k-1})_{ji}. \tag{3.65}$$

To derive this we combine $\boldsymbol{\mathcal{L}} = (I - A^T)^{-1}\mathbf{1}$ with Eq. (3.28) to obtain $\boldsymbol{\mathcal{L}} = (I - A^T)^{-2}\tilde{\boldsymbol{\ell}}$ and note that the matrix $(I - A^T)^{-2}$ has an infinite series expression $\sum_{k=1}^{\infty} k \left(A^T\right)^{k-1}$. Applying this produces Eq. (3.50). This expression makes it easy to see that $\mathcal{L}_i$ is a mean path length. As discussed in Section 3.2.2, the elements of A and $\tilde{\boldsymbol{\ell}}$ can be interpreted as transition probabilities in a network of money flows. The inner sum above is the probability that a random walk in this network starting at node i will arrive at the household node after exactly k steps. With the sum over k, Eq. (3.50) gives the mean number of steps the random walk takes to reach households.

A second way of writing $\mathcal{L}_i$ appears when rearranging the definition of $\boldsymbol{\mathcal{L}}$ as $\boldsymbol{\mathcal{L}} = \mathbf{1} + A^T\boldsymbol{\mathcal{L}}$ and writing in index form we have

$$\mathcal{L}_i = 1 + \sum_j \mathcal{L}_j a_{ji}. \tag{3.66}$$

3.B Invariance of aggregate output multiplier under aggregation

We show here that the aggregate output multiplier $\bar{\mathcal{L}}$ is equal to gross output over net output and is therefore invariant under aggregation. Let $\mathbf{M} = (M_1, \ldots, M_N)^T$ be the vector of gross outputs of each industry and let $\boldsymbol{\Pi} = (\Pi_1, \ldots, \Pi_N)^T$ be the vector of revenues from final consumption of each good. The total revenue of industry i is equal to the sum of revenues from final consumption and intermediate consumption,

i.e.

$$M_i \equiv \Pi_i + \sum_j M_{ij}. \tag{3.67}$$

Using the matrix A to express the intermediate consumption payments as $M_{ij} = a_{ij} M_j$ this can be written in matrix form as

$$\mathbf{M} = \mathbf{\Pi} + A\mathbf{M}. \tag{3.68}$$

Taking the transpose of both sides and solving for $\mathbf{M}^T$ gives

$$\mathbf{M}^T = \mathbf{\Pi}^T (I - A^T)^{-1}. \tag{3.69}$$

Multiplying both sides by $\boldsymbol{\gamma}/\tilde{\gamma}$, writing $\mathbf{\Pi} = \Pi\boldsymbol{\theta}$, and applying Eq. (3.39) this becomes

$$(1/\tilde{\gamma})\mathbf{M}^T\boldsymbol{\gamma} = \Pi\boldsymbol{\theta}^T\tilde{\boldsymbol{\mathcal{L}}}. \tag{3.70}$$

Defining the gross output of the economy as $\mathcal{O} = \sum_{i=1}^{N} M_i$ and $\tilde{\gamma}$ as the output-weighted average growth

$$\tilde{\gamma} \equiv \sum_i \frac{M_i}{\mathcal{O}} \gamma_i, \tag{3.71}$$

Eq. (3.70) becomes

$$\boldsymbol{\theta}^T\tilde{\boldsymbol{\mathcal{L}}} = \bar{\mathcal{L}} = \frac{\mathcal{O}}{\Pi}. \tag{3.72}$$

This shows that the aggregate output multiplier is simply the ratio of gross output to net output, which is independent of the growth rate vector $\boldsymbol{\gamma}$ and is invariant under aggregation. This demonstrates that the output-weighted average growth is

the natural choice for $\tilde{\gamma}$. It means that

$$\bar{\mathcal{L}} = \boldsymbol{\theta}^T \boldsymbol{\mathcal{L}} = \boldsymbol{\theta}^T (1 - A^T)^{-1} \mathbf{1} = \frac{\boldsymbol{\theta}^T (1 - A^T)^{-1} \boldsymbol{\gamma}}{\tilde{\gamma}} = \boldsymbol{\theta}^T \tilde{\boldsymbol{\mathcal{L}}}. \tag{3.73}$$

Chapter 4

Empirical analysis

4.1 Introduction

In this chapter we test the ideas we have developed previously using empirical data. To give a feeling for how trophic depth can be insightful about the properties of the economy, we graphically compare the USA and China using a network representation similar to a food web that we call trophic structure. We find that the two largest economies in the world have surprisingly significantly different trophic depths. The analysis of the trophic structure provides some insights on the production chain of the two economies.

We find that the four predictions under the homogeneous assumption made earlier on economic growth and price returns are validated by the data. The trophic depth is highly correlated with economic growth as we find a highly statistically significant R-squared equal to 0.38. We show that trophic depth is capturing something else that is not already captured by other variables used in standard growth models.

Finally, we also compare the trophic depth with two other complexity variables. We do not find any correlations between the trophic depth and any of the two variables. Therefore, it suggests that they can be complementary nonetheless we find

that on the subset of countries we have, the trophic depth has significantly more explanatory power when doing some growth regressions.

First, we compute some summary statistics for countries and industries and we examine the distribution of local improvement rates. Then, we check the important result that the trophic depth is invariant under aggregation and we test our predictions on prices and growth. Finally, we include the trophic depth in standard growth models and we make comparisons with other complexity variables.

4.2 Data

4.2.1 Presentation of the data

We use data from the World Input-Output Database (WIOD); see Timmer et al. (2015). This consists of input-output tables for 40 countries and 35 industries and covers the period from 1995 to 2009, for a total of 14 years. The 40 countries together capture about 85% of world GDP. The data includes industry production price indices that allow us to compute the vector rate of changes of prices $\mathbf{r}$ in each industry and each country. We also use data from the World Bank (World Bank, 2015) and Penn World Tables (Feenstra et al., 2015) to compare to standard factors used in growth models, and where explicitly stated we use the input-output tables for the US from the Bureau of Economic Analysis (BEA) (US Bureau of Economic Analysis, 2015).

4.2.2 Interpretation of national accounting in the context of our model

Our model makes simple assumptions, including neglecting investment flows and international trade in intermediate goods. Because of this the interpretation of our model in the context of national accounting statistics is not obvious.

The first question is how to interpret the labour term $\tilde{\boldsymbol{\ell}}$. The monetary flows to

households have several components, including payments for labour. Should one include all payments, or only labour? We have done this both ways,[1] and find that the results are qualitatively similar. For example, the univariate regression of $\bar{\mathcal{L}}$ against GDP growth presented in Figure 4.7, which is based only the labour share, gives $R^2 = 0.384$. Including all payments to households gives $R^2 = 0.418$. Similarly there is very little change for the regressions presented in Table 4.4. The main difference is that output multipliers are smaller when one includes all payments to households due to the fact that this increases the flow of money to the household sector and thus shortens path lengths.

To account for international trade, as described in Section 4.3, we treat the world as one large economy and compute the output multipliers of every industry-country pair. Unfortunately the WIOD does not contain data for labour and capital income separately for the Rest Of the World (ROW) region. We therefore neglected ROW when computing the output multipliers based on labour share alone, and for consistency we also neglected it when we included value added (though we did test and including it makes little difference). We obtain the aggregate output multiplier of each country c by weighting the output multipliers of industries by their share of GDP in country c. GDP shares are computed accounting for consumption and investment. We exclude net exports because some countries have industries in which imports would cause industry shares to be negative.

[1] To account for the labour term $\tilde{\boldsymbol{\ell}}$ we used the field titled "Labour compensation" and we use the exchange rates provided by the WIOD to convert to US dollars. For comparison we also used value added at basic prices (row code r64). All results here use the former unless otherwise noted. On the consumption side, e.g. to compute the vector θ describing GDP shares of industries, we used the sum of four terms: Final consumption expenditure by households (column code c37), Final consumption expenditure by non-profit organizations serving households (column code c38), Final consumption expenditure by government (column code c39) and Gross fixed capital formation (column code 41).

4.3 Summary statistics

We first compute some summary statistics to give a feeling for output multipliers and their relationship to GDP and growth. We treat the world as one big economy, constructing the 1400×1400 matrix A of input coefficients corresponding to all possible combinations of 40 countries and 35 goods. We then take the Leontief inverse and compute the 1400-dimensional vector $\boldsymbol{\mathcal{L}} = (I - A^T)^{-1}\mathbf{1}$ of output multipliers whose components correspond to each industry-country pair.

Our theory assumes a closed economy. This is highly restrictive as it yields only one data point for comparison. To make a wider set of predictions we project the vector of output multipliers onto individual countries, ignoring for the moment the fact that national economies are not closed. To compute the aggregate output multiplier $\bar{\mathcal{L}}_c$ for country c we compute the GDP weighted sum of the industry output multipliers for that country, $\bar{\mathcal{L}}_c = \boldsymbol{\theta}_c^T \boldsymbol{\mathcal{L}}$. The GDP weighting factor $\boldsymbol{\theta}_c$ is a 1400 dimensional vector that is normalized to sum to one, and whose entries are zero except for the GDP share of each industry in country c.

To compute the global output multiplier $\mathcal{L}_i$ for industry i we take GDP-weighted averages across countries, i.e. $\mathcal{L}_i = \boldsymbol{\theta}_i^T \boldsymbol{\mathcal{L}}$, where $\boldsymbol{\theta}_i$ is a 1400 dimensional vector that is normalized to sum to one, and whose entries are zero except for the GDP shares of each country in industry i. Note that this is equivalent to ignoring national boundaries and simply computing the 35 dimensional vector of output multipliers for the world.

Unless otherwise stated these computations are done for each year and averages are taken over the entire 14 years period from 1995 to 2009. The countries and their output multipliers are listed in Table 4.1 along with other characteristics such as GDP growth rate, and the industries and their output multipliers are listed in Table 4.2.

Unlike growth rates, which can fluctuate significantly from year to year, aggregate output multipliers vary slowly. Figure 4.1 shows the aggregate output multiplier

Code	Country	GDP per cap ($)	GDP growth (%)	γ_c (%)	$\bar{\mathcal{L}}_c$	$\bar{\mathcal{L}}_c^U$
AUS	Australia	33,510	2.07	1.26	2.84	2.70
AUT	Austria	34,831	1.65	0.81	2.45	2.64
BEL	Belgium	33,129	1.32	-0.02	2.67	2.58
BGR	Bulgaria	3,496	2.74	0.28	3.31	2.80
BRA	Brazil	5,764	1.60	1.24	2.73	2.72
CAN	Canada	32,034	2.22	1.15	2.53	2.64
CHN	China	1,969	9.11	2.62	4.21	2.84
CYP	Cyprus	14,928	1.51	1.63	2.21	2.65
CZE	Czech Republic	11,461	2.54	2.20	3.35	2.75
DEU	Germany	32,429	1.10	-0.15	2.40	2.70
DNK	Denmark	43,216	0.95	0.95	2.30	2.57
ESP	Spain	22,384	1.70	1.08	2.70	2.71
EST	Estonia	8,947	5.30	3.09	2.99	2.68
FIN	Finland	35,473	2.40	0.93	2.66	2.67
FRA	France	32,160	1.13	0.51	2.48	2.67
GBR	United Kingdom	32,000	1.61	0.95	2.42	2.62
GRC	Greece	19,048	2.84	2.04	2.55	2.67
HUN	Hungary	8,686	2.70	1.24	2.81	2.73
IDN	Indonesia	1,454	2.01	2.76	2.81	2.86
IND	India	736	5.77	2.62	2.67	2.84
IRL	Ireland	39,258	3.62	1.47	2.68	2.65
ITA	Italy	27,657	0.44	0.04	2.66	2.70
JPN	Japan	36,610	0.42	-0.13	2.60	2.71
KOR	Korea	15,885	4.44	1.61	2.75	2.73
LTU	Lithuania	6,962	5.66	4.43	2.72	2.72
LUX	Luxembourg	72,162	2.58	2.07	2.46	2.58
LVA	Latvia	6,858	5.87	2.46	2.97	2.63
MEX	Mexico	6,989	1.11	0.52	3.03	2.79
MLT	Malta	14,470	1.92	0.86	2.74	2.66
NLD	Netherlands	37,881	1.85	0.54	2.46	2.57
POL	Poland	7,436	4.32	1.20	2.78	2.70
PRT	Portugal	16,384	1.70	0.25	2.68	2.65
ROM	Romania	4,360	3.40	1.82	3.06	2.85
RUS	Russia	5,275	3.37	3.31	2.68	2.71
SVK	Slovak Republic	8,648	4.39	2.58	3.81	2.71
SVN	Slovenia	16,586	3.25	2.02	2.57	2.65
SWE	Sweden	37,744	1.99	0.73	2.53	2.62
TUR	Turkey	6,227	3.30	2.34	3.49	2.83
TWN	Taiwan	17,111	1.87	1.05	2.52	2.63
USA	United States	40,243	1.66	0.84	2.47	2.65
	World	29,998	2.00	0.96	2.70	2.70

Table 4.1: *Relevant statistics for countries in the WIOD dataset for 1995 - 2009.* The first column gives the ISO code of each country, and the third and fourth columns give the average GDP per capita in dollars and the annualized growth rate of GDP per capita. The fifth column gives the estimated local rate of improvement based on industry price indices using Eq. (4.1). The sixth and seventh columns list the aggregate output multiplier and for comparison the universal output multiplier (see Eqs. (3.36) and (4.2)).

ISIC	Code	Industry	$\mathcal{L}_i$	γ_i (%)	r_i (%)	$\sigma_{\mathcal{L}_i}$	σ_{γ_i}	σ_{r_i}
23	Cok	Coke, Refined Petroleum and Nuclear Fuel	3.76	3.02	0.93	0.64	0.45	0.45
34t35	Tpt	Transport Equipment	3.68	-0.36	-1.94	0.56	0.13	0.21
24	Chm	Chemicals and Chemical Products	3.56	0.97	-1.72	0.45	0.09	0.12
15t16	Fod	Food, Beverages and Tobacco	3.40	-1.32	-1.98	0.34	0.08	0.12
25	Rub	Rubber and Plastics	3.40	2.63	-2.12	0.49	0.10	0.13
27t28	Met	Basic Metals and Fabricated Metal	3.38	-2.70	-0.28	0.50	0.10	0.13
61	Wtt	Water Transport	3.32	2.97	-2.81	0.43	0.21	0.54
29	Mch	Machinery, Nec	3.27	1.22	-1.23	0.45	0.10	0.13
E	Ele	Electricity, Gas and Water Supply	3.27	4.52	-0.89	0.57	0.11	0.13
30t33	Elc	Electrical and Optical Equipment	3.27	3.50	-6.60	0.53	0.11	0.16
20	Wod	Wood and Products of Wood and Cork	3.22	-4.10	-0.93	0.43	0.11	0.23
17t18	Tex	Textiles and Textile Products	3.21	-1.99	1.67	0.43	0.08	0.11
19	Lth	Leather, Leather and Footwear	3.21	-0.64	1.41	0.57	0.13	0.15
36t37	Mnf	Manufacturing, Nec; Recycling	3.17	5.33	-1.18	0.49	0.13	0.17
70	Est	Real Estate Activities	3.17	10.19	-3.63	0.59	0.12	0.16
21t22	Pup	Pulp, Paper, Paper , Printing and Publishing	3.16	0.45	-1.08	0.45	0.07	0.11
26	Omn	Other Non-Metallic Mineral	3.15	0.54	-0.07	0.42	0.12	0.16
62	Ait	Air Transport	3.09	5.73	-1.55	0.51	0.18	0.20
C	Min	Mining and Quarrying	3.00	-1.88	2.21	0.40	0.15	0.17
F	Cst	Construction	2.93	-8.96	-1.94	0.50	0.14	0.14
AtB	Agr	Agriculture, Hunting, Forestry and Fishing	2.81	-1.62	-1.35	0.56	0.10	0.14
64	Pst	Post and Telecommunications	2.75	9.76	-5.07	0.52	0.10	0.12
H	Htl	Hotels and Restaurants	2.71	-0.64	-2.80	0.44	0.11	0.13
60	Ldt	Inland Transport	2.66	0.99	-1.51	0.49	0.16	0.12
63	Otr	Other Supporting and Auxiliary Transport Activities; Activities of Travel Agencies	2.60	1.36	-3.88	0.64	0.14	0.29
J	Fin	Financial Intermediation	2.55	-5.36	-3.84	0.44	0.10	0.16
O	Ocm	Other Community, Social and Personal Services	2.50	0.42	-2.50	0.45	0.08	0.15
51	Whl	Wholesale Trade and Commission Trade, Except of Motor Vehicles and Motorcycles	2.42	-18.10	-3.48	0.48	0.12	0.13
71t74	Obs	Renting of M&Eq and Other Business Activities	2.36	5.33	-5.73	0.49	0.08	0.11
50	Sal	Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel	2.33	1.39	-3.24	0.60	0.13	0.17
N	Hth	Health and Social Work	2.33	3.31	-3.51	0.48	0.06	0.09
52	Rtl	Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods	2.25	-9.14	-3.10	0.48	0.11	0.13
L	Pub	Public Admin and Defence; Compulsory Social Security	2.13	0.62	-2.93	0.41	0.07	0.07
M	Edu	Education	1.81	0.41	-2.41	0.38	0.05	0.08
P	Pvt	Private Households with Employed Persons	1.03	4.26	-0.61	0.20	0.08	0.09

Table 4.2: *Industries from the WIOD dataset.* The ISIC code corresponds to the ISIC Rev. 3 code. Column 3 gives the multiplier of each industry, computed by averaging its multiplier in each country weighted by world GDP share; industries are ordered according to their multiplier. Column 4 gives the local rate of improvement, and column 5 gives the real annual return of the price index averaged across countries (weighted according to GDP share). Columns 6-8 give the standard deviations of these variables across countries.

versus time for a few selected countries. In most cases it remains relatively constant, but for a few countries, including China, Turkey, and South Korea, there are periods where it increases more rapidly. The aggregate output multiplier of many countries drops after the global financial crisis in 2008.

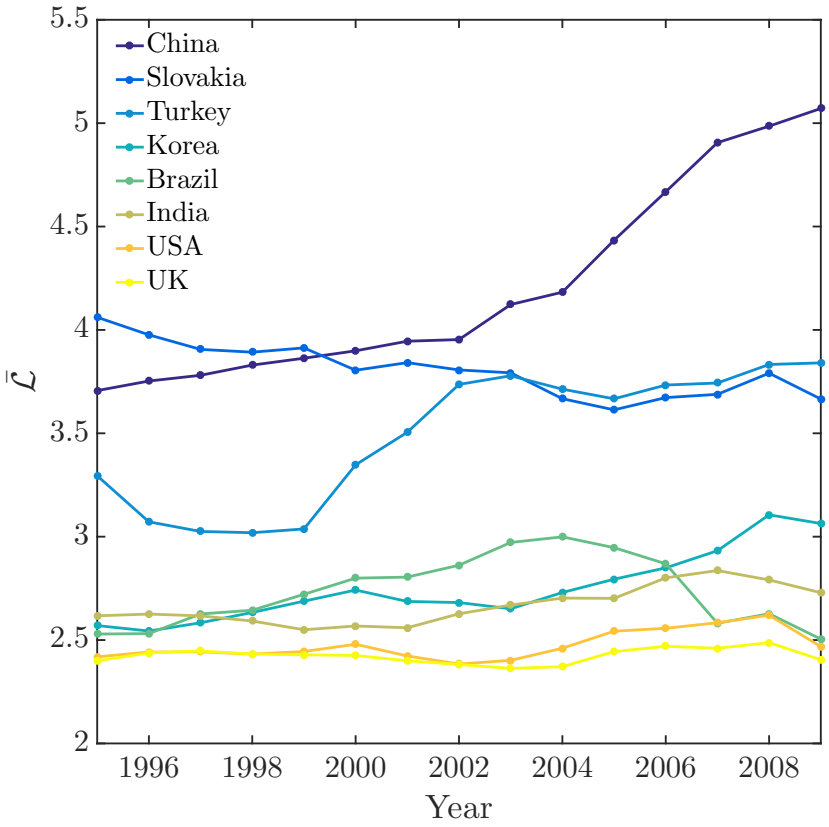


Figure 4.1: *The output multiplier versus time for a selection of countries.*

An interesting pattern emerges if we look at the aggregate output multiplier as a function of GDP per capita as shown in Figure 4.2. Highly developed countries have relatively low multipliers in a range from 2.4 – 2.8. Less developed countries show more variability, ranging from Cyprus, with a multiplier of 2.2 to China, with a multiplier of 4.2. Unfortunately undeveloped economies are not well represented in the data.

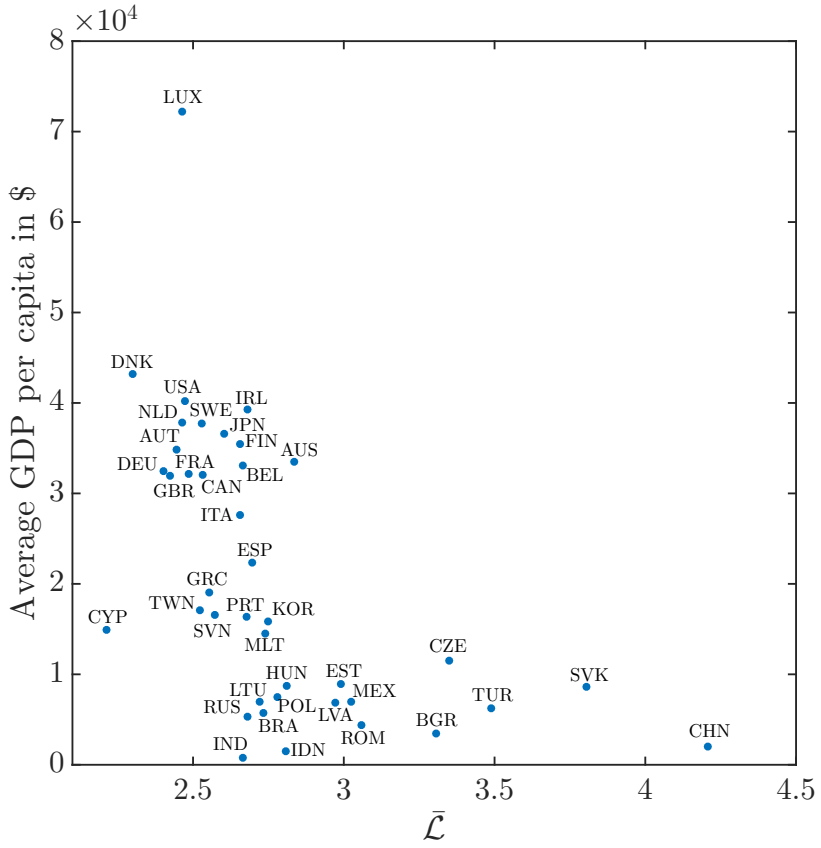


Figure 4.2: *Average GDP per capita in US dollars versus the output multiplier for each country in the WIOD database.* Country codes are listed in Table 4.1.

4.4 Two examples: the USA and China

To provide a feeling for the trophic structure of real economies and to illustrate how it can be insightful, in Figure 4.3 we show the input-output network for the USA and China.[2] These are chosen because they are the two largest economies in the world and because they provide a good contrast. On the vertical axis industries are positioned according to their output multiplier. On the horizontal axis they are positioned based on their industry classification; we use the ISIC codes (revision 3) from the WIOD,

[2]The trophic structure refers to the network representation of an economy while the trophic depth is the aggregated parameter that characterizes the trophic structure.

which are similar to those of the NAICS classification scheme.[3]

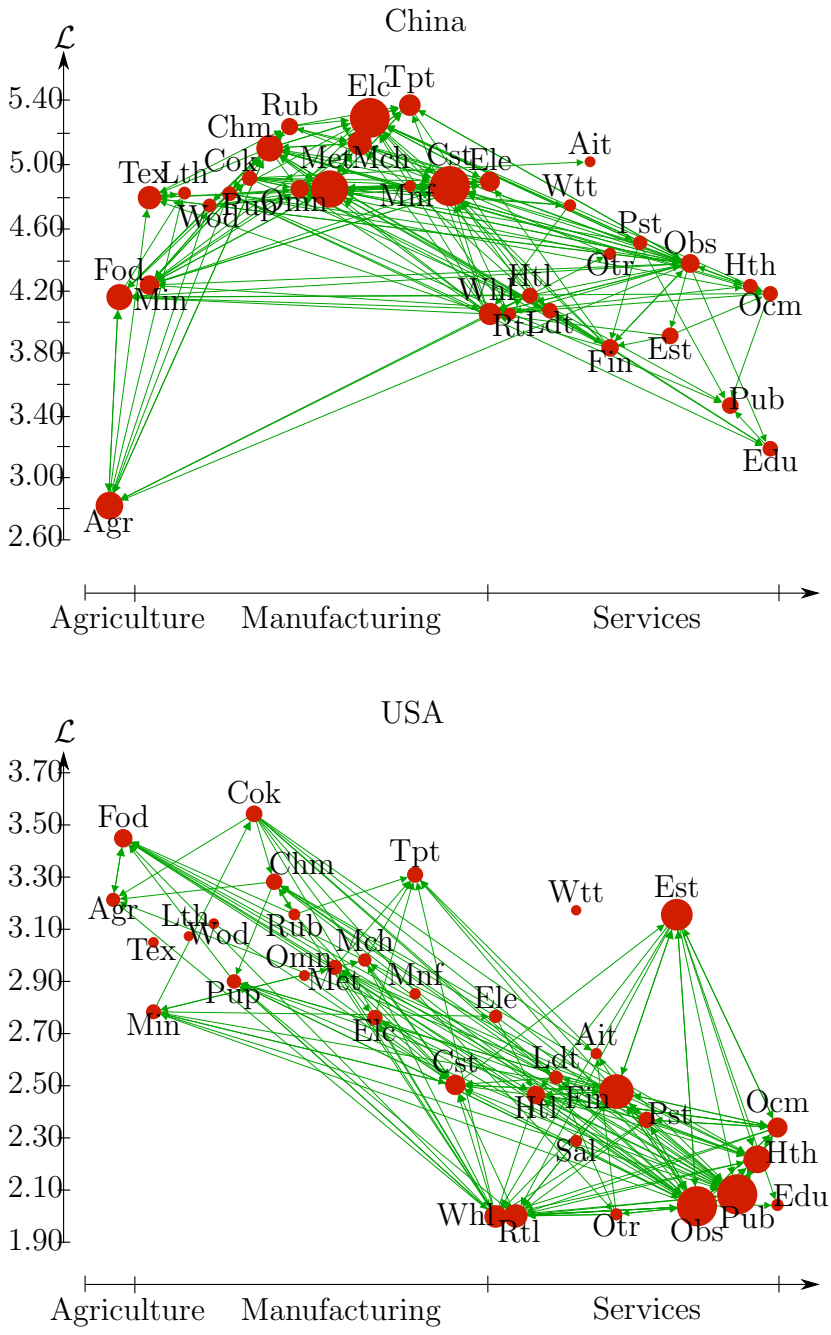


Figure 4.3: *Trophic structure of the Chinese economy (top) and of the American economy (bottom).* Industries are vertically positioned according to their output multiplier and horizontally positioned by industry classification. The node size corresponds to the industry GDP share; only links with money flows higher than $10 billion are shown. For the meaning of the industry codes see Table 4.2.

[3]Due to lack of data the Private Households with Employed Persons sector is not shown for either country, and the Sale, Maintenance and Repair of Motor Vehicles sector is not shown for China.

For the Chinese economy the industry at the top of the trophic structure is Tpt (Transportation Equipment), with a multiplier of 5.4; other large industries with high multipliers include Elc (Electrical and Optical Equipment), Chm (Chemicals and Chemical Products), Met (Basic Metals and Fabricated Metals), and Cst (Construction). The industry with the lowest multiplier is Agr (Agriculture). Most of the largest industries have multipliers near five.

The trophic structure of the USA economy is very different. The industry at the top of the trophic structure is Cok (Coke, Refined Petroleum and Nuclear Fuel), with a multiplier of 3.6. This is closely followed by Fod (Food, Beverages and Tobacco). In contrast to China, Agr (Agriculture) has a relatively high multiplier of about three. Of the four largest industries, Est (Real Estate Activities) has the highest multiplier, also roughly three. The other three largest industries are Fin (Financial Intermediation), with a multiplier of about 2.5, as well as Obs (Renting of M&Eq and Other Business Activities) and Pub (Public Admin and Defense; Compulsory Social Security), which have multipliers near 2.

To understand why the Chinese and the American economies look so different it is useful to recall the two principles mentioned in the discussion of the chain economy. To have a high multiplier, a dollar spent in an industry must change hands many times before it reaches the household sector. Thus to have a high multiplier an industry must have a low labour share and it must have inputs from other industries with high multipliers. Such an industry might supply a final good with a minimum of labour and a long chain of inputs, or it might produce an intermediate good with a low labour share and a long chain of inputs. For example, Agriculture in the USA is a highly mechanized and capital-intensive industry, and hence has a relatively high multiplier; Agriculture in China is more labour intensive. The high multiplier of Cok (Coke, Refined Petroleum and Nuclear Fuel) in the USA illustrates that to have a high multiplier is it not essential to produce a final good, but rather it is sufficient to

have inputs which themselves have large multipliers.

One striking difference between the USA and China is that in general multipliers in China are much higher than they are in the US. There are two reasons for this: One is that the Chinese economy is more concentrated in industries that tend to have high multipliers and the other is that the labour share is much lower in China. We will give a more quantitative explanation of this shortly.

In Appendix 4.A, we represent the trophic structure of all the economies listed in the WIOD in 2009.[4] For clarity, we only represent the top 100 links. The size of nodes captures the relative share of an industry in the country GDP. We note that in every country, services have the lowest trophic levels. The more developed an economy is, the bigger the share of services will be and we also observe that most of the links are from or towards a service industry. On the other hand, developing economies will have most of their links connected across the manufacturing sectors.

4.5 Test of invariance under aggregation

In Section 3.3.2.4 we showed that the aggregate output multiplier can be written as the ratio of the gross output to the net output, and that it should be invariant under aggregation. To test the latter property we use the input-output table from the BEA of the USA economy for the year 2002 (US Bureau of Economic Analysis, 2015). In Figure 4.4, we describe the hierarchy of the NAICS classification.

The NAICS classification is based on 6 layers such that the 6-digit number of an industry indicates the corresponding branch at any level of the classification. Based on this tag, it is possible to know at which sector an industry belongs to at any level of aggregation. For instance, the Medicinal and Botanical manufacturing industry has a code equal to 325411; thus, the sector that it belongs to at a 3-digit level is

[4] In Figure 4.3, we plot the average trophic structure from 1995 to 2009, while in the appendix, we only show a snapshot of each economy in 2009.

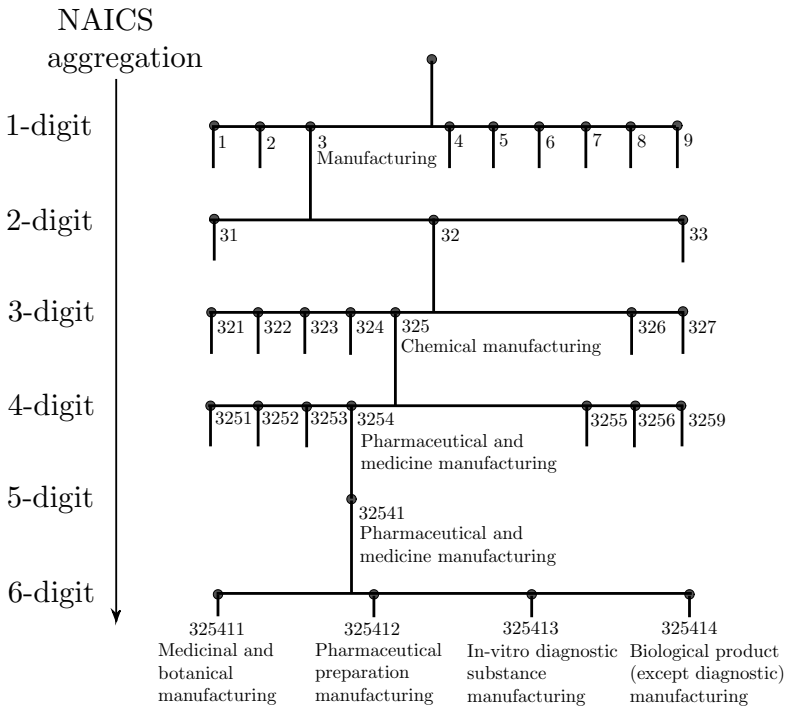


Figure 4.4: *Presentation of the NAICS classification* and a detailed ramification of the Pharmaceutical and Medicine manufacturing sector.

simply given by the first three digits of the code, 325, and refers to the Chemical manufacturing sector. We use this methodology to compute the trophic depth of a country at different level of aggregation.

The output multiplier of industries is computed for different levels of aggregation in Figure 4.5. The output multiplier is roughly invariant, decreasing only slightly when all industries are aggregated into a single node.[5]

[5]If we instead use value added to compute the $\tilde{\boldsymbol{\ell}}$ term rather than only labour, the output multiplier is almost exactly constant and equal to the gross output over the net output, even when the economy is aggregated to a single node (see Section 4.2.2). Note that the industry that has a multiplier of one for aggregation with three or more digits is called Private Households. It summarizes the employment of individual workers by private households and its multiplier is one by definition. The fact that the Agriculture, Forestry, Fishing and Hunting industry has the highest output multiplier here, in contrast to its position in Figure 4.3, is due to the difference in definition and level of aggregation in the WIOD database.

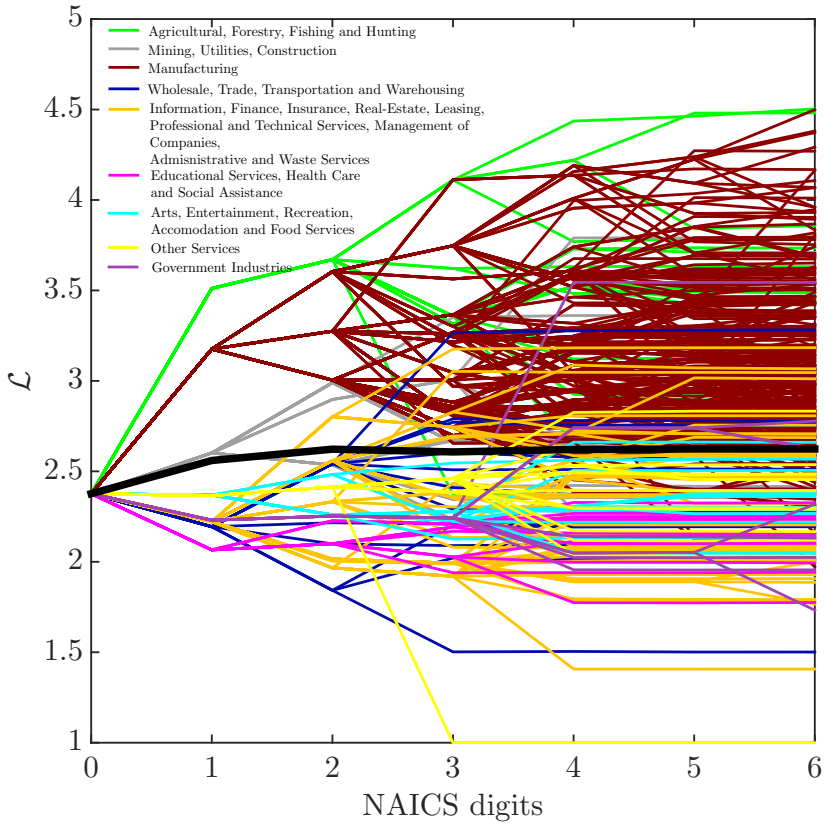


Figure 4.5: *Output multipliers of industries in the U.S. economy at different levels of aggregation*, based on the BEA input-output table for 2002. The black line represents the aggregate output multiplier. The 0-digit refers to the case when all industries are aggregated into a single node.

4.6 Local improvement rates

An estimate $\hat{\boldsymbol{\gamma}}$ for the local improvement rates can be made by rearranging Eq. (3.29) to read

$$\hat{\boldsymbol{\gamma}} = (A^T - I)\mathbf{r}. \tag{4.1}$$

This provides an estimate of the local improvement rates for all of the 1400 industry-country pairs without any free parameters. The estimated local improvement rate $\hat{\gamma}_j$ for industry-country pair j provides an estimate of its improvement that factors out

the improvements that it inherits from its chain of inputs. The distribution of $\hat{\gamma}_j$ at an annual time-scale is shown in Figure 4.6.

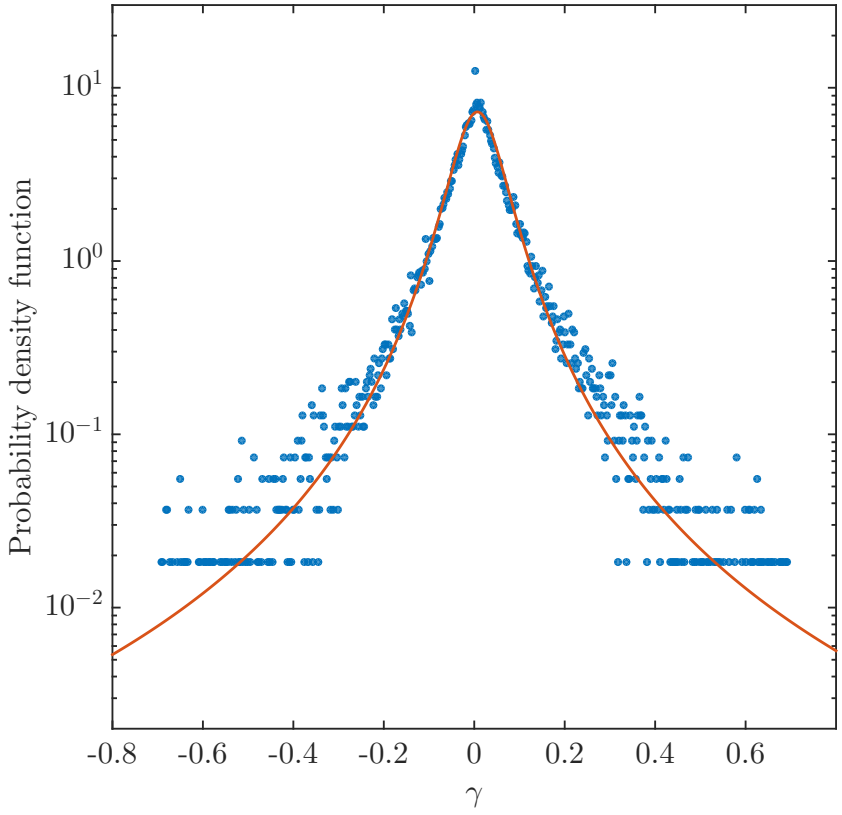


Figure 4.6: *An empirical estimate of the probability density function of the local improvement rate $\hat{\gamma}_j$ based on each industry-country pair j in each year.* For comparison the orange line is a Student distribution with 1.97 degrees of freedom, a mean of 0.71% and a scale parameter of 4.9%.

For comparison we show a Student distribution. The mean is 0.71% and from Table 4.1 the output multiplier of the world is $\bar{\mathcal{L}} = 2.7$. Multiplying the two we get 1.92% which is approximately equal to the average GDP growth rate of 2.00%. The fitted Student distribution has 1.97 degrees of freedom, which indicates that the distribution is sufficiently heavy tailed that the variance does not exist. The scale parameter is 4.9%, which is about seven times as big as the mean. Local improvement rates are observed as high as 70% and as low as -70%.

The average local improvement rate $\hat{\gamma}_c$ for country c can be estimated using Eq. (3.38), which gives $\hat{\gamma}_c = \mathbf{\Theta}_c^T \hat{\boldsymbol{\gamma}}$, where $\mathbf{\Theta}_c$ is the normalized vector of gross output weights for the industries of country c. Similarly the local improvement rate for industry i, averaged over countries, is $\gamma_i = \mathbf{\Theta}_i^T \hat{\boldsymbol{\gamma}}$. These are listed in Table 4.1 and Table 4.2.

As a self-consistency check we test the relationship $g = -r$ of Eq. (3.33). This compares the observed GDP growth for each country to the average GDP-weighted change in its price indices. There is a clear relationship with a high degree of statistical significance, with an $R^2 = 0.529$ and a p-value of $1.04 \cdot 10^{-7}$. However the slope is 0.495, indicating a substantial deviation from the identity line. Thus the average price index tends to overestimate growth for countries higher aggregate multipliers. We do not know whether this discrepancy is due to the poor measurement of the price indices or whether it is due to other assumptions of the theory, e.g. prices that may be substantially out of equilibrium. This should be born in mind when interpreting the estimates of $\hat{\boldsymbol{\gamma}}$, which depend on price indices.

We also observe deviations due to the fact that the economies of individual countries are not closed. As we have shown, if the economies of individual countries were closed, the relationship $\bar{\mathcal{L}} = \boldsymbol{\theta}_c^T \boldsymbol{\mathcal{L}} = \boldsymbol{\theta}_c^T \tilde{\boldsymbol{\mathcal{L}}} = \mathcal{O}/\Pi$ is an identity. Unsurprisingly this is not true for individual countries because their economies are open. Nonetheless, a comparison of any pair of these against each other shows that they have a strong correlation. For example, the regression of the country output multiplier versus the ratio of gross output over net output gives an R-squared equal to 0.50 and a slope of 0.47. If we use value added instead of labour share to compute the output multipliers the R-squared increased to 0.78 and the slope becomes 0.65 (see Section 4.2.2). Some of these deviations can be understood from first principles; for example we show in Section 3.3.2.7 that the multiplier of a given country becomes higher if it trades with other countries with high multipliers.

4.7 GDP growth versus trophic depth

The model predicts that for closed economies GDP growth rates should be positively correlated to aggregate output multipliers (Eq. (3.37) or alternatively Eq. (3.40)). We test this prediction at the level of individual countries, ignoring the fact that they are not closed economies. We regress GDP growth against aggregate output multipliers, allowing for the possibility of growth factors external to the model by adding a constant term. This gives a regression of the form $g_c = a\bar{\mathcal{L}}_c + b$. The factor g_c for country c is the annualized growth rate of GDP per capita between 1995 and 2009.[6]

The results are shown in Figure 4.7, where we plot GDP growth rates for 40 different countries against their output multipliers and compare to the regression line. The output multipliers range from roughly 2.2, corresponding mostly to highly developed countries, and 4.2 at the high end, corresponding to China. The growth rates range from Japan (0.4%) to China (9.1%). The slope of the regression is $a = 2.75\%$ and the intercept is $b = -4.87\%$ and the p-values are $2.03 \cdot 10^{-5}$ and $3.78 \cdot 10^{-3}$ respectively. The R-squared of the fit is 0.38.[7] We thus find a strong and highly statistically significant relationship between the GDP growth of a country and its output multiplier.

We also find that there is a substantial correlation between local improvement rates and output multipliers. This is evident from the fact that the correlation between the local improvement rates γ_c and the output multipliers $\bar{\mathcal{L}}_c$ given in Table 4.1 is 0.36 with a p-value of 0.02. This suggests that the correlation between GDP growth rates and output multipliers shown in Figure 4.7 has two sources. The first is our theoretical prediction that even if local improvement rates are the same, countries with

[6] We take per capita GDP as a proxy for GDP per labour share. It is tempting to interpret the slope coefficient a as a typical country improvement rate $\bar{\gamma}$. However we expect the improvement rate to vary across countries, and by not capturing this variation we would potentially introduce an important omitted variable bias. Given this we avoid interpreting the regression in causal terms and simply ask how well $\bar{\mathcal{L}}_c$ is correlated with g_c.

[7] Removing China from the fit changes the slope to 1.86% but does not substantially change the significance of the regression.

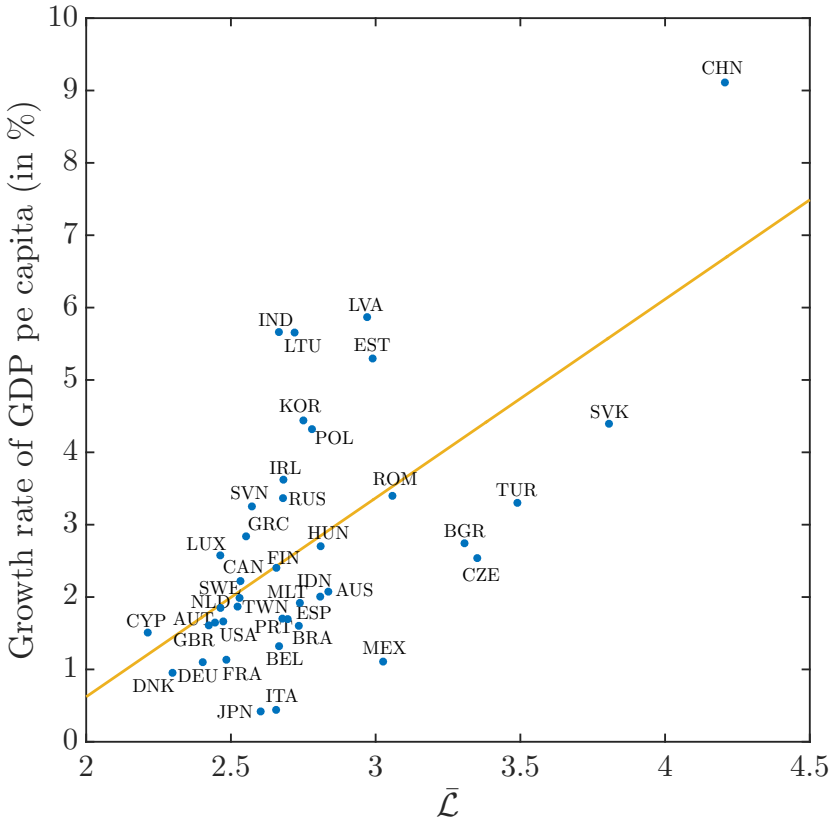


Figure 4.7: *Average growth rate of real GDP per capita versus aggregate output multiplier of countries.* The straight line is a linear regression with a constant term. Three-letter country codes are listed in Table 4.1.

larger output multipliers should grow faster. The second is the empirical observation that countries with larger output multipliers tend to have higher local improvement rates. Our theory says nothing about the second observation, though factors such as investment could increase both the length of a country's production chains and its rate of technological improvement at the same time.

A regression of the standard deviation of the GDP fluctuations for each country against their output multipliers based on Eq. (3.44) yields $R^2 = 0.11$ with a p-value $= 0.036$, indicating the predictions for variance are not equally strong but still significant. Moreover, to test the predictive power of the output multiplier, we

regress the overall economic growth rate versus the output multiplier at the beginning of the data. The slope is 2.17% associated to a p-value equal to $5.25 \cdot 10^{-5}$ and the R-squared is 0.274.

4.8 Price return and trophic level

In the model, we also predict the evolution of industry price indices. Regressing the annualized returns of industry-country pair price indices against the corresponding industry output multipliers following Eq. (3.41) yields a slope of -0.89% associated to a p-value $= 5.74 \cdot 10^{-7}$. The result is shown in Figure 4.8.

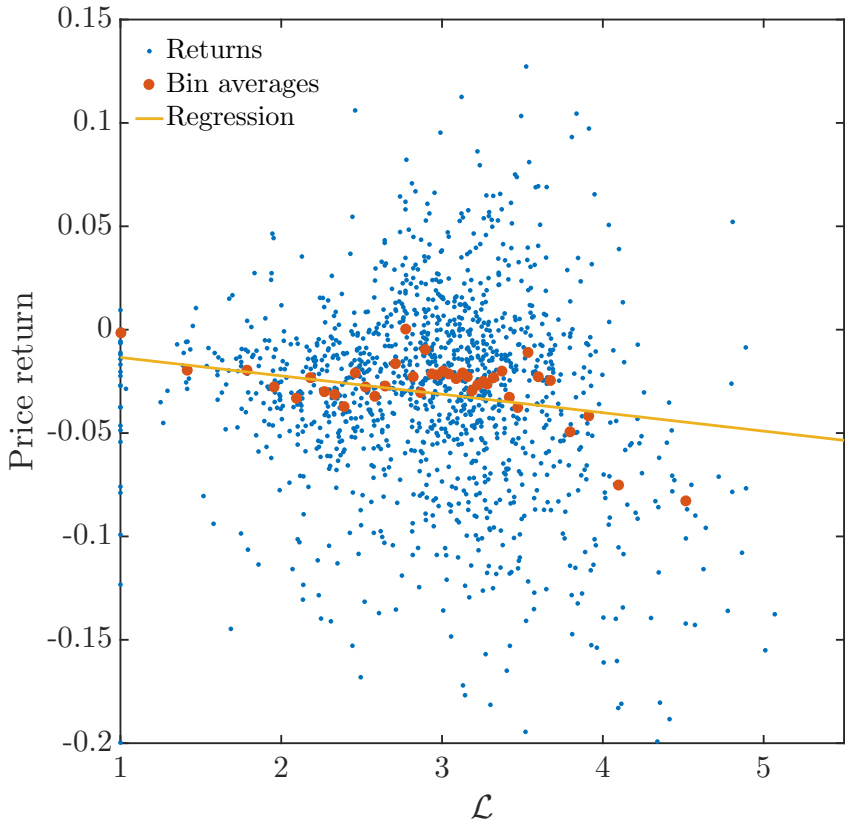


Figure 4.8: *Average industry-country pair price return versus trophic level.* The straight line is a linear regression with a constant term. The red dots show the bin averages.

The R-squared is small but it is not surprising given the large number of points. The intercept is not significant as expected in the model. A regression of the standard deviation of the price indices for each industry against their trophic level has a positive and highly significant relation. However, when we test these two relations at the aggregated level using the data in Table 4.2 only the results on the standard deviation hold. We believe these results reflect the inherent problems with the aggregation at the industry level. Regressing the variance of prices against the trophic level leads to a slope of 0.0665 associated to a p-value equal to $1.64 \cdot 10^{-14}$.

4.9 Relation to standard growth models

The fact that so much variance is captured by the output multiplier immediately raises the question of its relationship to the variables used in standard growth models. Is it capturing something new or is it simply a proxy for something that is well-known? To get insight into this we choose 14 variables that have been studied in growth models,[8] regress them against the country growth rates and estimate their correlation to the output multiplier, as shown in Table 4.3. We only focus on explanatory variables that appear in a Solow model while we could have controlled for other cultural, historical, and institutional parameters (Sala-i Martin, 1997).

We find that the output multiplier has the largest R-squared of any single variable. However this also reveals significant correlations with several variables, including positive correlation to the savings rate and negative correlations to total factor productivity, health expenditure, tax revenue and R&D workers. But the most striking result is the correlation of -92% between the output multiplier and labour share. A strong correlation is to be expected from the theory, as the labour share determines the fraction of the inputs that go directly to the household sector, making it an impor-

[8] We would like to have included the growth rate of total factor productivity as well, but we have not been able to obtain this. We do include R&D workers.

Explanatory variables	GDP per capita growth rate		$\bar{\mathcal{L}}$	
	R-squared	p-value	Correlation coefficient	p-value
$\bar{\mathcal{L}}$	0.384	$2.03 \cdot 10^{-5}$	1.00	0
Capital-labour ratio	0.381	$2.20 \cdot 10^{-5}$	-0.57	$1.27 \cdot 10^{-4}$
Urban population	0.330	$1.07 \cdot 10^{-4}$	-0.41	$9.37 \cdot 10^{-3}$
Health expenditure	0.268	$6.22 \cdot 10^{-4}$	-0.49	$1.42 \cdot 10^{-3}$
Total factor productivity level	0.264	$6.87 \cdot 10^{-4}$	-0.49	$1.41 \cdot 10^{-3}$
Labour income share	0.255	$8.92 \cdot 10^{-4}$	-0.92	$4.53 \cdot 10^{-17}$
Savings rate	0.172	$7.82 \cdot 10^{-3}$	0.30	$5.95 \cdot 10^{-2}$
log(GDP per capita in 1995)	0.141	$1.69 \cdot 10^{-2}$	-0.20	0.211
Researchers in R&D	$8.41 \cdot 10^{-2}$	$6.94 \cdot 10^{-2}$	-0.40	$9.72 \cdot 10^{-3}$
Population growth rate	$7.44 \cdot 10^{-2}$	$8.85 \cdot 10^{-2}$	-0.17	0.299
School enrollment (secondary)	$6.05 \cdot 10^{-2}$	0.126	-0.52	$6.69 \cdot 10^{-4}$
Index of human capital	$2.40 \cdot 10^{-2}$	0.340	-0.19	0.248
Inflation rate (GDP deflator)	$1.17 \cdot 10^{-2}$	0.506	0.39	$1.41 \cdot 10^{-2}$
Tax revenue	$8.97 \cdot 10^{-3}$	0.561	-0.25	0.120
Depreciation rate	$2.32 \cdot 10^{-4}$	0.926	-0.05	0.758

Table 4.3: *Relationship between the aggregate output multipliers of countries and fourteen variables associated with economic growth.* The first row shows the R-squared of a regression when the variables are used alone, and the second row gives the correlation.

tant determinant of the output multiplier. (See the discussion in Sections 3.3.2.5 and 3.3.2.6). With such a large correlation one might presume that the two variables are largely interchangeable, and that the output multiplier is simply a proxy for labour share.

Surprisingly this is not the case. The correlation of the output multiplier to growth rates is substantially larger than for labour share, and despite the fact that the two variables are closely related; the output multiplier contains information that is not included in labour share. This can be seen in several ways: If labour share is regressed against the growth rate by itself it has an $R^2 = 0.25$ with a p-value of $8.92 \cdot 10^{-4}$, which is substantially lower than for the output multiplier alone which has $R^2 = 0.38$ and a p-value $= 2.03 \cdot 10^{-5}$.

To understand this more comprehensively we perform regressions of growth rates against our set of 14 standard growth factors and then perform the same regressions with the output multiplier included as well, as shown in Table 4.4. We use three

different groupings with 3, 8, and 14 explanatory variables. The first grouping summarizes a basic growth model with only capital accumulation. The second set of variables contains variables of the Solow model and the last grouping contains explanatory variables of the Solow model plus some additional variables used in the growth literature. We then redo these regressions adding the output multiplier, i.e. with 4, 9 and 15 explanatory variables.

Including the output multiplier always substantially improves the goodness of fits. A standard growth model using three variables,[9] the logarithm of GDP in 1995, savings rate and labour share, gives an $R^2 = 0.50$, which increases to $R^2 = 0.61$ when the output multiplier is included; similarly in the regression with eight standard variables $R^2 = 0.67 \to 0.72$ and with fourteen variables $R^2 = 0.72 \to 0.77$.

Note that without the output multiplier there is a consistently positive y-intercept ranging from 4.0% to 8.3%, though of varying statistical significance. When the output multiplier is included the y-intercept becomes negative, with values ranging from -8.5% to -12.6%, but it is never statistically significant at the level of two standard deviations. The estimated coefficient for the output multiplier is about 2.8%, in comparison with 2.7% for the univariate regression.

To return to the relative contributions of the output multiplier and labour share, note that in Table 4.4 when the labour share is included without the output multiplier it is only marginally statistically significant at the level of two standard deviations with 3 variables, and it is far from statistically significant when 8 or 14 variables are included. When both labour share and the output multiplier are included the p-values of the output multiplier are always more significant than those of the labour share, in some cases by a substantial amount.

To directly illustrate that the output multiplier contains useful information that

[9]Because labour share and the output multiplier are strongly correlated, we also performed the regression omitting labour share. In this case the estimated slope for the output multiplier is $1.87 \cdot 10^{-2}$ with a p-value of $1.22 \cdot 10^{-3}$.

Explanatory variable	Standard growth model						Standard growth model+$\bar{\mathcal{L}}$					
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
$\bar{\mathcal{L}}^{\star}$							$3.82 \cdot 10^{-2}$	$3.43 \cdot 10^{-3}$	$2.81 \cdot 10^{-2}$	$2.54 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$3.90 \cdot 10^{-2}$
log(GDP per capita in 1995)*	$-9.37 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$-6.77 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$	$-5.41 \cdot 10^{-3}$	0.160	$-8.14 \cdot 10^{-3}$	$1.91 \cdot 10^{-3}$	$-6.15 \cdot 10^{-3}$	$1.52 \cdot 10^{-2}$	$-6.16 \cdot 10^{-3}$	$9.06 \cdot 10^{-2}$
Savings rate*	$1.31 \cdot 10^{-3}$	$2.22 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$2.17 \cdot 10^{-4}$	$1.24 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$1.25 \cdot 10^{-3}$	$1.31 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	$3.37 \cdot 10^{-4}$	$1.14 \cdot 10^{-3}$	$1.25 \cdot 10^{-2}$
Labour income share*	-0.105	$3.03 \cdot 10^{-2}$	$-1.16 \cdot 10^{-2}$	0.810	$-5.75 \cdot 10^{-2}$	0.342	0.174	$8.53 \cdot 10^{-2}$	0.176	$6.37 \cdot 10^{-2}$	0.135	0.206
Depreciation rate			$-7.79 \cdot 10^{-3}$	0.981	$-6.17 \cdot 10^{-2}$	0.871			$-2.32 \cdot 10^{-2}$	0.940	-0.139	0.697
Researchers in R&D			$-1.02 \cdot 10^{-6}$	0.457	$-4.89 \cdot 10^{-7}$	0.756			$-2.96 \cdot 10^{-7}$	0.821	$-2.79E-08$	0.985
Population growth rate			$-4.25 \cdot 10^{-3}$	0.190	$-2.96 \cdot 10^{-3}$	0.453			$-3.98 \cdot 10^{-3}$	0.189	$-2.58 \cdot 10^{-3}$	0.483
Total factor productivity level*			$1.91 \cdot 10^{-2}$	0.233	$1.70 \cdot 10^{-2}$	0.329			$1.40 \cdot 10^{-2}$	0.352	$1.51 \cdot 10^{-2}$	0.351
Capital-labour ratio*			$-1.30 \cdot 10^{-7}$	$3.24 \cdot 10^{-3}$	$-1.34 \cdot 10^{-7}$	$1.07 \cdot 10^{-2}$			$-1.09 \cdot 10^{-7}$	$8.54 \cdot 10^{-3}$	$-1.17 \cdot 10^{-7}$	$1.76 \cdot 10^{-2}$
Index of human capital					$4.10 \cdot 10^{-3}$	0.613					$2.15 \cdot 10^{-3}$	0.778
School enrollment (secondary)					$1.27 \cdot 10^{-4}$	0.631					$2.19 \cdot 10^{-4}$	0.384
Health expenditure*					$1.94 \cdot 10^{-4}$	0.898					$8.16 \cdot 10^{-5}$	0.954
Urban population*					$-3.15 \cdot 10^{-4}$	0.129					$-2.47 \cdot 10^{-4}$	0.205
Tax revenue					$1.38 \cdot 10^{-4}$	0.589					$-9.78 \cdot 10^{-6}$	0.968
Inflation rate (GDP deflator)					$-2.14 \cdot 10^{-4}$	0.242					$-2.16 \cdot 10^{-4}$	0.207
Constant	$8.33 \cdot 10^{-2}$	$3.98 \cdot 10^{-4}$	$3.97 \cdot 10^{-2}$	0.125	$5.71 \cdot 10^{-2}$	0.166	-0.126	$7.83 \cdot 10^{-2}$	-0.105	0.122	$-8.51 \cdot 10^{-2}$	0.268
R-Squared	0.497		0.669		0.725		0.608		0.721		0.770	
Adjusted R-Squared	0.455		0.584		0.570		0.563		0.637		0.627	

Table 4.4: *Regressions of the observed growth rates for 40 countries against a variety of explanatory variables.* This is done for different groupings of standard growth factors excluding the country output multipliers (first three columns) and including them (last three columns). Variables that are statistically significant at the 5% level or lower when they are the sole regression variable are identified with a star.

is not included in the labour share, we create a universal version of the output multiplier that explicitly excludes the contribution of labour share. This is based on the fact that, all else equal, countries with a greater concentration in industries with high output multipliers will have higher aggregate output multipliers. Letting $\boldsymbol{\mathcal{L}}^I$ be the vector of global industry output multipliers $\mathcal{L}_i$ defined in Section 4.3 and shown in Table 2, we compute a universal output multiplier for each country as

$$\bar{\mathcal{L}}_c^U = \boldsymbol{\theta}_c^T \boldsymbol{\mathcal{L}}^I, \tag{4.2}$$

as shown in the last column of Table 4.1. By construction the relative values of $\bar{\mathcal{L}}_c^U$ for different countries do not depend on labour share. Regression of growth rates on universal output multipliers gives an $R^2 = 0.145$ with a p-value of $1.52 \cdot 10^{-2}$, in contrast to $R^2 = 0.25$ for labour share. We thus see that labour share explains about two thirds of the variance while industry composition explains about a third. Thus the output multiplier contains useful information that is not in the labour share.

4.10 Other complexity variables in the economic literature

In this section, the trophic depth is compared to two other indexes used in the economic complexity literature to perform cross-country analysis. These indexes are the Economic Complexity Index and the Global Competitive Index.

The first variable is the Economic Complexity Index designed by Hausmann and Hidalgo (Hausmann et al., 2011). It provides a measure of the knowledge of an economy by looking at the complexity of products the economy is exporting. The complexity is not only defined by how complex the exports of a country are but also the number of products exported. Thus, the complexity index is a measure of the "diversity" of a country and of the "ubiquity" of the products it exported.

The Global Competitive Index (Schwab and Sala-i Martin, 2014) is based on 12

pillars that form subindexes. They are classified into three categories: basic requirements subindex (institutions, infrastructure, macroeconomic environment, and health and primary education), efficiency enhancers subindex (higher education and training, goods market efficiency, labour market efficiency, financial market development, technological readiness, and market size), and innovation and sophistication factors subindex (Business sophistication and innovation). The coefficients associated to each category to compute the global competitive index depend on the stage of development of economies.

Contrary to the trophic depth, none of these two indexes are directly related to GDP growth or to the level of GDP. They give no clear evidence on how these variables should be regressed with economic growth. However, ordinary least squares regression is used to regress the economic growth with the Economic Complexity Index and the authors compared and analysed their results with other regressors in section 4 of the first part of their book. A direct comparison between the trophic depth and the economic complexity index is possible. Finally, the logarithm of Global Competitive Index is used instead of the raw Global Competitive Index when it regresses against the economic growth.

Table 4.5 summarizes the values of complexity indexes for 40 countries and the last row shows the R-squared of the regression of the GDP growth versus each regressor independently. The trophic depth performs incredibly well compared to the two other variables, then comes the Economic Complexity Index with an R-squared equals to 0.15 and finally the Global Competitive Index does not explain lots of variance. The correlation coefficient between the trophic depth and the Economic Complexity Index is -0.234 while it is equal to -0.322 with the Global Competitive Index. We note that they are both negative but not largely correlated. The negative sign means that countries with high trophic depths will have lower complexity indexes. The model using the trophic depth explains a much larger amount variance compare to the others.

Code	Country	GDP growth (%)	Trophic Depth	Economic Complexity Index	Global Competitive Index[1]
AUS	Australia	2.07	2.84	-0.10	1.64
AUT	Austria	1.65	2.45	1.81	1.64
BEL	Belgium	1.32	2.67	1.33	1.64
BGR	Bulgaria	2.74	3.31	0.52	1.42
BRA	Brazil	1.60	2.73	0.50	1.44
CAN	Canada	2.22	2.53	0.87	1.67
CHN	China	9.11	4.21	0.61	1.56
CYP	Cyprus	1.51	2.21	0.89	1.48
CZE	Czech Republic	2.54	3.35	1.58	1.52
DEU	Germany	1.10	2.40	2.11	1.70
DNK	Denmark	0.95	2.30	1.37	1.69
ESP	Spain	1.70	2.70	1.05	1.53
EST	Estonia	5.30	2.99	0.69	1.54
FIN	Finland	2.40	2.66	1.86	1.70
FRA	France	1.13	2.48	1.54	1.64
GBR	United Kingdom	1.61	2.42	1.78	1.68
GRC	Greece	2.84	2.55	0.24	1.39
HUN	Hungary	2.70	2.81	1.24	1.46
IDN	Indonesia	2.01	2.81	-0.12	1.47
IND	India	5.77	2.67	0.19	1.46
IRL	Ireland	3.62	2.68	1.45	1.59
ITA	Italy	0.44	2.66	1.45	1.48
JPN	Japan	0.42	2.60	2.39	1.69
KOR	Korea	4.44	2.75	1.39	1.63
LTU	Lithuania	5.66	2.72	0.51	1.49
LUX	Luxembourg	2.58	2.46	na[2]	1.61
LVA	Latvia	5.87	2.97	0.52	1.46
MEX	Mexico	1.11	3.03	0.97	1.45
MLT	Malta	1.92	2.74	na[2]	1.47
NLD	Netherlands	1.85	2.46	1.19	1.69
POL	Poland	4.32	2.78	1.02	1.48
PRT	Portugal	1.70	2.68	0.64	1.49
ROM	Romania	3.40	3.06	0.70	1.41
RUS	Russia	3.37	2.68	0.58	1.44
SVK	Slovak Republic	4.39	3.81	1.33	1.45
SVN	Slovenia	3.25	2.57	1.49	1.48
SWE	Sweden	1.99	2.53	1.98	1.71
TUR	Turkey	3.30	3.49	0.29	1.46
TWN	Taiwan	1.87	2.52	0.73	1.66
USA	United States	1.66	2.47	1.67	1.72
	R-squared		0.384	0.162	0.107

[1] Logarithm in base 2 of the Global Competitive Index as used in the reports to regress the GDP growth

[2] na: not available

Table 4.5: *List of complexity indexes with their explanation of economic growth.* The three complexity indexes are: the trophic depth, the economic complexity index (Hausmann et al., 2011), and the global competitive index (Schwab and Sala-i Martin, 2014). The values are the average between 1995 and 2009. The R-squared is the amount of variance explained when regressing the economic growth against each complexity indexes independently.

It can be explained by the fact that it is only one for which a linear relation between GDP growth and trophic depth is predicted. One possible reason of the difference between trophic depth and the Economic Complexity Index could be related to the interpretation of the services sector. It is discarded when computing the Economic Complexity Index as no data are available to include it. This can introduce a large bias as it accounts for a large part of the GDP composition of developed economies.

4.11 Conclusion

In this chapter, we used empirical data to test all the predictions made in the previous chapter. We emphasised the relation between trophic depth and economic growth and we tested our model after controlling for other parameters. In all these cases, the trophic depth is an important variable and always remains significant. This finding is essential as we prove that the depth of the production chain is a key parameter when comparing the economic growth of countries. According to our model countries with a small output multiplier will experience low economic growth. On the other hand, if the trophic depth is high the economic growth will be significantly higher as local technological change gets more amplified. In addition, we demonstrated that the trophic depth is also predictive of future growth. Finally, we compared the trophic depth of countries with two other indexes used in complexity economics and noted that the trophic depth explains more variance.

This chapter summarises some of the main results of the thesis that trophic depth is a key parameter when comparing the economic growth of countries. Despite the fact that the model is relatively simple as we only considered one particular type of technological progress and assumed that it is the same across all countries and all industries, our results are really robust and we managed to illustrate the role of the production chain. The framework we have developed is promising as it provides a

new approach to understand other key drivers of economics growth. In future work we would need to include other important effect such as capital investment in the framework we developed.

Appendix

4.A Network representation of the trophic structure of countries in 2009

On the vertical axis, we position each industry according to its country trophic level and on the horizontal axis industries are ranked based on the ISIC industry classification (revision 3); see Table 4.2 for the meaning of the industry codes. The trophic structure of the USA and China differ slightly from Figure 4.3 as we now only plot the year 2009 while previously it was the average trophic structure over the 14 years. Only the top 100 links are highlighted for clarity.

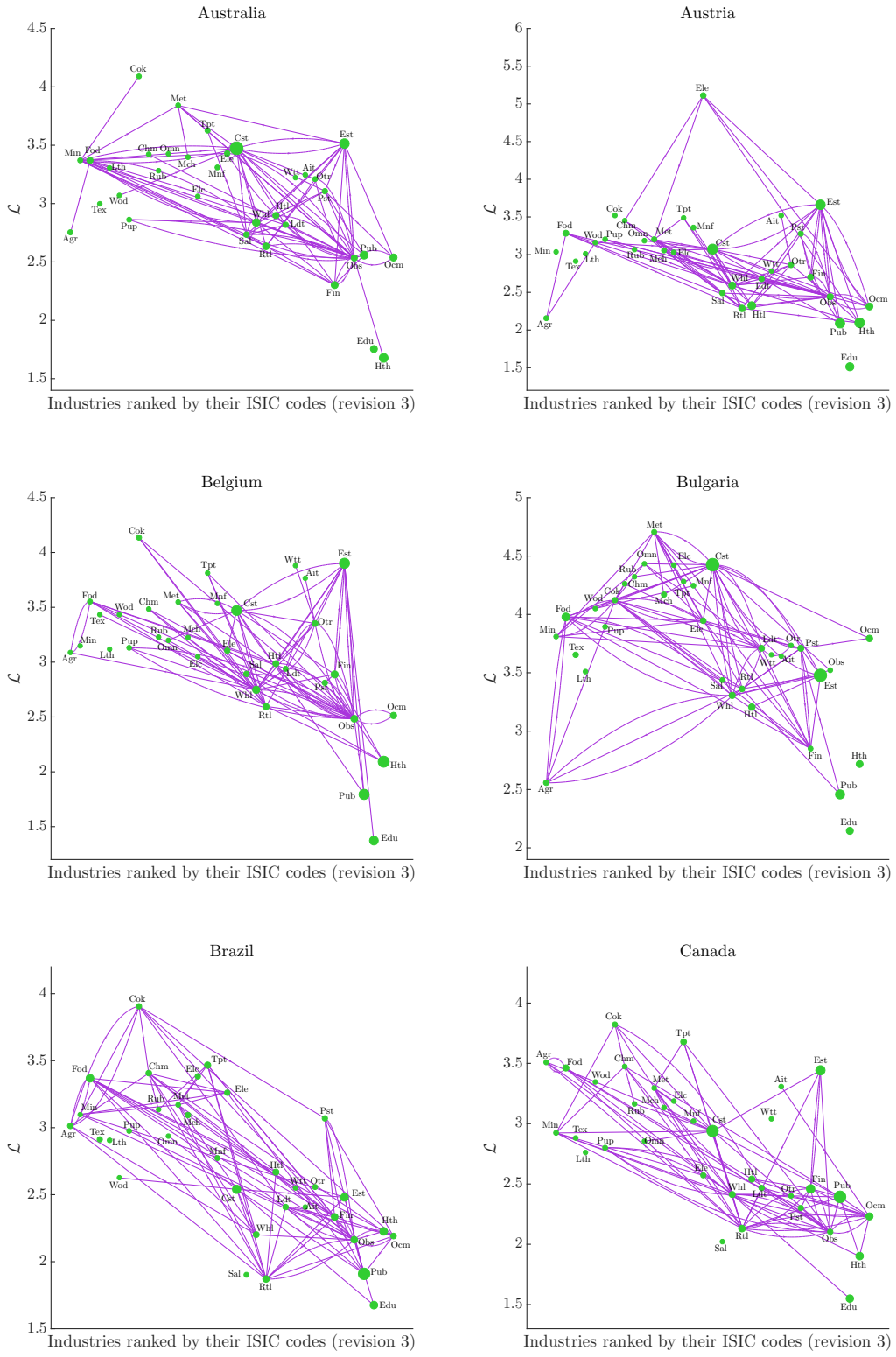


Figure 4.9: *Trophic structure of Australia, Austria, Belgium, Bulgaria, Brazil, and Canada.*

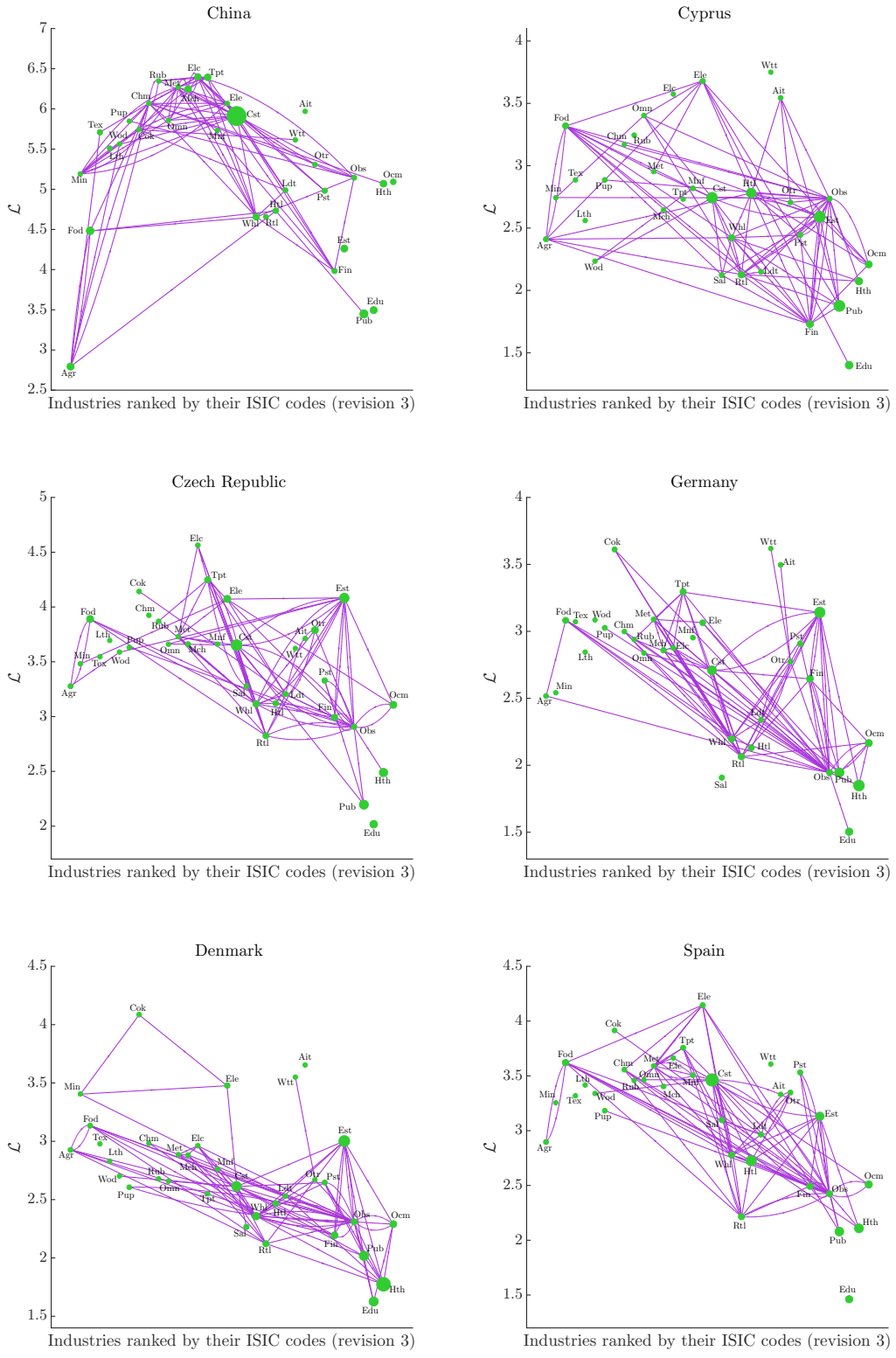


Figure 4.10: *Trophic structure of China, Cyprus, Czech Republic, Germany, Denmark, and Spain.*

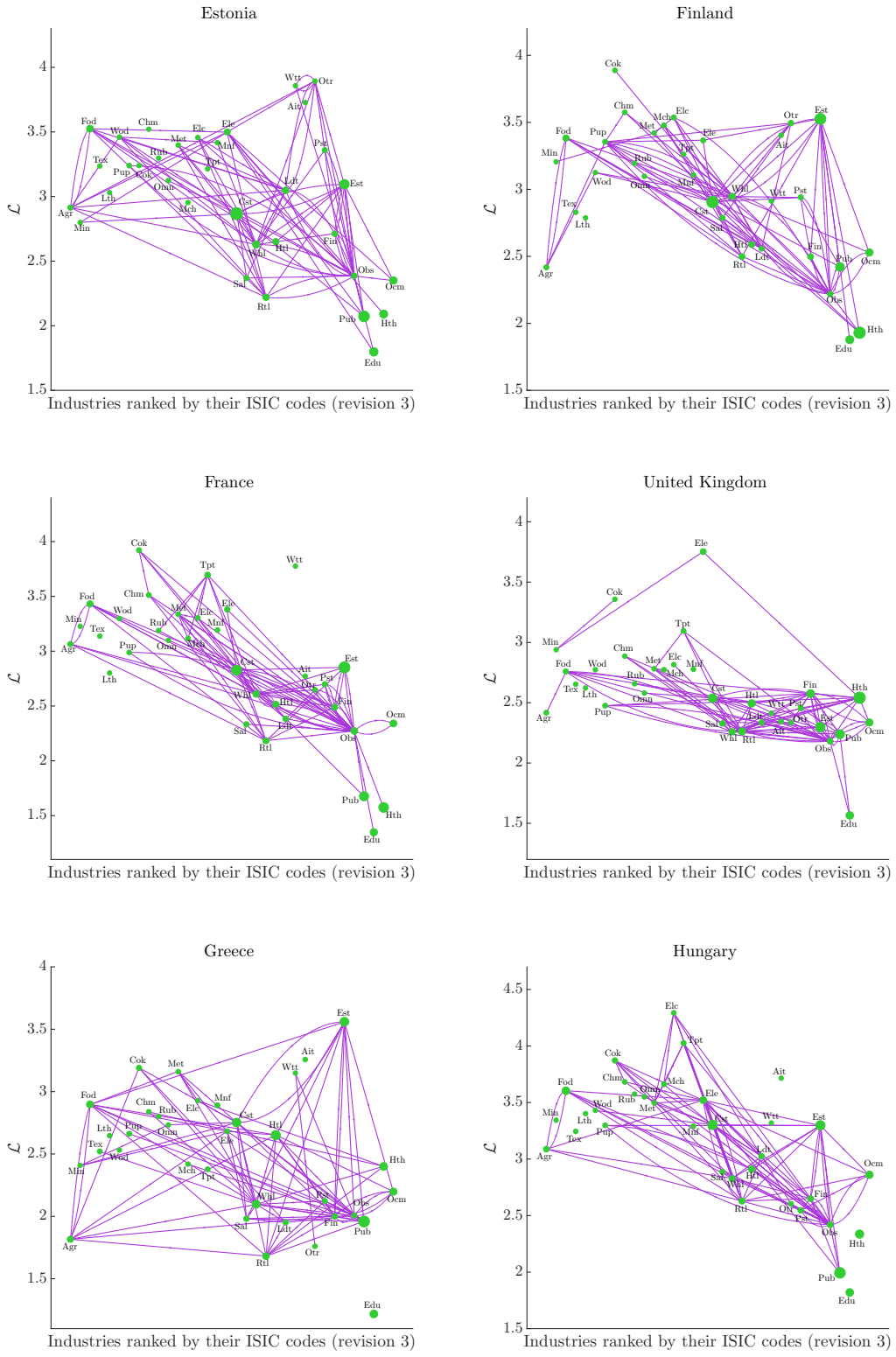


Figure 4.11: *Trophic structure of Estonia, Finland, France, United Kingdom, Greece, and Hungary.*

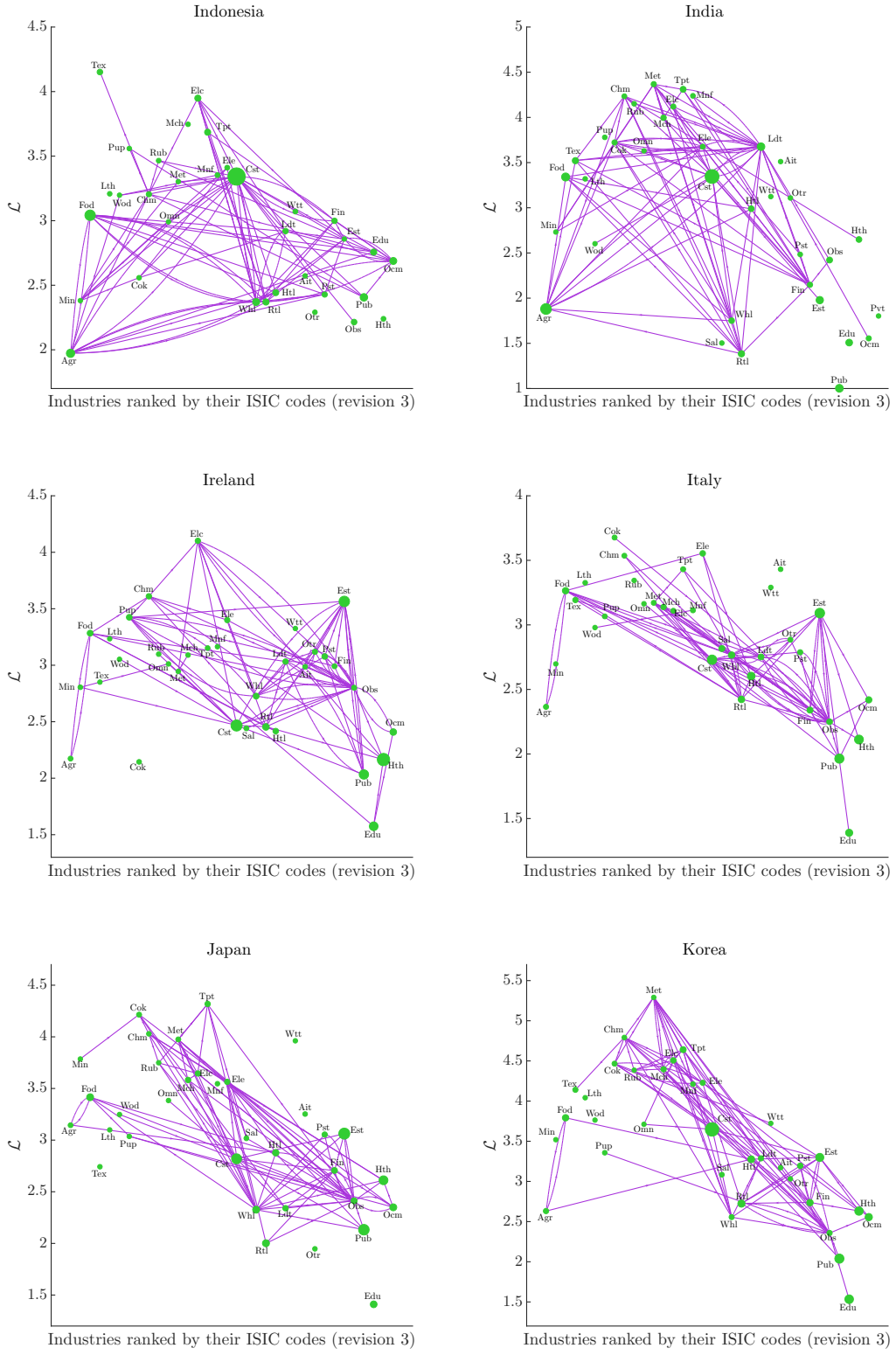


Figure 4.12: *Trophic structure of Indonesia, India, Ireland, Italy, Japan, and Korea.*

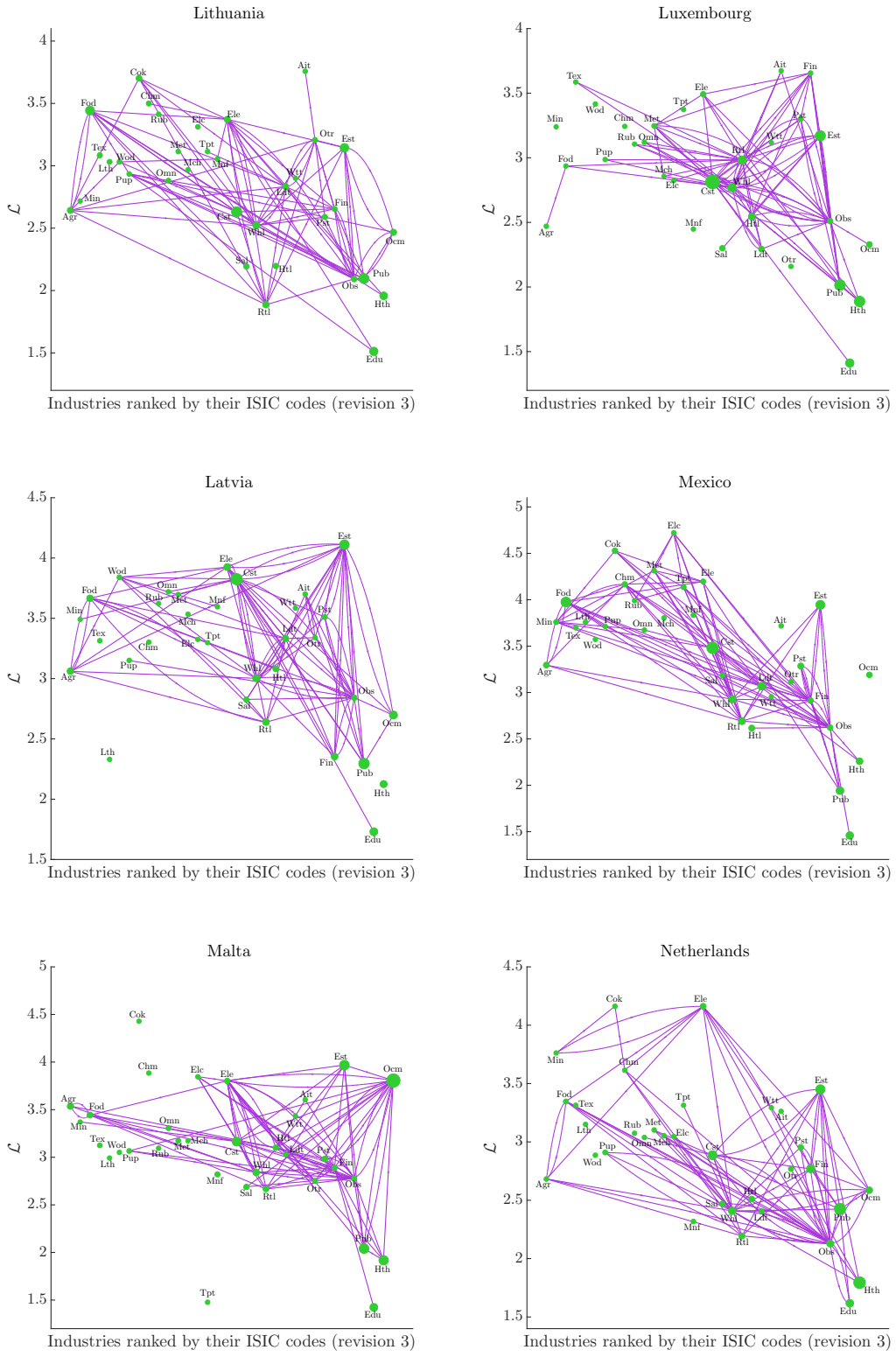


Figure 4.13: *Trophic structure of Lithuania, Luxembourg, Latvia, Mexico, Malta, and Netherlands.*

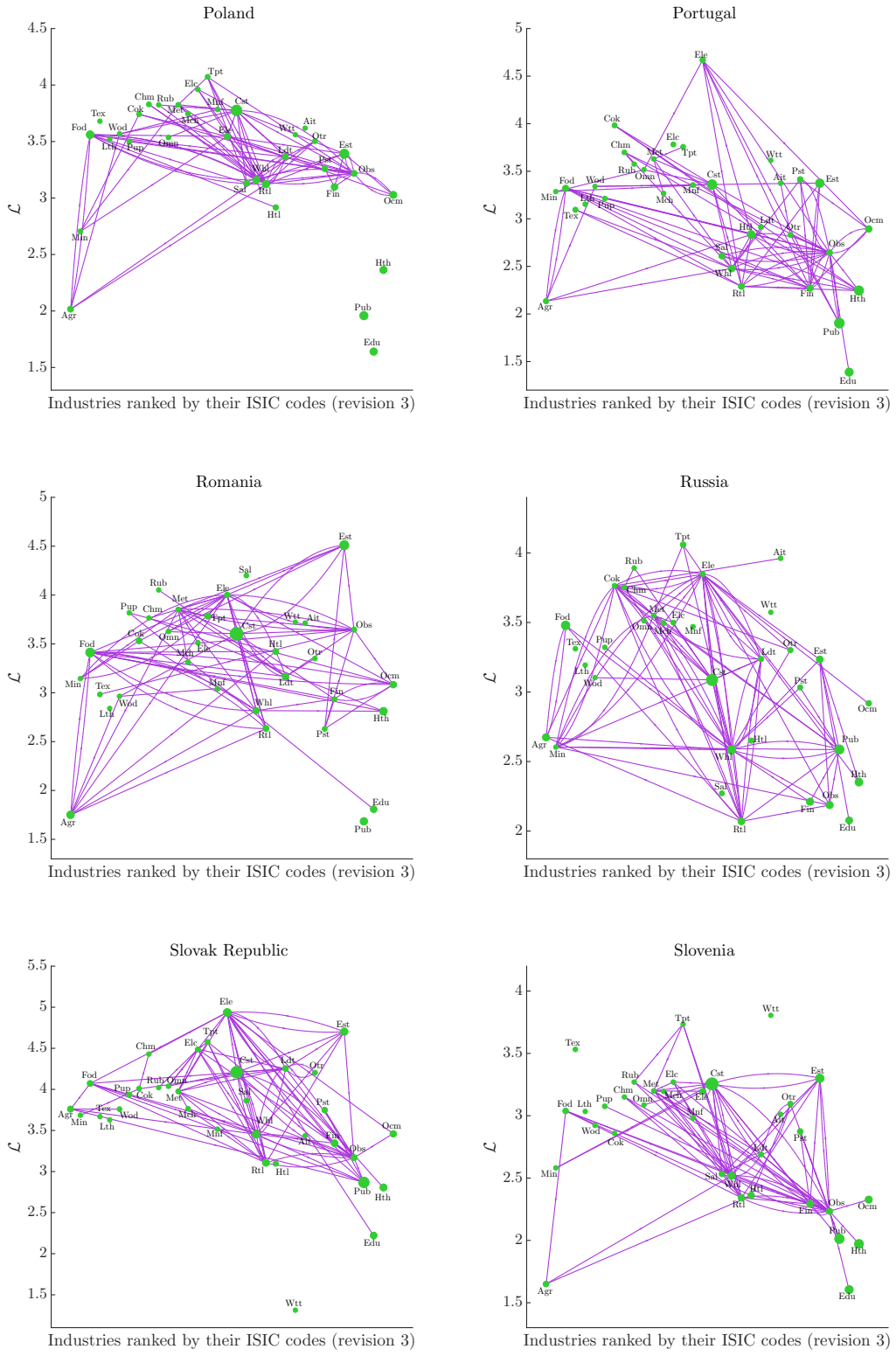


Figure 4.14: *Trophic structure of Poland, Portugal, Romania, Russia, Slovakia, and Slovenia.*

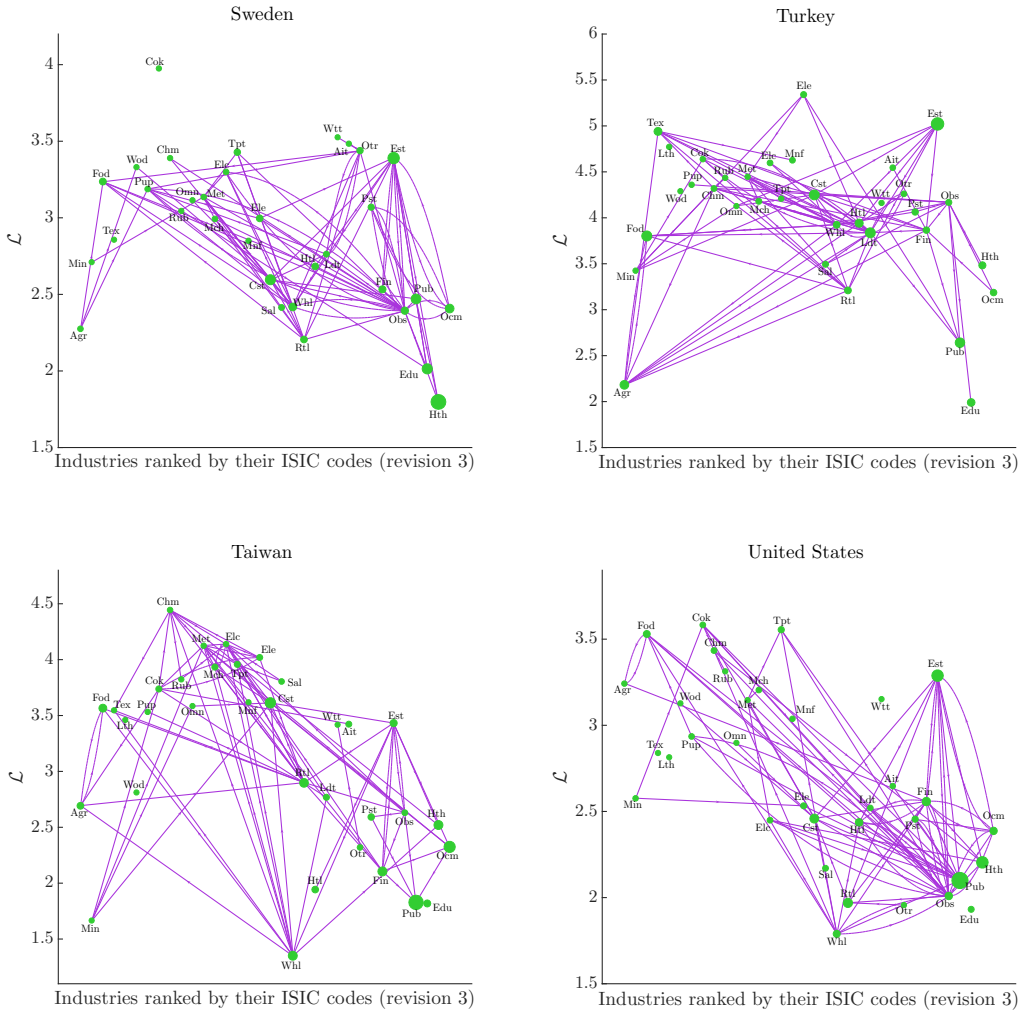


Figure 4.15: *Trophic structure of Sweden, Turkey, Taiwan, and the USA.*

Chapter 5

Structural change and trophic depth

5.1 Introduction

In the previous chapter, we highlighted the role of the trophic depth as a key driver to explain the difference of growth rates across countries. We also noted that developed economies have similar trophic depths while emerging countries have their trophic depths spread over a large range of values. Along the development path, the rate of economic growth accelerates and decelerates (Hausmann et al., 2004; Pritchett, 2000). One question we would like to answer is whether the trophic depth can explain some long run trend of acceleration and deceleration of growth. This would meant that the amplification of technological shocks by the production chain network varies with the country's development. If it explains some acceleration and deceleration of growth, we would like to know whether the dynamic of trophic depth along the development is predictable and whether we can identify any patterns.

Those questions are related to the field of structural change of the economy. Kuznets (1973) stated that due to technological change and the large variety of impacts it has on various sectors, the depth of the economy changes rapidly. Several mechanisms have been identified based on the supply or demand sides, or both. The common idea between them is that the evolution happens at the micro level and then

propagate to the macro level. Krüger (2008) realised a complete review of the different approaches to study structural change. He identified four categories: the 3-sector growth models, the multi-sector growth models, the evolutionary theories, and the reallocation approach. The first approach consists in modelling the 3 main sectors (agriculture, manufacturing, and services) and studying the evolution of the share of each sector. On the other hand, the multi-sector growth approach does not restrict the study to 3 sectors but instead considers a continuum of sectors. In the evolutionary approach, the economy is considered as being far from equilibrium and the economic agents are largely heterogeneous. In this case, analytical solutions do not exist but instead solutions are obtained via simulations. Finally, the last approach is about modelling the economy at the firm level to study the shift in the allocation of resources.

In this chapter, I will focus on the evolution of trophic depth at different stages of economic development. Our approach is in the same vein as the evolutionary structural change method. We assume that there exists an underlying mechanism that explains the evolution of the trophic depth. While we cannot model it at the micro level we will use historical data to exhibit an empirical model for the dynamic of the trophic depth at the macro level. I show that there exist two potential growth paths. Along the first growth path, a country's trophic depth remains in the same range while along the second growth path the trophic depth increases first and then declines once a certain level of wealth is reached. I also emphasize that developed economies followed an arch-shape growth path that leads to growth acceleration first and then deceleration. Emerging countries that are following a different path, do not experience large growth rate in the long run and are not likely to catch up the most wealthy economies.

5.2 Intuitions

First, I present some intuitions of what the trophic depth could be along the growth path and how the output multiplier of the USA can be inferred over the last two centuries.

5.2.1 Assumption

Gross domestic production can be computed in three ways: the production approach, the income approach or the expenditure approach. In theory, each method gives the same result when aggregated as long as the accounting is done correctly. However, once we want to go down to the GDP share of an industry, using one approach instead of another is no longer equivalent. For instance, let us consider two industries with industry 1 producing only an intermediate good and industry 2 producing only a final good. Each of them hires the same amount of labour. In this case, the GDP share of industry 1 and of industry 2 based on the income approach is the same for both of them and is equal to 0.5. But if we now compute the GDP share based on the expenditure approach, we no longer get the same result. The latter would assign a GDP share of 0 to industry 1 and of 1 to industry 2.[1] This basic example highlights the problem arising once we want to compute the GDP shares of sectors.

In the model derived earlier, it was explicit which one has to be chosen: the expenditure approach. Unfortunately, long term time series or cross-country data on GDP shares are mainly done through the income approach. In the following, when needed, we will assume that the GDP shares computed via the expenditure or the income approaches are equivalent. This is a rather crude assumption but at a high level of aggregation we expect it not to change much. To test the assumption, we compute the trophic depths of countries in the WIOD using both approaches and we

[1] When using the expenditure approach, the GDP share of an industry is given by the household final consumption for this product and by definition of an intermediate good its GDP share will then be 0.

find that the correlation coefficient is equal to 0.93. This high correlation shows the assumption is legitimate.

5.2.2 Estimation of the trophic depth

We now estimate the shape of the growth path for countries that went through industrialization. One would guess that as the level of GDP per capita increases, trophic depth evolves. In the early stage of economic development, agriculture was the only economic sector. The only input needed in traditional agriculture is labour. Thus, the output multiplier is equal to 1, meaning that the growth rate of the economy is simply equal to labour efficiency. This is fully consistent with the approach of the Malthusian stagnation. Before the Industrial Revolution, despite the existence of some manufacturing industries, we expect the trophic depth not to have radically changed over a long time, remaining close to 1.

At the beginning of the Industrial Revolution, some technologies become viable, allowing two effects to happen. The first effect is that new technologies are spread to existing industries such that labour is no longer the main input. Thus, their trophic level is increasing and technological change gets amplified. The second effect is that more and more people are moving to the manufacturing sector, allowing a long run of technological progress. Finally, the more industrialized the economy becomes, the larger its trophic depth is. According to our model, if the rate of technological change remains constant but the trophic depth is increasing, then we predict that the economic growth is accelerating, leading to super-exponential growth.

Developed economies entered into a new phase where there is a shift from manufacturing towards services. By definition, for the service sector labour is the largest input needed, suggesting that the trophic level of the service sector should be much smaller than that of the manufacturing industry. Assuming that technological progress remains the same, the model predicts that growth rates of countries should slow down

and then stagnate when the long run balance between manufacturing and services is found. The argument that the rate of improvement remains the same is quite unrealistic, as while technological progress for some technologies is around 10% per year, the increase rate of labour productivity in developed economies is closer to 1%, amplifying the economic slowdown of advanced countries.

With this simple approach, we expect countries that have had an industrial revolution to experience rapid increases in their trophic depth followed by a decrease at the end of the period due to a shift to services.

5.2.3 Application to the USA

In Section 4.9, we emphasized that two variables (the fraction of expenditures paid to employee compensation and the GDP composition of countries) can approximate the trophic depth of countries. Using this information, the past trophic depth of the American economy will be inferred. The first assumption is that the universal trophic level of an industry is a notion that does not depend on time. Because it was defined across countries at different stages of development this hypothesis is reasonable at the first approximation but is quite crude. In the previous chapter, we highlighted some important differences of the trophic level of the agriculture sector between the USA and China. We related that to the fact that in China agriculture is still relatively traditional compared to the USA where it is highly mechanized. We believe it will have no impact on the trend but only on the amplitude of the trophic depth.

The estimated trophic depth is computed as follows:

$$\bar{\mathcal{L}}_{est}(t) = \alpha(1 - \text{Labour Index}(t)) \sum_i \theta_i(t) \mathcal{L}_i^I, \tag{5.1}$$

where α is a scaling parameter such that the average trophic depth computed with this method is the same as the average trophic depth computed directly from the

WIOD over the period from 1995 to 2009. The labour index adjusts for the change of cost of labour compared to the cost of capital. This term controls for the fact that all else equal the trophic depth will increase if the cost of labour decreases.

The proportions of GDP by sector of origin are given by the BEA from 2013 to 1947 and then by International Historical Statistics (2013) for the period from 1939 to 1869. The remaining GDP shares from 1869 to 1790 were estimated by regressing the trend over the following century.

It is also necessary to adjust both for the labour share and the evolution of wages compared to the inflation to capture the change in labour costs. The labour share from 1947 to 2014 is given by the US Bureau of Labor Statistics (2015) and then is extrapolated using a backward regression. The employee compensation and the inflation rate are available from 1790 to 2014 (Officer and Williamson, 2013, 2015).

Combining all these data, Figure 5.1 shows the trophic depth of the USA inferred back to 1790.

The estimated growth path has a characteristic arch-shape. From 1790 to 1900, the trophic depth increased and peaked in the early 20th century. Then, it declines by nearly 40%. This observation emphasizes what was mentioned earlier that along the development path the trophic depth varies. At the beginning of the development, the trophic depth is low but is increasing. Then, when a certain level of wealth is reached, the trophic depth decreases. This also confirms that the network depth amplifies shocks more during the Industrial Revolution compared to the second half of the 20th century. This provides an explanation for the economic slowdown of developed economies. The further we go in the past, the less confidence we have in our predictions. Nonetheless, the observed shape is aligned with our intuition.[2]

[2]We have tried to perform the same analysis for Korea, Japan, and the United Kingdom but we did not find data over a period that is long enough to observe any change in trend.

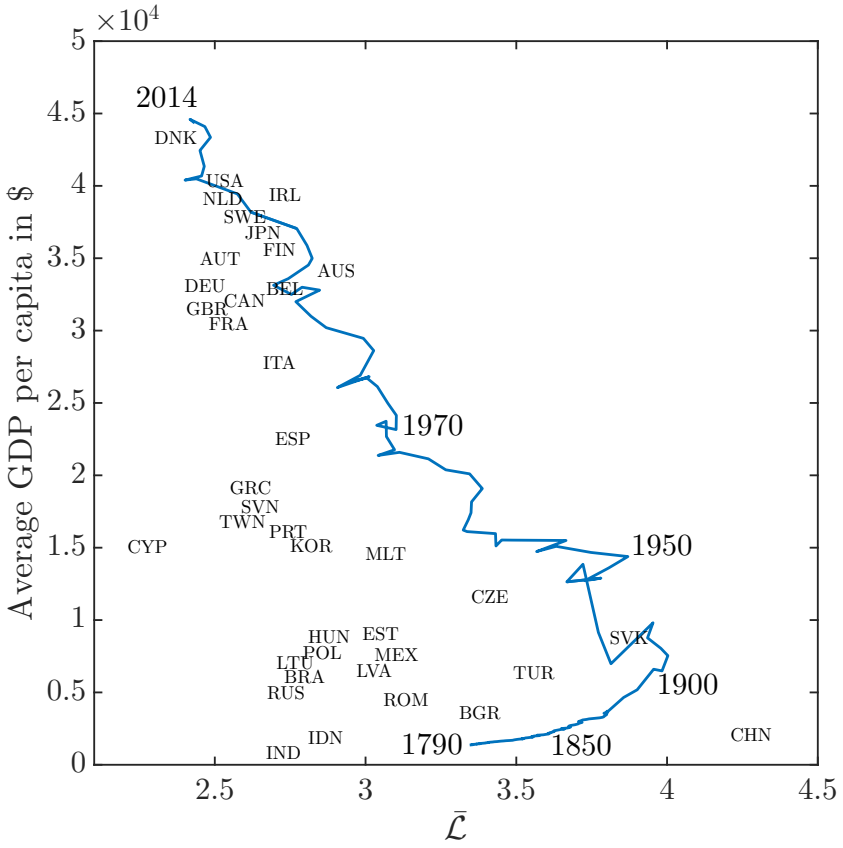


Figure 5.1: *Estimated growth path of the USA from 1790 to 2014.* The GDP per capita is given by Bolt and Van Zanden (2014). The continuous line is the estimated path for the USA. We also indicates the average position of the 40 countries in the WIOD (Figure 4.2).

5.3 Growth path analysis

In the previous section, I presented the intuitions of what the trophic depth can be along the growth path of countries. I also demonstrated that the intuition is supported by the estimation of the USA trophic depth from 1790 to 2014. In this section, we analyse the trajectory of the 40 countries in the WIOD from 1995 to 2009.[3] Despite of

[3]We are aware that there exists another database called EORA (Lenzen et al., 2012) that covers 187 countries over more than 40 years. However, when computing the trophic depth over the same period and the same countries as in the WIOD we find no correlation between the two. We also observe that the trophic depth of countries can fluctuate by order of magnitudes in EORA, highlighting some quality issues in this database.

the really short period we expect that some countries will highlight different sections of the arch-shape growth path.

5.3.1 General observation

Figure 5.2 summarizes the trajectories of all countries in the trophic depth-income plane from 1995 to 2009. We note that developed countries are all localised in the same region of the phase-space while emerging economies can have significantly different trajectories. In the following, we comment on each category of countries. The categories are arbitrarily defined by looking at their level of income and the expected shape of their trajectories. Despite the lack of resource-rich economies, we would expect these countries to be located in the upper-right region associated to high income per capita but with also relatively high trophic depth. For some countries, we observe a significant decrease at the end of the dataset and it is likely to be a consequence of the 2008 financial crisis.

5.3.2 Developed economies

Developed economies represented in Figure 5.3.a have a clear linear upward trend and countries' trophic depth remains between 2 and 2.5. This emphasizes that those economies have reached a steady state where trophic depth is constant. The fluctuations of growth rate are either due to the fluctuations of the technological progress or to other mechanisms not represented here.

5.3.3 Constant trophic depth growth path

We observe a second type of economies that have trophic depths between 2.5 and 3 and are represented in Figure 5.3.b. Their trophic depth is slightly larger than developed economies but still comparable. Over the period from 1995 to 2009, their trophic

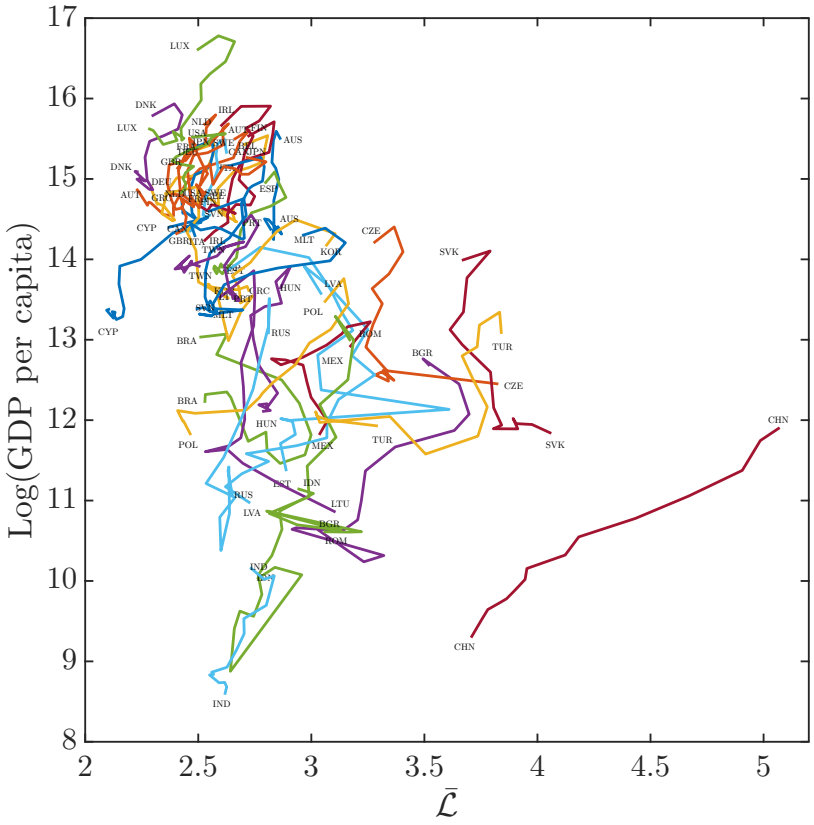
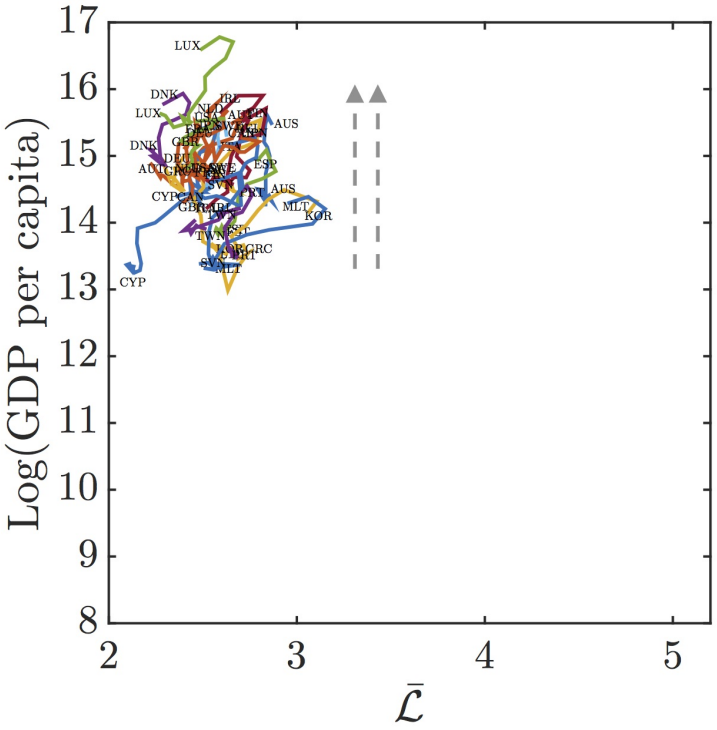


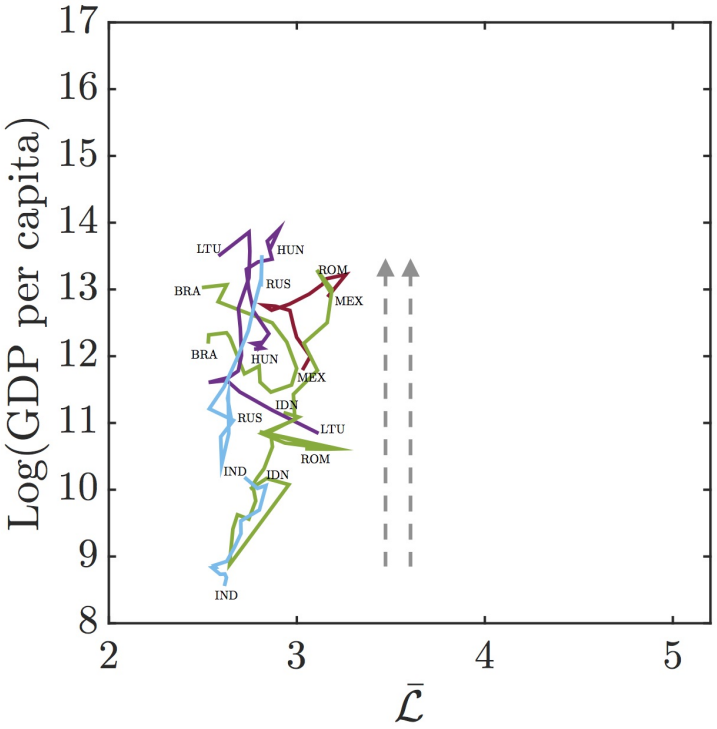
Figure 5.2: *Growth paths of the countries listed in the WIOD from 1995 to 2009.* We show all the 40 countries. Country codes are listed in Table 4.1.

depth remains in the same range despite some large fluctuations. Thus, according to the model, their growth rates would be comparable to developed economies. Hence, they will never catch up to developed economies but instead will fall behind.

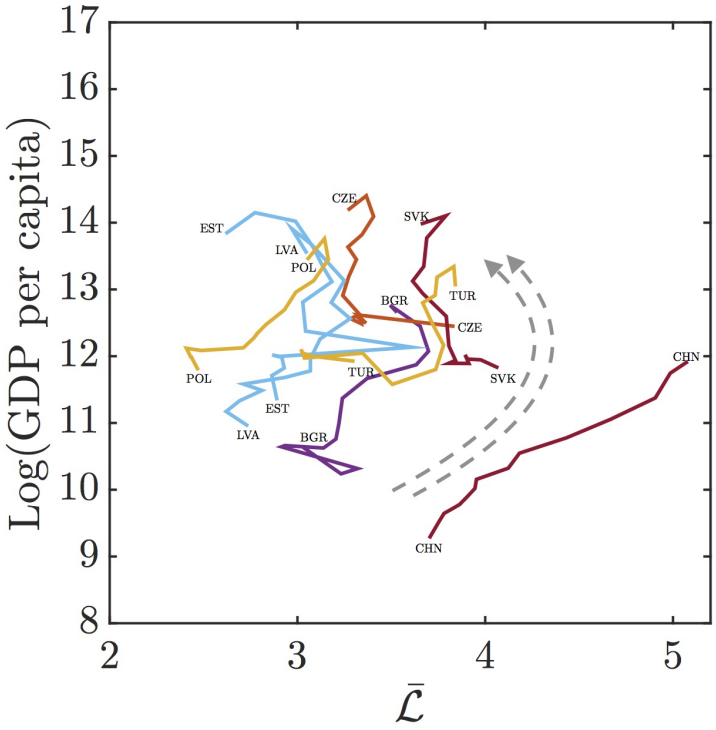
However, two countries might still follow a different trajectory as they are still at an early stage of development. These economies are India (IND) and Indonesia (IDN) and they still have the possibility to follow a different path. We will come back to this point in the second part of this chapter as it is a crucial question concerning their economic development and whether we can predict their future evolution.


Log(GDP per capita)
17
16
15
14
13
12
11
10
9
8
2
3
4
5
$\bar{\mathcal{L}}$
LUX
DNK
AUS
ESP
KOR
MLT
CYP
GRC

(a)


Log(GDP per capita)
17
16
15
14
13
12
11
10
9
8
2
3
4
5
$\bar{\mathcal{L}}$
LTU
HUN
ROM
RUS
BRA
MEX
IDN
IND

(b)



(c)

Figure 5.3: *Hypothesized growth trajectories of the countries listed in the WIOD from 1995 to 2009.* The arrows in grey show the hypothesized trajectory for each subset of countries. a) Trajectories of developed economies; b) Trajectories of developing economies following a growth path with no change of trophic depth; and c) Trajectories of developing economies following a potential arch-shape path. Country codes are listed in Table 4.1.

5.3.4 Potential arch-shape growth path

Finally, in the last group, we expect countries to follow an arch-shape trajectory in the phase-space and are represented in Figure 5.3.c. First, their trophic depth increases and according to our model they should experience some super-exponential growth then after reaching a certain level of wealth the trophic depth decreases. Estonia (EST), Latvia (LVA), and Bulgaria (BGR) are the only countries for which we can observe the climax of the trajectory; for the others like China (CHN) we will expect it in the future and for Slovakia (SVK) we expected in the past. For most countries, the trophic depth highest value is reached when the logarithm of GDP

per capita is between 11.5 and 12.5. It seems that this is a turning point for the economic growth path where super-exponential growth slows down and finally the trophic depth converges to a value around 2.5 that seems to be a convergence point. There is an interesting fact about the position of China because at the end of the dataset it reaches the critical wealth per capita where trophic depth starts to reduce and according to the model its economic growth should mechanically slow down.

5.3.5 Discussion

We note that the dynamic of countries in the phase-space given by the trophic depth and the log of GDP per capita is quite complex as there are several potential growth paths. This description is useful as it provides a new tool to compare the potential of growth of countries and at the same time we can provide a classification of countries influenced by the evolution of their trophic depth. For instance the BRIC countries can be split into subcategories.[4] Brazil and Russia would be in the same one as they both follow the similar growth path while China has a completely different one. India is still at an early stage so we cannot classify it yet.

The case of China is particularly interesting given the behaviour of other countries that followed a similar path and the trajectory of the USA that we inferred. To comment the future evolution of China, we need to develop a model of structural change.

5.4 Model selection analysis

In the previous sections, we highlighted the existence of a complex non-linear relation between the logarithm of GDP per capita and the trophic depth. One difficulty comes from the relatively short period when input-output tables are available. Thus,

[4]The BRIC countries are Brazil, Russia, India, and China. It is a common classification in economics.

we only observe incomplete development trajectories. The method introduced to infer the past trophic depth of the USA is not satisfying as well because it would require data across many countries and over a long time period. Most of these types of data are not available.

In the following section, we want to use the historical data from the WIOD and do some statistics to conduct an empirical model selection to identify the model with the most explanatory power but including the minimal number of terms by penalising for the addition of an extra term. Thus, we would be able to perform some projections on the future trajectories of countries and compare our predictions with global observations made previously about the arch-shape growth path of countries. The role of this approach is to capture the complex and non-linear relation between GDP per capita and trophic depth. We also want to analyse the case of China and whether the model can predict any deceleration of growth and its amplitude. This approach is in the same vein as Cristelli et al. (2015) and Ranganathan et al. (2014). Cristelli et al. (2015) defined a complexity index called the Fitness that is closely related to the Economic Complexity Index mentioned in 4.10. They look at the trajectories and position of countries in the phase-space drawn by the level of GDP per capita and the Fitness of countries. Then, using the theory of dynamical systems, they defined a selective predictability scheme to infer the evolution of countries in this phase-space. This approach is not optimal for what we want to do as it requires data over the entire phase-space. In the case of Ranganathan et al. (2014), they study the relation between democracy and level of wealth.

We will first describe our approach and compare it to a Bayesian model selection used in Ranganathan et al. (2014). The only difference between the two approaches is the criteria used to compare models. Then, we will use the model to estimate future trajectories of countries.

5.4.1 Method

We now describe the method used to estimate the non-linear relation between Log(GDP per capita) and the trophic depth.

5.4.1.1 Our approach

We assume that the dynamic of countries in the phase-space is two dimensional, described by Log(GDP per capita) and the trophic depth, $\bar{\mathcal{L}}$. Their joint dynamic is described by a set of two equations:

$$\frac{d\bar{\mathcal{L}}}{dt} = f_1(\bar{\mathcal{L}}, Log(GDP)), \tag{5.2}$$

$$\frac{dLog(GDP)}{dt} = f_2(\bar{\mathcal{L}}, Log(GDP)), \tag{5.3}$$

where f_1 and f_2 are two non-linear functions. Estimating the model consists of finding the best functions using empirical observations. To estimate these functions, we approximate them with a polynomial form

$$f_i(\bar{\mathcal{L}}, Log(GDP)) = \sum_{j=-N_i}^{N_i} \sum_{k=-N_i}^{N_i} a_{j,k} \bar{\mathcal{L}}^j Log(GDP)^k. \tag{5.4}$$

The number of terms, noted m, is the number of non-zero coefficients. Solving the problem consists of finding N_i and the number of terms that best fit the data. For each combination, we perform a multiple linear regression to compute the coefficients. We arbitrarily choose the potential terms.

The second step consists of defining a criterion to select the best model. The adjusted R-squared is a good candidate to compare models with different numbers of terms (Theil, 1961) and is equal to

$$R^2_{adj} = 1 - \frac{n-1}{n-p} \frac{SS_{res}}{SS_{tot}}, \tag{5.5}$$

with SS_{res} the sum of squares of residuals and SS_{tot} the total sum of squares. The adjusted R-squared increases only when the additional explanatory variable increases the coefficient of determination, R^2, more than what it would have been by chance. This criterion is the one we pick to compare models with various number of terms.[5]

Our approach is different from a stepwise regression procedure. Given the number of terms specified, m, we estimate all the possible models while in a stepwise regression procedure, the algorithm starts with the best model with $m-1$ terms and chooses the best additional term. In a stepwise framework, the order in which you introduce terms matters while we do not have this problem in our approach.

5.4.1.2 Comparison with Bayesian model selection

Another approach often used to perform some model selection analysis is the Bayesian selection model; this is the one used by Ranganathan et al. (2014). In the following, we show that a Bayesian model selection with no prior knowledge of the distribution of the estimates is equivalent to the adjusted R-squared method.

Similarly to the R-squared, the log-likelihood value increases with the number of terms but the Bayesian marginal likelihood would introduce a penalty for any additional term that does not improve the model. Thus, it is similar to the adjusted R-squared. Nonetheless, the Bayesian approach is also comparing the estimates with some prior distributions. This has the advantage that in some field specialists might have some prior knowledge about the value of coefficients; thus the Bayesian approach would allow them to add this extra bit of information in the model selection procedure.

In some cases, prior knowledge does not exist or is not available. From a Bayesian perspective it can be interpreted as an non-informative prior distribution and implies that the prior distributions of the estimates and of the logarithm of the standard

[5]The adjusted R-squared is based on the coefficient of determination of a regression is $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$. The R-squared captures the share of explained variance by the model. It has the inconvenient feature of increasing with the number of predictor variables and therefore does not fit the requirement for model selection.

deviation of the errors are uniform distributions. Then, the explained variance given by a Bayesian selection model with non-informative prior distribution is (Gelman et al., 2003)

$$R^2_{Bayes} = 1 - \frac{n-3}{n-p-2}\frac{SS_{res}}{SS_{tot}}. \tag{5.6}$$

The adjusted R-squared and the Bayesian selection model with non-informative prior density are equivalent if $n \gg 1$ and $p \ll n$. Our data perfectly match those conditions as $n = 560$ and our maximum number of explanatory variables is 17. Thus, a Bayesian model selection and the adjusted R-squared method are equivalent for our analysis. The adjusted R-squared is not a direct measure of the predictability of a model. Later, we will introduce a cross-validation method to analyse the model predictability and we will find that the best model based on the adjusted R-squared is the same as the one given by the cross-validation method.

5.4.2 Application

We apply the procedure described earlier to model the dynamic of countries in the phase-space given by the Log(GDP per capita) versus the trophic depth.[6] Then, we will used it to perform some predictions.

[6]For concision, in the equations, we replace Log(GDP per capita) by Log(GDP).

5.4.2.1 Model

We first specify the polynomial function that we pick. In the following we use

$$\begin{aligned} f_i(\bar{\mathcal{L}}, Log(GDP)) &= a_{0,0} + a_{1,0}\bar{\mathcal{L}} + a_{0,1}Log(GDP) + a_{1,-1}\frac{\bar{\mathcal{L}}}{Log(GDP)} \\ &+a_{-1,1}\frac{Log(GDP)}{\bar{\mathcal{L}}} + \frac{a_{-1,0}}{\bar{\mathcal{L}}} + \frac{a_{0,-1}}{Log(GDP)} + a_{2,0}\bar{\mathcal{L}}^2 \\ &+a_{0,2}Log(GDP)^2 + \frac{a_{-1,-1}}{\bar{\mathcal{L}}Log(GDP)} + \frac{a_{-2,0}}{\bar{\mathcal{L}}^2} + \frac{a_{0,-2}}{Log(GDP)^2} \\ &+a_{1,1}\bar{\mathcal{L}}Log(GDP) + a_{3,0}\bar{\mathcal{L}}^3 + a_{0,3}Log(GDP)^3 \\ &+\frac{a_{-3,0}}{\bar{\mathcal{L}}^3} + \frac{a_{0,-3}}{Log(GDP)^3}. \end{aligned} \tag{5.7}$$

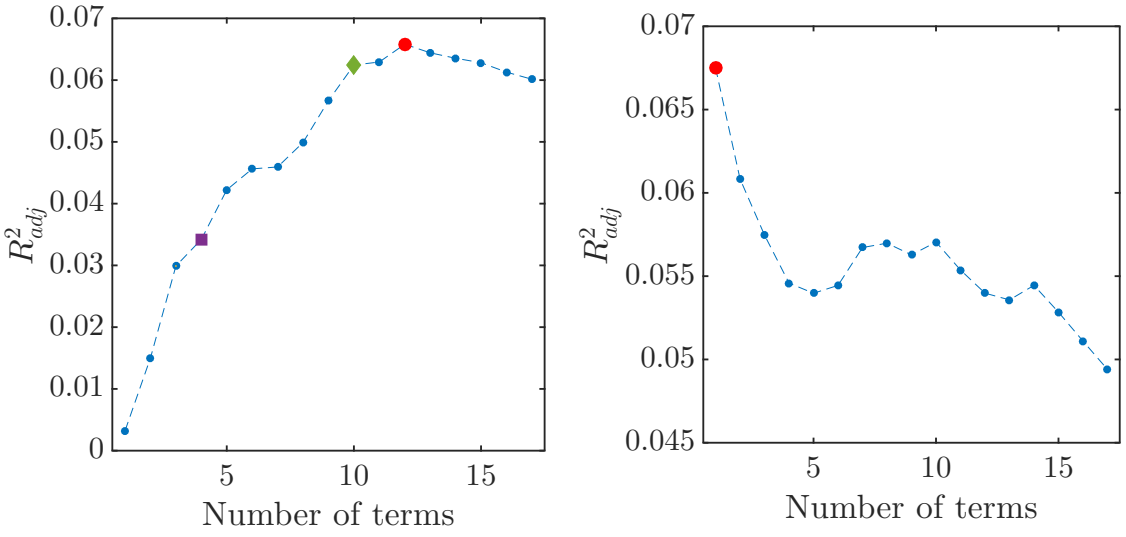
The previous equation highlights the potential 17 terms of the model. After we ran the model selection algorithm, we find that we need 10 terms for the dynamic of the trophic depth and 1 term for the dynamic of the Log(GDP per capita). The two equations are

$$\begin{aligned} f_1(\bar{\mathcal{L}}, Log(GDP)) &= 43.85 - 25.13Log(GDP) - \frac{250.61}{\bar{\mathcal{L}}} - 0.110Log(GDP)^2 \\ &+0.900\bar{\mathcal{L}}Log(GDP) - \frac{4590.20}{Log(GDP)^2} + \frac{1715.40}{\bar{\mathcal{L}}Log(GDP)} \\ &+172.52\frac{\bar{\mathcal{L}}}{Log(GDP)} + 9.06\frac{Log(GDP)}{\bar{\mathcal{L}}} \\ &+\frac{5767.70}{Log(GDP)^3}, \end{aligned} \tag{5.8}$$

$$f_2(\bar{\mathcal{L}}, Log(GDP)) = 3.57 \cdot 10^{-3}\bar{\mathcal{L}}^3. \tag{5.9}$$

As expected, the dynamic of the trophic depth is complex and the equation is not helpful to understand its dynamic. On the other hand, the dynamic of the Log(GDP per capita) is best explained by only one term that only depends on the trophic depth, meaning that its dynamic is independent of the the stage of development because it is independent of Log(GDP per capita). We now detail how we ended up with these two equations.

The adjusted R-squared as a function of the number of terms for the dynamic of the trophic depth and of the Log(GDP per capita) are shown in Figure 5.4.



(a) Trophic depth

(b) Log(GDP per capita)

Figure 5.4: *Model selection algorithm for the trophic depth and the Log(GDP per capita).* The adjusted R-squared, noted R^2_{adj}, is the variable used for the model selection. The best model is the one with the highest adjusted R-squared and is highlighted by a red dot. The green point indicates the model used to estimate the dynamic of the trophic depth. The purple point highlights a minimalist model.

It is clear that the dynamic of the Log(GDP per capita) is best explained by a model that only contains 1 term and any extra term does not increase the explanatory power of the model. On the other hand, the dynamic of the trophic depth is looking significantly more complex as the best model requires 12 terms. However, while the model with 12 terms is the best model according to the algorithm, we might also think that 10 terms is enough as from 10 terms to 12 terms the adjusted R-squared does not change much as it reaches a plateau.

To investigate this question, we plot a heatmap that shows the dynamic of the trophic depth in the phase-space for the best model associated with different number of terms, see Figure 5.5.

We now compare the heatmaps with 11 and 12 terms with the one related to

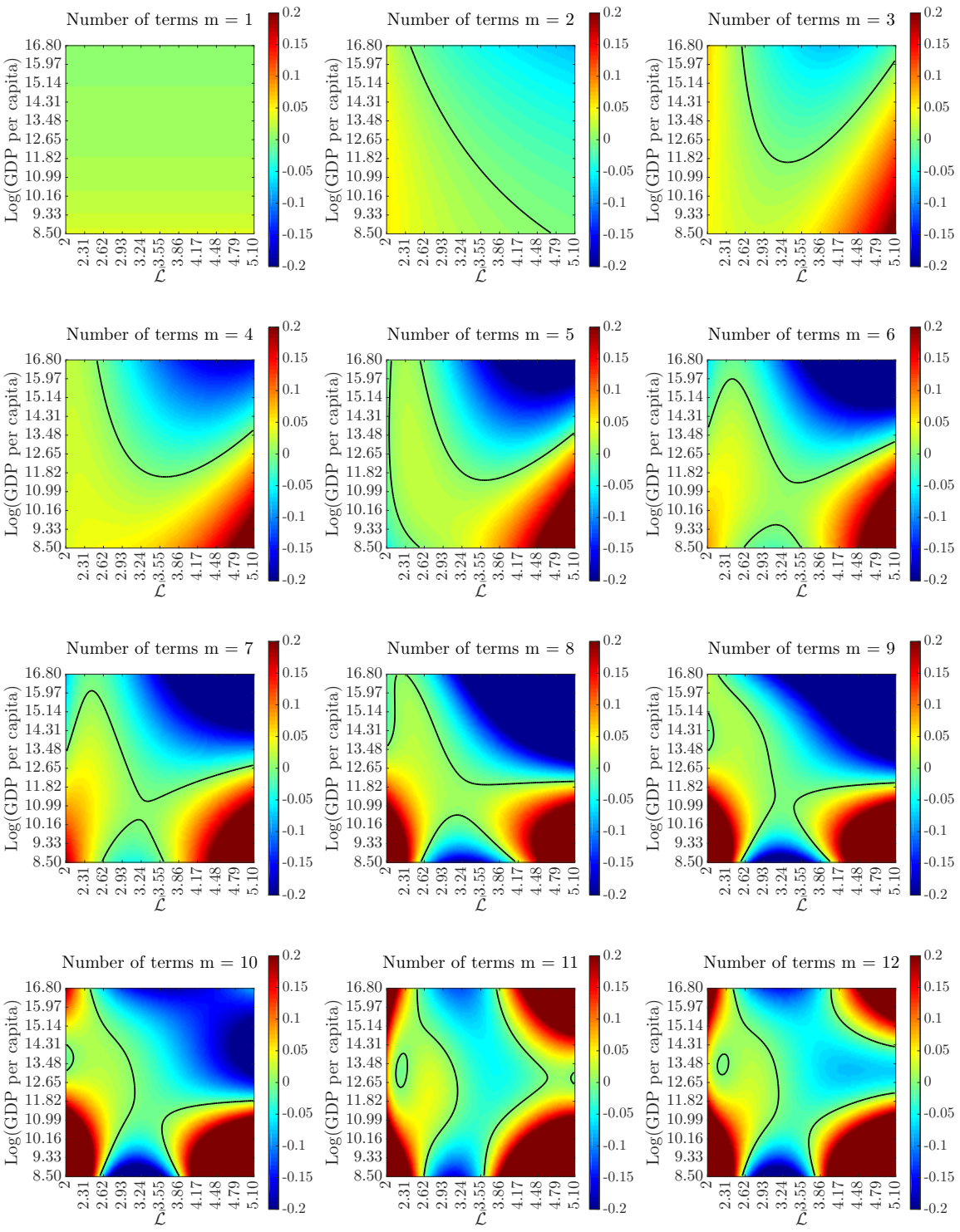


Figure 5.5: *Dynamic of the trophic depth for the best model associated with different number of terms.* We create the heatmap representing $f_1(\bar{\mathcal{L}}, Log(GDP))$ for different models with different terms. The colormap indicates the sign and the amplitude of $\frac{d\bar{\mathcal{L}}}{dt}$ at every point in space. The regions in red indicate that the trophic depth is increasing and regions in blue that the trophic depth is decreasing. The continuous black line represents the contour plot for $\frac{d\bar{\mathcal{L}}}{dt} = 0$.

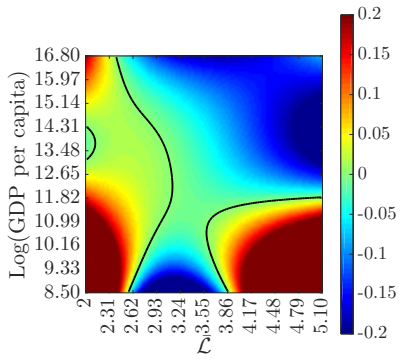
10 terms and we observe that the upper-right corner is the region that is the most affected. While with any model with 10 or less terms this region is associated to a negative growth rate, it becomes largely positive with 11 and 12 terms. As we do not have data in the upper-right corner, it suggests that the models with 11 and 12 terms are over fitting the data. Thus, the best model for the dynamic of the trophic depth is the one that contains 10 terms.

To understand the dynamic of the trophic depth in better detail, we focus on the heatmap that shows its dynamic. In Figure 5.6.a, we plot the dynamic of the trophic depth for a model with 10 terms. We also plot some theoretical trajectories to emphasize what it means in term of growth path in Figure 5.6.c. The trajectories are completely determined by the initial conditions.

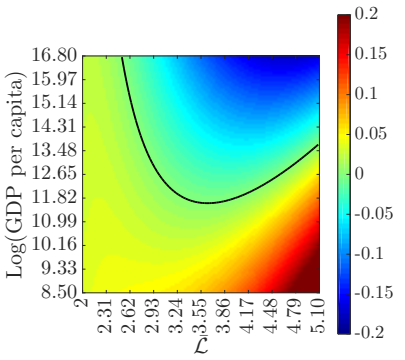
The first result is that with 10 terms we are able to replicate the fact that some countries will grow without any major change in trophic depth while others will have some radical changes. We also see that countries that follow an arch-shape growth path will develop faster than countries for which the trophic depth remains constant; see Figure 5.6.e.

The black lines on the heatmap represent two vertical isoclines and are defined as $\frac{d\bar{\mathcal{L}}}{dt} = 0$. The first growth path is associated with the black line starting at $\bar{\mathcal{L}} = 2.60$. If an economy deviates from it then the system will push it back to it. This development path is stable. One can notice that its trajectory is bent. Along its development, an economy on this growth path will increase its trophic depth up to 3.20 and once its level of Log(GDP per capita) is around 12 then the trophic depth decreases.

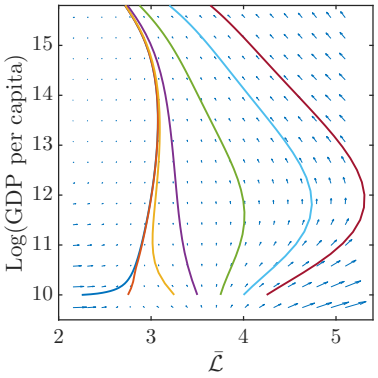
The second black line starting at $\bar{\mathcal{L}} = 3.90$ represents another vertical isocline but this one is unstable. Any small perturbation around it will push the economy away from the vertical isocline. According to our model, an economy on the right part of the growth path will experience a rapid and accelerating economic growth as $\frac{d\bar{\mathcal{L}}}{dt}$ is positive and once the economy moves on the left side of the vertical isocline it will



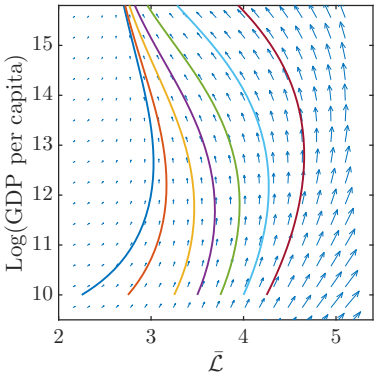
(a) Heatmap of $f_1(\bar{\mathcal{L}}, Log(GDP))$ with 10 terms.



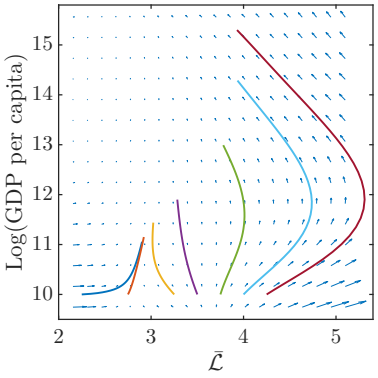
(b) Heatmap of $f_1(\bar{\mathcal{L}}, Log(GDP))$ with 4 terms.



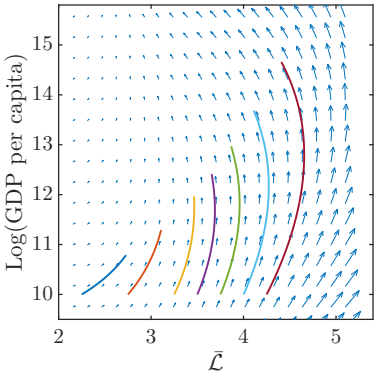
(c) Theoretical trajectories, $m = 10$



(d) Theoretical trajectories, $m = 4$



(e) Theoretical trajectories over 15 years, $m = 10$



(f) Theoretical trajectories over 15 years, $m = 4$

Figure 5.6: *Dynamic of the trophic depth for the best model with 10 and for a minimalist model with 4 terms.* The colormap indicates the sign and the amplitude of $\frac{d\bar{\mathcal{L}}}{dt}$. The regions in red indicates that the trophic depth is increasing and regions in blue that the trophic depth is decreasing. The continuous black line represents the contour plot for $\frac{d\bar{\mathcal{L}}}{dt} = 0$. In c) and d) we indicate 7 theoretical trajectories. For all of them, the initial Log(GDP per capita) is 10 and the initial trophic depths are: 2.25, 2.75, 3.25, 3.5, 3.75, 4, and 4.25. In e) and f) we show the same trajectories over a period of 15 years only.

return to the first growth path and its economy will slow down. The critical level of wealth is given by the flat portion of the second vertical isocline and is around 11.80. This means that once the Log(GDP per capita) is higher than 11.80 then the trophic depth of the country will slow down. This leads to the arch-shape trajectories observed in Figure 5.6.c.

The number of terms needed to describe the dynamic of the trophic depth dynamic is still relatively high so one might want to reduce it even more. We now focus on Figure 5.6.b and Figure 5.6.d. We now arbitrarily limit the number of terms at 4. The best model containing only 4 terms describing the dynamic of the trophic depth (Eq.5.8) is

$$\begin{aligned} f_1(\bar{\mathcal{L}}, Log(GDP)) &= 0.0986 Log(GDP) - 0.0262 \bar{\mathcal{L}} Log(GDP) + 0.0056 \bar{\mathcal{L}}^3 \\ &\quad -0.00952 \frac{Log(GDP)}{\bar{\mathcal{L}}}. \end{aligned} \tag{5.10}$$

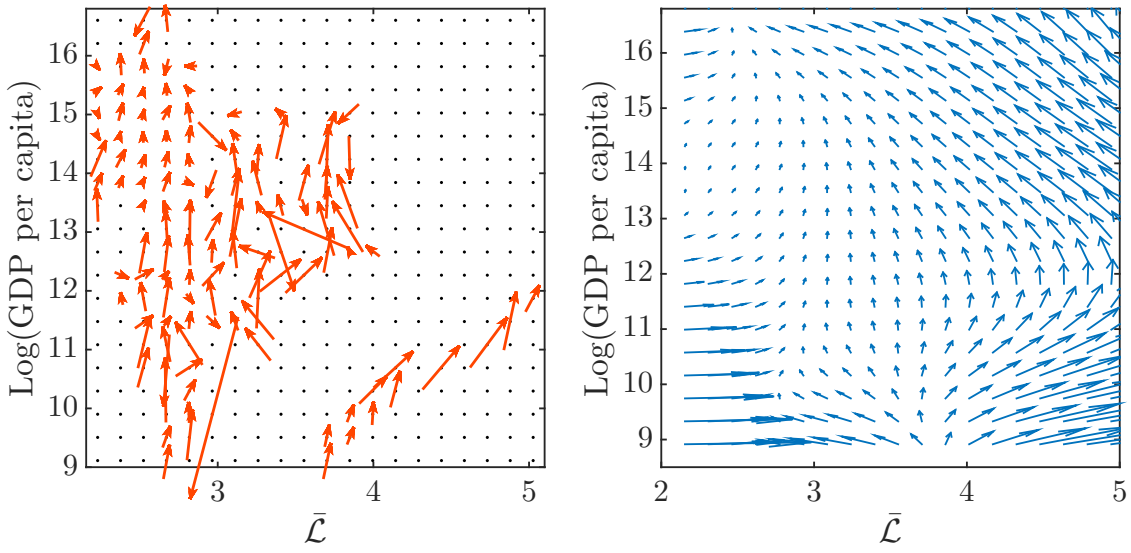
First, we analyse the stability of the vertical isocline given by the black line defined by $\frac{d\bar{\mathcal{L}}}{dt} = 0$. The left part of the curve represents a stable isocline, if an economy deviates from the isocline then the system brings it back. On the other hand, on the right section of the curve the isocline is unstable. To better understand the dynamics of countries similarly to Figure 5.6.c we plot the trajectories of some theoretical economies in Figure 5.6.d. The starting points are the same as before. The growth paths of countries are significantly different compared to the dynamic with 10 terms however they share some important common features. The most important observation is that countries with high trophic depth are developing at a much higher rate in both cases. In addition, in both cases, the trajectories of countries with high trophic depth are bent. The main difference between the two dynamics is the amplitude of the curvature of the bend; Figure 5.6.c is closer to what we observe in Figure 5.2.

There is one open question on how an economy moves from one growth path to

another. This seems strongly influenced by the early stage of development. According to Figure 5.6.a, there is a barrier represented by the blue region between the two paths at low level of GDP. With this approach we do not know if this is actually a barrier or if this is a consequence of the lack of data in this region. To understand why these could be a barrier, we refer to the work of McMillan and Rodrik (2011) on globalisation and structural change. They find that structural change has been growth reducing in Latin America. Their explanation is that labor is moving from industries with high productivity to industries with low productivity; in another words from manufacturing to services. In Asia, they observe the opposite effect. The mechanism they have highlighted here is linked to globalisation.

To summarise the new opportunities introduced by the model selection we compare the outcome of the model with another approach: the coarse-graining. Similarly to weather forecasts and the maps of the wind that describe the evolution of the air flows, we plot the dynamic of countries using a vector field in Figure 5.7.

The coarse-graining approach consists on dividing the phase-space into smaller regions and in each of them we aggregate the dynamic of countries. It requires some extra care about the level of aggregation at which countries are coarse-grained given that countries are heterogeneously spread. Here we divide the phase-space into 400 boxes and we check that the probability distribution functions of the dynamic along the two directions are preserved by the coarse-graining. One significant advantage of the model selection approach is that it allows us to estimate the dynamic at any point in space even in regions where we do not have any observations. On the other hand, one has to be careful on using the model selection in region with no data as over fitting issues are likely to occur. Our goal is to perform some out sampling predictions about the future trajectories of countries thus we need to know their dynamic in any point in space.



(a) Coarse-graining (b) Model selection algorithm

Figure 5.7: *Comparison between the model selection algorithm and the coarse-graining of trajectories.* The vector field represents the dynamic of countries in the phase-space Log(GDP per capita) versus the trophic depth. For the coarse-graining the phase-space is divided into 400 boxes. On the other hand, in the case of the model selection the solution of the problem is continuous so we can plot the dynamic at any scale.

5.4.2.2 Prediction analysis

Before making some predictions, we first want to check whether the selection criteria used in our model, the adjusted R-squared, selects the model with the best out sampling predictive power. Then, in a second part we will compare the predictions of the model with a null model.

To answer the first task, we perform a cross-validation method and train the algorithm over a period of 9 years and compare its predictions 5 years ahead to actual data. The training period is from 1995 to 2004. The adjusted R-squared criteria indicates that 8 terms are sufficient to model the dynamic of the trophic depth and only 1 term for the dynamic of the Log(GDP per capita). The number of terms needed to model the dynamic of the trophic depth is different from the previous section. This effect is due to the fact that the sample size to train the algorithm is

not the same. We only focus on the number of terms for the dynamic of the trophic depth as for the dynamic of the Log(GDP per capita) it is already at its minimum. In Figure 5.8 we plot the errors between the projections made for 2009 and the actual data.

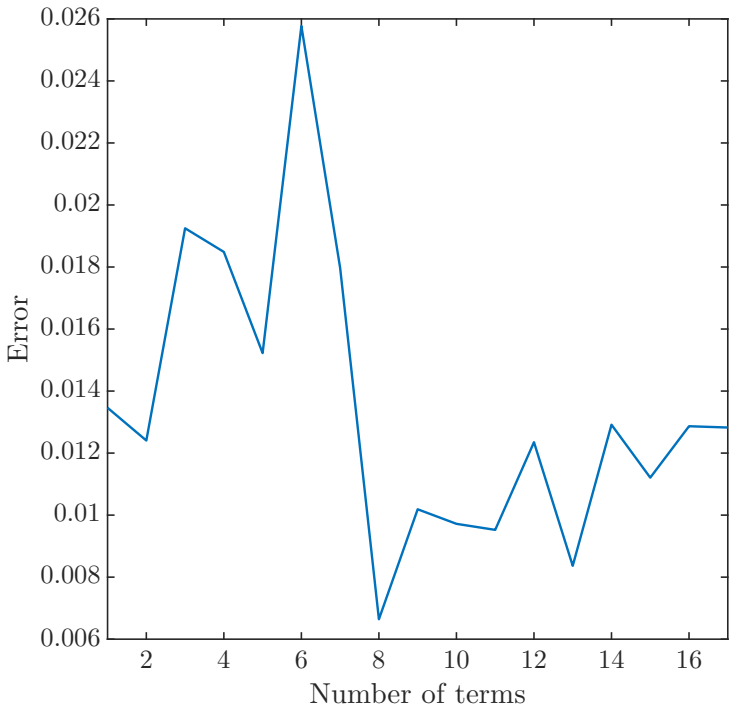


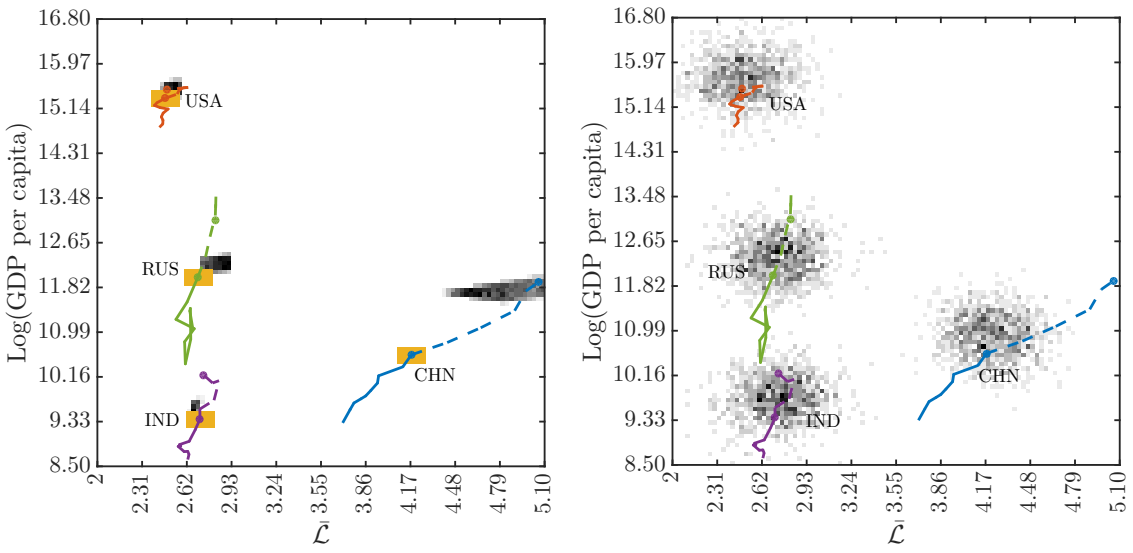
Figure 5.8: *Amplitude of the errors for different number of terms in the dynamic of the trophic depth.* We train the model selection with the data from 1995 to 2004 and compare the predictions for 2009 with actual data. We do that for different number of terms for the dynamic of the trophic depth and plot the errors as a function of the number of terms.

The predictions are the best when we minimise the errors made by the model during the cross-validation procedure. Thus, we find that with 8 terms in the dynamic of the trophic depth we best capture the dynamic of countries in the phase-space. This corresponds to the exact number of terms recommended by the adjusted R-squared criterion. This result emphasizes that the criterion used is relevant for our purpose.

Before comparing our approach to a null model, we need to detail how we perform

predictions with the model selection. The model given by our algorithm is deterministic as for any point in space there exists a unique trajectory. However, we want to introduce some variability. For this purpose, we consider that the last position of a country is only known at a certain level of precision. Thus, when we do some predictions, we first define a region around the position of a country then we solve the model for many starting points in this region. Finally, the prediction for a given country will be the median position of all the predictions made for the points located around the country position. The width of the box is 0.2 and its height 0.3 thus, along the $\bar{\mathcal{L}}$-axis we have 15 boxes and 30 along the $Log(GDP)$-axis.[7]

In Figure 5.9.a, we show how the predictions perform for the USA, India (IND), Russia (RUS), and China (CHN).



(a) Model selection algorithm.

(b) Null model.

Figure 5.9: *Comparison of predictions between the model selection algorithm and a model of random walk with drift* of the position of countries in 2009 using the two models trained from 1995 to 2004 and the observed position. The yellow box represents the uncertainty we have about the last position of a country and the grey area represents the distribution of the predictions for each country. The darker the region is, the more predictions converge to it.

[7]The choice for the width and the height is such that they correspond to two standard deviations of the empirical displacements along each dimension.

The yellow box represents the uncertainty we have about the last position of a country. The continuous lines represent the trajectories of the 4 countries from 1995 to 2004 and the dotted lines are the actual trajectory of countries from 2004 to 2009. The grey area represents the distribution of the predictions for each country. The darker the region is, the more predictions converge to it. We observe that our model is performing well in the region of high trophic depth while the null model seems to perform badly. On the other hand, in the region of low trophic depth, null model appears to perform similarly to our model. It suggests that the null model is biased against countries with high trophic depth. In the following, we compare the two models quantitatively.

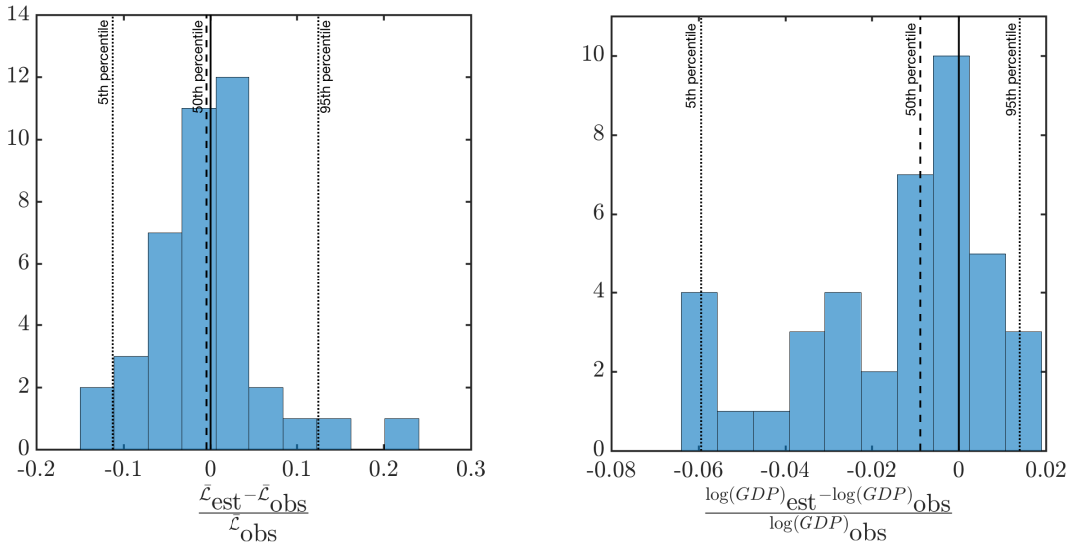
To compare the predictions with observed data we compute the errors along each axis. The errors along the $\bar{\mathcal{L}}$-axis are defined as the difference between the predicted Log(GDP per capita) and the observed one divided by the observed Log(GDP per capita) and we proceed similarly for the trophic depth. In Figure 5.10, we summarize the performance of the model.

On the two histograms, we have also plotted the 5th, 50th and 95th percentile of the distribution. We note that the 50th percentile is close to 0 indicating that the distribution of errors are centred around 0 and the difference between the 95th and 5th percentile is an indication of the how the errors are spread.

Now we compare our model selection approach with a null model. The null hypothesis is that each country's GDP and trophic depth move randomly. We assume the mean and the variance to be the same for all countries and constant in time. The mean and the variance are computed over the period 1995-2004. In this case, the dynamic is simply given by

$$d\bar{\mathcal{L}} = \mu_{\bar{\mathcal{L}}} dt + \sigma_{\bar{\mathcal{L}}} dW_t^{\bar{\mathcal{L}}}, \tag{5.11}$$

$$dLog(GDP) = \mu_{Log(GDP)} dt + \sigma_{Log(GDP)} dW_t^{Log(GDP)}, \tag{5.12}$$



(a) Comparison of the performance for the trophic depth.

(b) Comparison of the performance for the Log(GDP per capita).

Figure 5.10: *Histogram of the errors made using the model selection algorithm at a 5 years time horizon.* The continuous line represents the case where the predictions are exact. We also plot the 5th, 50th, and 95th percentile of the distribution of errors to indicate how they are spread.

where W_t is a Brownian motion. For simplicity, we assume that the noises are independent. Similarly to Figure 5.9.a, Figure 5.9.b shows the outcome of the random model for the USA, India, Russia and China.[8]

We now compare the two models by first looking at the distribution of the errors given by both models, see Table 5.1.

We note that the model selection predicts better the evolution of the trophic depth than the null model while its seems that the random walk with drift performs better for the Log(GDP per capita). However, there is a problem with the predictions made using the null model. In fact, the random model fails to predict the evolution of countries for which the trophic depth changes over time as we find a significant negative relation between the errors and the trophic depth. The random walk not only fails for the dynamic of the trophic depth but also for the evolution of the

[8]The simulations are done with a timestep of 1 year.

Variables	Errors from the selection model		Errors from the random model	
	$\bar{\mathcal{L}}$	Log(GDP per capita)	$\bar{\mathcal{L}}$	Log(GDP per capita)
Mean	-0.0019	-0.0054	0.0083	-0.0008
Median	-0.0012	-0.0054	0.0079	-0.0009
Std	0.0741	0.0231	0.0686	0.0277
Δ 95-5	0.2595	0.0755	0.2345	0.0861

Table 5.1: *Statistics about the distributions of errors derived from the selection model algorithm and the random model for the trophic depth and the Log(GDP per capita).* Std is the standard deviation and Δ 95-5 is the length of the interval delimited by the 95th and the 5th percentile of the distribution.

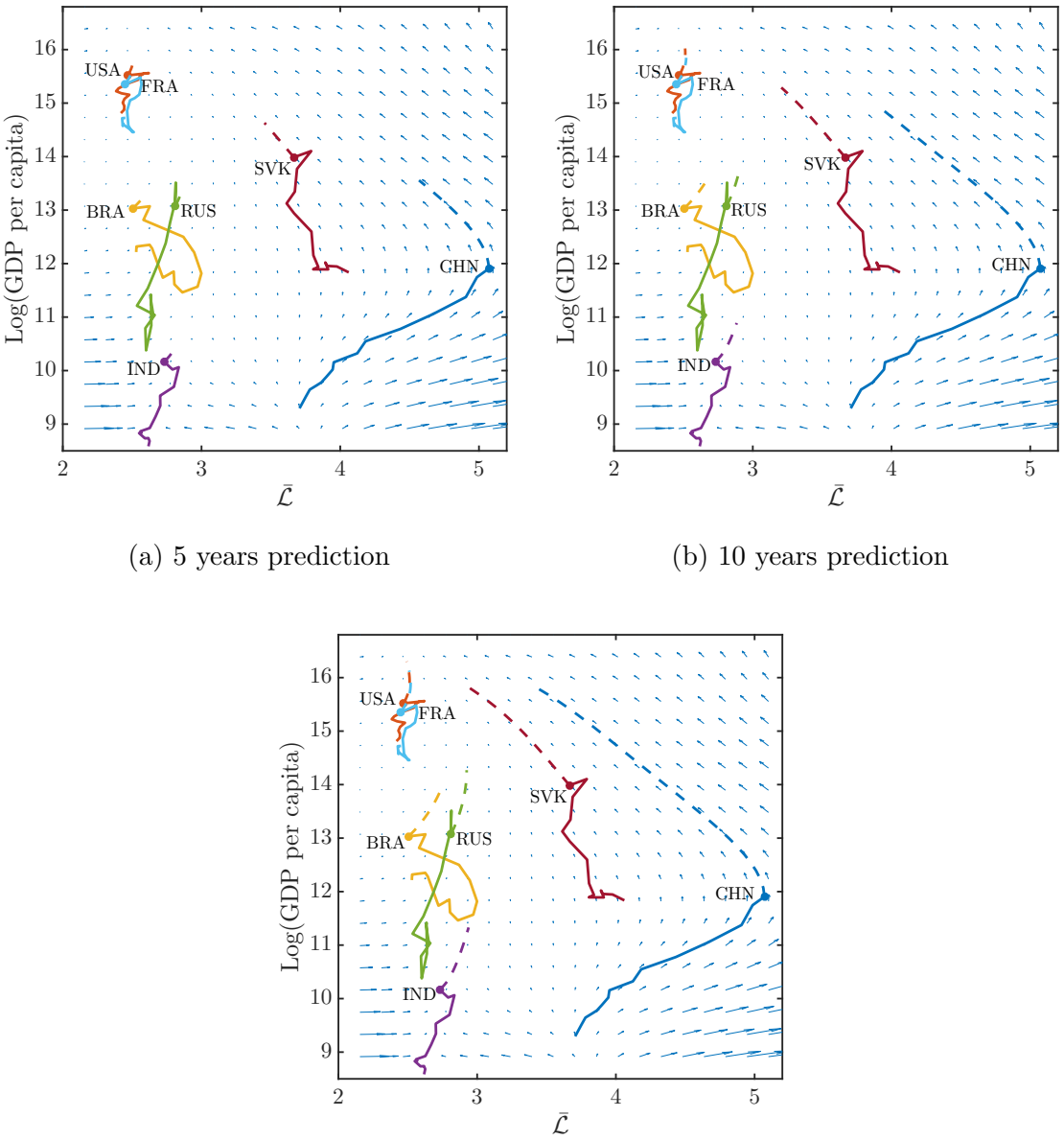
Log(GDP per capita). Instead with our model, the dynamic of countries with high trophic depth is as well modelled as the dynamic of countries for which the trophic depth remains constant. It is not surprising to see that the random model performs well when the trajectory of countries follows a random walk with drift as it is the case for any economy with a low trophic depth. Because it significantly fails predicting the evolution of countries in the arch-shape regions, we can reject the null model.

In this section, we analysed our model via a cross-validation method. We found that the model selection criteria we picked is relevant and leads to the most powerful model. Moreover, our model predicts the evolution of countries independently of where they are located in the phase-space contrarily to a random walk with drift that was our null model.

5.4.2.3 Future forecasts

Now that we have analysed the performance of the algorithm and compared it with a null hypothesis we will train the algorithm over the entire sample period and use it to do some forecasts about the position of countries in the phase-space. In Figure 5.11, we perform a 5, 10 and 15 years prediction for the USA, France (FRA), China (CHN), India (IND), Brazil (BRA), Russia (RUS), and Slovakia (SVK). In the following we

compare the outcome of our model with forecasts made by the IMF.



(a) 5 years prediction

(b) 10 years prediction

(c) 15 years prediction

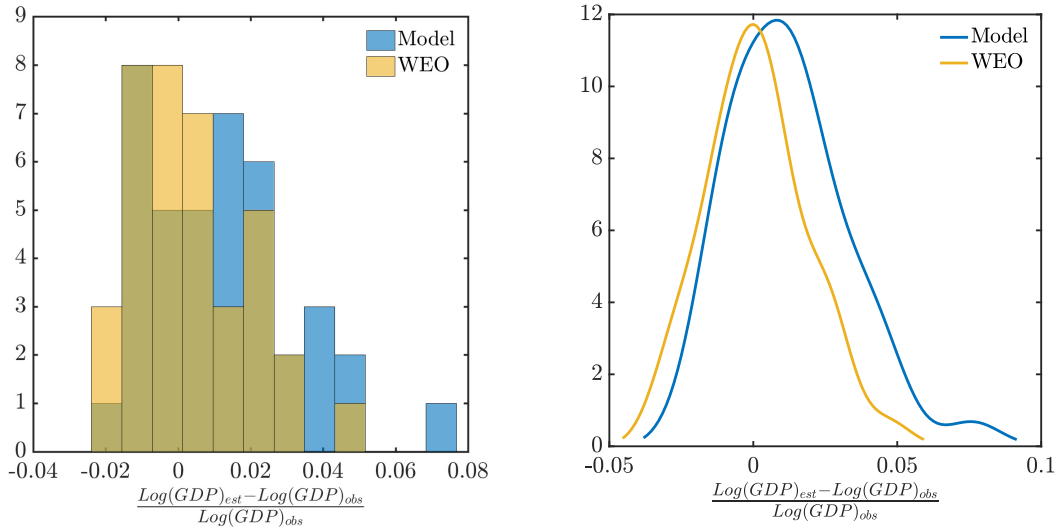
Figure 5.11: *Forecast of the trajectories of 7 countries in the phase-space for different time scale using the model selection algorithm.* The countries are the USA, France, Slovakia, Russia, Brazil, China, and India. The continuous lines represent the evolution of countries from 1995 to 2009. The dashed lines show the predicted trajectories at a 5, 10, and 15 years time horizon. The arrows represent the vector field derived from the model selection algorithm.

The IMF's World Economic Outlook is both analysing the actual state of the

world economy but is also making forecasts at different time horizons. The method they developed follows a "bottom-up" approach. The methodology used is country and time specific but overall the strategies follow a common pattern. It consists in making forecasts at the individual level assuming some global parameters. Then, the forecasts are aggregated to estimate the global parameters before they are plugged back into the first step. This is reiterated up to when the predictions converge.

We compare forecasts made in 2009 about the states of the 40 economies mentioned in the WIOD in 2014 both using our model and the predictions made by experts. We use the data from the International Monetary Fund (2009) for the forecasts and from the International Monetary Fund (2016) to compare the actual value of the GDP per capita in 2014. The most interesting result is the Chinese predictions. In 2009, the IMF predicts that the economic growth of China will increase from 8.5% to 9.5% in 2014. On the other hand, the model predicts that the trophic depth of China will decrease therefore China's growth will slow down. In fact, in 2014, the Chinese economic growth rate has decreased to 7.5% aligned with the trend predicted by the model. The errors made are summarised in Figure 5.12. We both show the overlapping histograms and the kernel density estimation of the errors.

The median of the errors made by the IMF is $-6.40 \cdot 10^{-6}$ while it is $1.03 \cdot 10^{-2}$ for our model. Additionally, the standard deviations are 0.0173 and 0.0204 respectively for the experts and for our model. We note that experts better predict the evolution of the logarithm of GDP per capita; however our model performs reasonably well given the minimalist and universal approach that we used. Moreover, the model had better predictions for 16 countries out of the 40 in the WIOD. The main outlier in the model prediction is China for which the error is around 0.07. Thus, the observed GDP is much smaller than the one predicted by our model. Its consequence is that the observed slowdown of the Chinese economy is higher than the one predicted by our model. Contrarily to the random model, experts are able to make realistic predic-



(a) Overlapping histogram

(b) Kernel density

Figure 5.12: *Overlapping histogram and kernel density of the forecasting errors made by experts and our model in 2009 about the Log(GDP per capita) in 2014.* The errors made by the model are in blue and by the experts are in orange. For each approach we have 40 observations.

tions independently of the trophic depth of countries. We do not find evidence for the model to better predict a category of countries compared to experts but instead the model does as well as them, though it is simpler and is not country specific. Again we perform predictions based on a random walk with drift and we find that it performs not as well as our model or as the experts.

In Figure 5.11 we note that in about a 15 years time horizon, China and Slovakia would have reached the same level of wealth as the most developed economies. However, given that after 5 years the slowdown is already more important than the one we predict, it might take some additional years to actually reach developed economies. On the other hand Brazil, Russia, and India will fall behind developed economies. It is important to notice that while at the beginning of the chapter the arch-shape trajectory was either inferred or expected, in the model that we developed the arch-shape trajectory is now emerging from the combination of countries. Our model provides an explanation for the acceleration and deceleration of economic growth of economies

for which the trophic depth fluctuates.

These predictions are only based on the analysis of the dynamic of the trophic depth and its non-linear relation with the level of wealth but further analysis should be realized taking into account other drivers of growth. This can be done using the same framework as what we did and adding additional variables in the model. It will increase the dimensionality of the problem but improve its accuracy.

5.5 Conclusion

In this chapter, we studied the evolution of the trophic depth at different stage of development. Under some approximations, we inferred the trophic depth of the USA from 1790 to 2014 and noted that it followed an arch-shape trajectory. Then, we highlighted that in the WIOD some countries like China were following the same type of trajectory. On the other hand, some emerging economies are following trajectories closer to a laminar flow as their trajectories are straights in the phase-space given by the Log(GDP per capita) and the trophic depth. Finally in the remaining part of the chapter, we developed a model selection algorithm to forecast the dynamics of the trophic depth and of the income of countries. This approach confirmed our intuition that there are two potential growth paths. Moreover, we found that the dynamic of Log(GDP per capita) is best explained by a model that only depends on the trophic depth. This shows that the dynamic of the trophic depth provides one explanation for the acceleration and deceleration of growth rate of countries along their development path.

The forecast we made about the Chinese economy is interesting as it matches some recent observations about the slowdown of the Chinese economy. However, the model underestimates the deceleration. In 2015, China experienced its slowest growth rate over the last 25 years. In the same time, the China's Five-Year Plan (2016-2020)

confirmed the shift towards moderate economic growth and related that to the development of better services that we know are related to lower trophic levels.

In the coming months, the WIOD will release an updated version of the database with three additional countries: Norway, Croatia, and Switzerland. The period covered will be extended up to 2014 and the number of sectors will increase from 35 to 56. It will still not include any African country but Norway is a resources rich country and therefore it would be interesting to look at its dynamic. Moreover, we would be able to improve our predictions and to check the forecasts made earlier.

The speed of evolution of countries in the phase-space is likely to be time dependent because it is related to the set of technologies available at one time. Then, we expect China to evolve along the arch-shape path at a faster rate than the USA. We cannot test that but to do it, we would need to perform several model selections with the same terms for different periods. Then, we would study how the coefficients evolve with time.

The approach we followed did not provide any answer about how to move from one growth path to another. To answer this question, we will need to dig into the black box to understand the structural change at the firm level and to model the evolution of the input-output network.

Nonetheless, we can ask ourselves what would be the impact of robotisation on the trophic depth of developed economies. Two effects can occur and both will increase the trophic depth of countries. First, the replacement of humans by robots in service industries would increase their trophic level and consequently the trophic depth of countries that are largely dependent of services will increase (Frey and Osborne, 2017). Moreover, if manufacturing is robotised therefore it might become viable to reinstall manufacturing industries in developed economies. Thus, their trophic depth would also increase. However, this could also create some additional issues like wealth redistribution depending on who owns the robots.

Chapter 6

Linkage dynamic of the input-output network

6.1 Introduction

The model of technological improvement introduced in chapter 3 is only focused on material or labour efficiency: to produce the same quantity of output, producers need less inputs or labour. Despite the simplicity of the mechanism, we have been able to identify the key role of the trophic depth to understand economic growth.

In this chapter, we focus on another model of technological progress that is connected to the dynamic of the input-output network. Indeed, if the recipe to create one good changes it will be translated into the input-output network framework by the adoptions of new links between industries. Here we make the assumption that the complexity of products are increasing over time, meaning that once a product has been integrated into a production process it will not be removed. Hence the only possibility to change the recipe is by adopting new products. Thus, we study the linkage dynamic of the input-output network and therefore we are looking at the parameters explaining the probability of adoptions.

This study is in the same vein as the work of Carvalho and Voigtländer (2014); however, in our approach we use some additional variables directly related to the results derived in the previous chapters. These additional variables turn to be not

only complementary to those used in Carvalho and Voigtländer (2014) but also they have significantly more explanatory power. While in the previous works of this thesis the technological progress was an exogenous variable in the model, the analysis in this chapter can be viewed as an attempt to endogenise it using the input-output framework.

The underlying idea in our approach is that the existing structure of the network contains information about future potential adoptions. Hence, the probability of adoption is a function of the network proximity. The economic intuition is that the suppliers of your suppliers are more likely to become a new direct supplier compared to any other industry assuming that it is not a direct supplier already (see Carvalho et al. (2014) for a network representation in the case of semiconductors). By default, we do not have any weight threshold to define the adoption, however, we perform some consistency checks and impose a non-zero threshold.

In general, studies on network dynamics make the assumption that while the network evolves, its structure remains consistent over time; we note that the node degree distribution is often a key parameter. Thus, the potential new links are those that preserve the network properties. Lü et al. (2015) focus on the link predictions across various complex networks. Their strategy is to perform some random shocks on the adjacency matrix and then use a structural consistency parameter to estimate the link predictability. They have shown that their algorithm outperforms the state-of-the-art link prediction methods. A major difference between them and us is that in the 13 real networks they consider, the average node degree is much smaller than the total number of nodes in the network while the density of input-output networks that we consider is close to 1.[1]

Jackson and Rogers (2007) made the distinction between links that are created randomly and those for which the structure of the network is predictive. They use the

[1]We use the most disaggregated data that we are aware of.

node degree distribution to evaluate the fraction of links that are randomly adopted compared to those that are formed via the structure of the network. Carvalho and Voigtländer (2014) used this approach to derive some theoretical results and to explain the dynamic of the input-output network. It can be justified at the firm level analysis where the network is disaggregated but we believe that given the level of aggregation of the sectoral input-output network we need a theory that is independent of the node degree distribution, as in a dense network, a node's degree is not informative.

In this chapter, we first describe the different variables that can be used to approach the network proximity. Then, we present the model used to investigate the role of the network structure to explain the adoption of products. Finally, we test the model using time series data of the USA input-output network.

6.2 Model

In the first section, we describe the model and we introduce the different variables associated with the idea of network proximity.

6.2.1 Evolutionary dynamics

Rogers (1995) describes the adoption process as a series of choices and actions to evaluate a new idea before deciding to integrate it or not. It is a long time process and it can be divided into five different stages: knowledge, persuasion, decision, implementation, confirmation. A review of the determinants for innovation adoptions is done by Frambach and Schillewaert (2002). In this chapter, we focus on some aspects of the adoption mechanism that fall into the categories labelled Social Network and Environmental Influences. These two factors are related to network proximity. We will also control for other key determinants like the relative economic advantage

and the size of the adopters.

The probability of an adoption by sector i of goods produced by sector j at time t given that at time $t - \Delta t$ there was no link[2] is given by

$$Prob(a_{ji}(t) > 0 | a_{ji}(t - \Delta t) = 0) = f(X_{ij}, X_i, X_j), \tag{6.1}$$

where a_{ij} is the technical coefficient,[3] X_{ij} is a set of variables that depends on the network proximity, X_i (resp. X_j) are a set of variables that depend only on the adopting (resp. producing) sector. The functional form of f will be discussed in section 6.3.2.2. When studying the adoption by i of a good produced by j we are interested in the influence that industry j has on industry i. In the next section we define precisely the variables capturing the network proximity between two industries.

6.2.2 Definition of network proximity

In the previous section, we relate the environmental influences to the network proximity. We now describe three different approaches to approximate the network proximity. The first two are the same as in Carvalho and Voigtländer (2014) while the choice of the third variable is one of our main contributions. Our approach is radically different.

First, we can compute the network distance between two nodes. There are two ways of doing that. The first method simply requires a binary adjacency matrix. The geodesic distance between two nodes given all the possible paths between them is the number of jumps along the shortest path. In dense networks, as is the case for input-output networks at the sectoral level, this value will be 1 for most nodes. This

[2]Here we do not impose a weight threshold. However, we can do it in two ways. We either impose that the new good has an expenditure share higher than a certain level or we impose that the money flows is above a certain amount. In the future, we choose the second method.

[3]The technical coefficients or input coefficients are the expenditure on good i per dollar of production by j. Hence, when looking at the adoption of good j by i, the money flow is from i to j hence it is represented by a_{ji}.

definition ignores the weights of the edges. It is a crude assumption as in input-output networks the weights distribution is closely fit by a log-normal. This definition is not appropriate for our work.

The second method includes the weight of edges. Instead of using the raw money flows, we used its normalized version: the technical coefficients. If a_{ij} is the input coefficient and is non-zero, the directed network distance from i to j is $d_{ij} = \frac{1}{a_{ij}}$.[4] Because the input-output network is not symmetric d_{ij} is different than d_{ji} and by definition d_{ij} is greater than or equal to 1.[5] The matrix of distances has the same density as the input-output network. To compute the distance between any nodes in the network that are not directly connected, we use the shortest path algorithm and the distance from i to j can be written iteratively as

$$d_{ij} = \min_{\substack{k \neq i,j \\ k \in V}} \left\{ \frac{1}{a_{ik}} + d_{kj} \right\}, \tag{6.2}$$

with V the set of nodes. Thus, the distance from i to j is given by the sum of the weights that minimize the overall length. As long as a path exists the distance is defined. This approach is taking into account the weights' distribution. However, there is still one issue with this approach that is intrinsic to the density of the network. This algorithm will only consider the shortest path and will ignore all the other potential paths. This is not entirely satisfying for our work as it misses some additional information based on the network structure. In the following, we will fix this issue.

Finally, we introduce another variable that is taking into account all the possible paths from one node to another to compute their proximity. It is related to the Katz centrality (Katz, 1953). While the Katz centrality simply has an attenuation factor that will be independent of the strength of the links, we would need to change the

[4]While algebraic distance has to be symmetric, in graph theory the geodesic distance is not necessarily symmetric for directed graphs.

[5]By definition, a_{ij} is between 0 and 1.

approach such as we include their weights. The weight has to be representative of the influence that one node has on another. The Leontief inverse, noted L, is given by

$$L = (I - A^T)^{-1}, \tag{6.3}$$

where A is the matrix of money transition rates. The influence of node j on node i is equal to L_{ij} and we call it the *influence proximity parameter*. To see why this is a good candidate to capture the network proximity let us explain the meaning of the Leontief inverse. L_{ij} represents the production of sector j needed to produce one unit of final demand of good i. Even if there is no direct link between i and j, it is non zero as long as there is a path from j to i. Indeed, it includes the need of good j to produce all the intermediate goods needed to get one unit of good i. Thus, it takes into account all the possible paths from j to i including self-loops. We will discuss in section 6.2.3 why taking self-loops into account could be important. A high value of L_{ij} indicates that good j has an important role in the production of good i and if there is no direct link between the two it is an indicator of how sensitive sector i is to the production of good j. On the other hand, a small L_{ij} shows that j is not an important good in the production chain that is leading to one unit of production of i.

To compare the weighted distance proximity with the influence proximity we realize a case study. In 1972, 31 industries are using semiconductors while 25 years later, 117 industries have adopted semiconductors. For all the industries in 1972 that are not using semiconductors we compute their proximity with the Semiconductor industry using both proximity variables and then we rank the industries according to their score. The industry with the shortest distance will be ranked at the top. In Table 6.1, we summarize the ranking for both variables.

The distance proximity, denoted d_{ij}, explains well that the Office Machines in-

Ranking	Distance proximity	Influence proximity
1	261⋆	235
2	316⋆	261⋆
3	163⋆	255⋆
4	142	17
5	175⋆	311
6	174⋆	281⋆
7	164	263
8	321	308⋆
9	234⋆	221
10	136	242⋆

Table 6.1: *Comparison of the industry ranking for the distance proximity and the influence proximity in the case of semiconductor's adoption.* The distance and the influence are computed in 1972 and for each industry that was not originally directly using semiconductors we study whether it has adopted them any time later. The top industries are those that are the closest to semiconductors. The ⋆ indicates that the corresponding industry had adopted semiconductors in the following 25 years. See Table 6.17 for the meaning of each industry code.

dustry (code 234) and the Electronic Components industry (code 261) have adopted semiconductors. On the other hand, the influence proximity explained not only the adoption by the Electronic Components industry (code 261) but also the Radio and TV Receiving Sets industry (code 255) and the Automated Temperature Controls (code 281). We note that except for the Electronic Components industry, which appears in both rankings, there is no redundancy between the two proximity variables. It suggests that both variables are complementary but also that the influence proximity is capturing another type of information. To compare the ranking of industries given by the two methods, we compute the Spearman's rank correlation coefficient. It is equal to 0.2039 and the p-value is $2.54 \cdot 10^{-4}$. The correlation despite being significant is not really high.

6.2.3 Network aggregation

Network aggregation is often an issue in network analysis and only few properties are preserved by the aggregation process. For instance, it can introduce noise and randomness in the adoptions that would have initially been based on the network structure. The main consequence of network aggregation is an increase in the network density. For the purpose of consistency when two nodes are aggregated then the money flows between the two are transformed into a self-loop for the new node.

None of the three methods to compute the network proximity are preserved by the network aggregation. This could be problematic; nonetheless, the influence proximity has an advantage that none of the others have. When computing the first two network proximity variables, self-loops are simply discarded. But when computing the influence proximity, self-loops are taken into account and are treated equally compared to other links. Therefore, we believe that the influence proximity is more robust to network aggregation issues.

6.3 Application at the sectoral level

We now study the probability of adoption at the sectoral level for the USA over 25 years.

6.3.1 Description of the data

We analyse the probability of adoption of a new good into the production chain of any other goods. We use data from the US Bureau of Economic Analysis (2015) at the most disaggregated level. We are aware that while the BEA reports data they might have set a threshold below which they are reporting no transactions. The smallest new transaction observed is equal to \$27,000 and it suggests that their threshold is smaller than this value. Input-output tables contain 350 industries and the time

horizon is 25 years. While the BEA provides input-output tables for some other years; the classification has evolved and to be consistent we need to have the same industry classification over time. Therefore, we have data from 1972 to 1992 with a snapshot every 5 years. The time interval is noted Δt. In Table 6.2 we summarise some basic statistics on the data.

Variables	**Value-Mean**	**Standard deviation**
Number of years	5	–
Number of industries	350	–
Number of observations	324,859	–
Number of adoptions	46,175	–
Density of I-O	0.7138	0.1609
$\ln(d_{ij})$	3.24	0.50
$\ln(\mathcal{L}_{ij})$	-12.22	3.58

Table 6.2: *Statistics of the USA input-output networks from 1972 to 1992.* The density of the input-output network is the fraction of non-zero links. We also include the statistics of the two proximity variables introduced in Section 6.2.2 for all the potential new links. The number of observations is the number of unobserved links that form the set of potential new adoptions.

Overall there are 324,859 potential adoptions in the input-output network over the 25 years that corresponds to all the non existing links.[6] Over the 25 years, we have observed 46,175 new adoptions meaning that around 14% of the potential adoptions have been adopted. Moreover, as already mentioned earlier, the density of the input-output network is really high and in average is equal to 0.71.

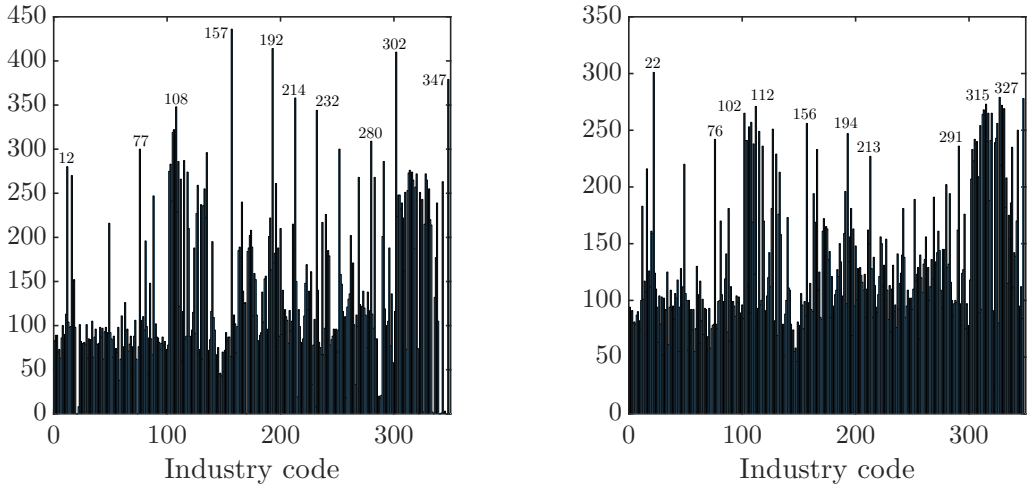
6.3.2 Approach to the probability of adoptions

We first present some more refined details about which industries produced the goods that are adopted and which industries adopted them. Finally, we describe the mathematical expressions used to study the probability of adoption.

[6]Note that because we are studying the adoption at each period if one link has never been adopted we have one observation at each year so in total this link appears 5 times in the dataset. Though because the network evolves, all properties associated to the link are time dependent.

6.3.2.1 Adoption properties

We show in Figure 6.1 the cumulative number of adoptions over the entire sample period both for producers and adopters. This number can be higher than the number of nodes in the network because some links can appear one year then disappear before reappearing again, and when looking at the cumulative number of adoptions this is counted twice.[7] In future analysis, we would be able to control for that by only including new links that exist for more than a certain time.



(a) Producers (b) Adopters

Figure 6.1: *Cumulative number of adoptions over the entire sample period for each industry.* We make the distinction between industries that produced the good (left) and those that have adopted a new input good (right). The industry code is given in Table 6.17.

We have highlighted some industries according to the amplitude of their cumulative number of adoptions. The number of adoptions does not distinguish small or large adoptions by small or large sectors. The rank-size distribution figure can be found in the appendix; Figure 6.3. On the producers' side, the goods that were the most adopted were produced by the Abrasive Products industry (code 157), the

[7]This behaviour can be assigned to experimentation, for instance, or it can also be due to the fact that the value is oscillating around the threshold of the reporting value.

Cutlery industry (code 192), and the Power Driven Hand Tools industry (code 214).[8] On the other hand, on the adopting side, we notice that the Maintenance and Repair industry (code 22) and the Hardware industry (code 194) have adopted a large panel of new goods. We notice that two groups of industries are particularly active: the first group has industry codes between 101 and 114 and the second group has industry codes higher than 303. The first group corresponds to the Paper and Allied Products manufacturing and the Printing, Publishing, and Allied Industries while the second group groups the Transportation, Communications, and Utilities industries, the Wholesale and Retail Trade industries, the Finance, Insurance, and Real Estate industries, and the Services industries.

6.3.2.2 Two mathematical forms for the regression functions

So far we have not specified the mathematical form of the function f used in Equation 6.1. Here, we focus on two functional forms. The first one is based on the Probit regression model (Bliss, 1934) and the second one is the ordinary-least-square (OLS) regression model. Given the binary response of our problem, adoption or not at every period, the most accurate approach is the one using the Probit regression model. The OLS approach is indicated for comparison.

In the Probit model, the dependent variable can only take two values hence it is a natural choice for our problem. The function f is then the cumulative distribution function of the standard normal distribution, noted Φ. It returns the probability to observe one phenomenon given the states of the explanatory variables. The equation of the problem is

$$Prob(a_{ji}(t) > 0 | a_{ji}(t - \Delta t) = 0) = \Phi(\boldsymbol{\beta}^T \boldsymbol{X}), \tag{6.4}$$

[8]Note that the industry code 302 corresponds to the Miscellaneous Products industries not elsewhere specified. It is not informative of what the goods are.

where $\boldsymbol{X}$ is the vector of explanatory variables and $\boldsymbol{\beta}$ the vector of unknown parameters. By definition, the function is bounded between 0 and 1.

To analyse the performance of a Probit model we compute the area under the receiver operating characteristic (ROC) curve, denoted AUC (Hanley and McNeil, 1983). It provides a number that summarises how well the model can separate adoptions versus non-adoptions using the rank of the scores given by $\Phi(\boldsymbol{\beta}^T\boldsymbol{X})$ for all the potential adoptions. In the case of a random process, the AUC is equal to 0.5. If the AUC of the model is higher than 0.5 it means that the model is performing better than at random. This can also be represented in a graph when plotting the true positive rate versus the false positive rate as in Figure 6.2.

The dashed line corresponds to the performance of a random model and the associated AUC is 0.5. The continuous blue line is the ROC curve for the complete adoption model specified in Table 6.8. The AUC value is 0.7653.

To compare the performance of the Probit model, we will also perform an OLS regression. Note that in the case of the OLS regression the function is no longer bounded between 0 and 1 and thus cannot be interpreted as the probability of adoption. In this case, the function f is simply the identity. The equation of the problem becomes

$$P(a_{ji}(t) > 0 | a_{ji}(t - \Delta t) = 0) = \boldsymbol{\beta}^T\boldsymbol{X}, \tag{6.5}$$

where P is not a probability but instead a function that provides a high score if a link is likely to appear in the future and low score if a new link is very unlikely. To compare different models we use the R-squared as the AUC is not appropriate for an OLS regression.

For both models, we also want to compare the contribution of each variable and not just how the overall model performs. For this purpose we will also study the increase or decrease of the probability of adoption if the explanatory variable increases by one standard deviation. Thus, we would be able to identify the most important variables.

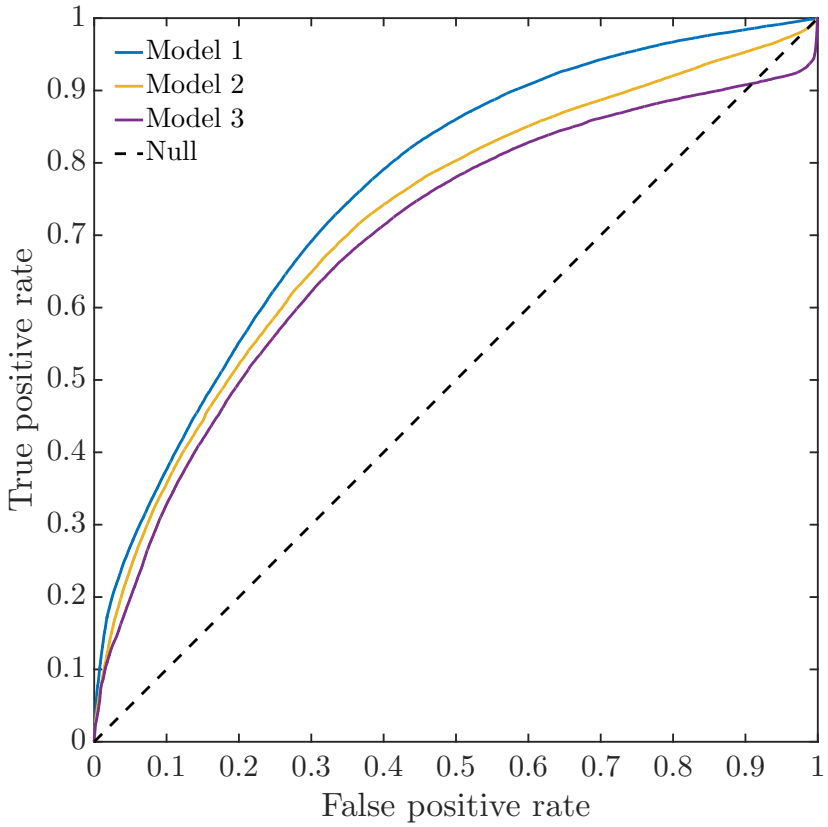


Figure 6.2: *Receiver operating characteristic for classification by Probit regression of the complete adoption model.* The model 1 is given by the regression model specified in Table 6.8 and the AUC value is 0.7653. In model 2, we have removed the influence variable. In model 3, we have removed the influence variable and the trophic level, such that it correspond to the model without our contribution. The dashed line is $x = y$ and represents a random model: the null hypothesis.

In the following sections, every regression will be performed according to both models.

6.3.3 Distance proximity versus influence proximity

We first compare the two terms capturing the network proximity between industries. To prevent any outliers from drastically influencing the regressions we use the logarithm of the distance and of the influence. Then, the probability of adoption model

using a Probit regression is given by

$$Prob(a_{ji}(t) > 0 | a_{ji}(t - \Delta t) = 0) = \Phi(\beta_0 + \beta_1 \ln(d_{ji}) + \beta_2 \ln(L_{ij})). \tag{6.6}$$

In the following, we analyse how each explanatory variable performs alone by setting either β_1 or β_2 equal to 0 and we also include a constant term. Then we study the significance of each of them when they are both included in the regression analysis. Table 6.3 shows the results for each network proximity variable independently.

Explanatory Variables	**Probit**		**OLS**	
	1	2	3	4
$\ln(d_{ji})$	-0.2086^{***} (0.0055) [−8.26%]		-0.0566^{***} (0.0012) [−2.81%]	
$\ln(L_{ij})$		0.0920^{***} $(8.271 \cdot 10^{-4})$ [25.78%]		0.0213^{***} (0.0002) [7.60%]
Constant	-0.3995^{***} (0.0178)	0.0029 (0.0098)	0.3256^{***} (0.0040)	0.4019^{***} (0.0021)
AUC	0.6335	0.6406	—	—
R-squared	—	—	0.0065	0.0473

Table 6.3: *Comparison between the distance proximity and the influence proximity to explain new adoptions.* Regression table with a single interaction term with a Probit and an OLS regression model. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We first notice that for any of the two functional forms considered the result is qualitatively similar. In the regression tables, the figures in brackets correspond to the increase or decrease of the probability of adoption if the explanatory variable increases by one standard deviation. They are calculated via normalized regressions. We find that according to both mathematical expressions the influence variable is always the variable that increases the probability of adoption the most due to a

one standard deviation increment. For the Probit model, a one standard deviation increase augments the probability of adoption by 25.78% for the influence variable and reduces it by 8.26% for the distance variable. We also note a difference of sign between the two estimates. The longer the distance between two nodes, the less the probability of adoption of a new input is. Then, we expect a negative sign for the estimate of $\ln(d_{ji})$. However, the influence variable has a positive sign meaning that the more a node is influential, the higher the probability to have a new link between any two nodes if there is no existing link yet. All estimates have a p-value smaller than 0.01 except the constant term of the Probit model with the influence term only. The OLS regressions provide qualitatively similar results when comparing estimates except that the amplitude of change in adoption probability due to one standard deviation increase in the explanatory variables is about three times smaller than with the Probit model.

In Table 6.4, we include both variables in the same regression and also introduce a fixed year effect (FYE) control parameter. The FYE allows us to remove any year-specific effect that might exist.

First, we notice that all the estimates are statistically significant. This highlights the combining effect of the distance and of the influence variables that we have already raised when we looked at the particular case of semiconductor's adoption. Moreover, both variables are even more significant when added together. Indeed, the change of probability of adoption due to an increase of one standard deviation of the influence parameter is 15% higher than before and a decrease of one standard deviation of the distance parameter augments the probability of adoption by 80% compared to when it was not coupled with the influence parameter. Nonetheless, the influence is still the most significant parameter as a one standard deviation increase augments the probability of adoption twice as much as a one standard deviation decrease of the distance variable. Finally, when we compare the AUC of a model including both

Explanatory Variables	Probit		OLS	
	1	2	3	4
$\ln(d_{ji})$	-0.3804^{***} (0.0056) [−14.99%]	-0.3853^{***} (0.0056) [−15.18%]	-0.0764^{***} (0.0012) [−3.80%]	-0.0798^{***} (0.0012) [−3.96%]
$\ln(L_{ij})$	0.1062^{***} $(8.78 \cdot 10^{-4})$ [29.60%]	0.1070^{***} $(8.93 \cdot 10^{-4})$ [29.80%]	0.0226^{***} $(1.67 \cdot 10^{-4})$ [8.06%]	0.0231^{***} $(1.72 \cdot 10-4)$ [8.27%]
Constant	1.3831^{***} (0.0224)	1.3536^{***} (0.0227)	0.6656^{***} (0.0047)	0.6542^{***} (0.0047)
FYE		√		√
AUC	0.6981	0.6979	—	—
R-squared	—	—	0.0590	0.0597

Table 6.4: *Comparison between proximity variables with fixed year effect control parameter.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

proximity variables with one including only one variable we note that the power of the model has increased. Again, there are no big differences between the Probit model and the OLS model other than the amplitude of the estimates. Moreover, the FYE control parameter has very small impacts on the regressions. It increases the probability of adoption of both variables by a tiny amount. In the following we will not include the FYE control parameter as its contribution is small.

6.3.4 Original values

We now focus on whether the variables in 1972, at the beginning of the dataset, have some explanatory power for the new adoptions over the following 25 years covered by the data. Table 6.5 represents the results of the regressions.

First, we notice that for the Probit model the sign of the constant term is now negative while it was positive before. The initial distance alone performs better than the distance in the previous 5 years alone. As a one standard deviation decrease

Explanatory Variables	Probit			OLS		
	1	2	3	4	5	6
$\ln(d_{ji})_{1972}$	-0.1732 (0.0024) [-13.74%]		-0.1045 (0.0028) [-8.32%]	-0.0483 ($6.07 \cdot 10^{-4}$) [-4.82%]		-0.0321 ($6.75 \cdot 10^{-4}$) [-3.21%]
$\ln(L_{ij})_{1972}$		0.0558 ($7.11 \cdot 10^{-4}$) [16.20%]	0.0405 ($8.24 \cdot 10^{-4}$) [11.79%]		0.0138 ($1.66 \cdot 10^{-4}$) [5.05%]	0.0099 ($1.84 \cdot 10^{-4}$) [3.62%]
Constant	-0.5789 (0.0072)	-0.4839 (0.0078)	-0.3476 (0.0085)	0.2836 (0.0019)	0.2925 (0.0019)	0.3440 (0.0022)
AUC	0.5941	0.6080	0.6395	—	—	—
R-squared	—	—	—	0.0191	0.0209	0.0277

Table 6.5: *Adoption properties using values of the proximity variables at the beginning of the dataset.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to a one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***.

in the distance in 1972 would increase the probability of adoption by 13.74% while it would increase by 8.26% using the distance 5 years before the adoption. On the other hand, the initial influence parameter is less pertinent than before however it still performs better than the distance variable. When both variables are included in the same regression the influence variable performs 50% better than the distance variable. This is significantly less than the factor two we had in the previous regression in Table 6.4. When comparing the AUC values we find that overall a model using the values of the explanatory variables in 1972 are not performing well. It means that the current state of the network can only be used to predict new adoptions in the short term.

6.3.5 Complete adoption model

We now focus on a broader adoption model, meaning that we also include in the regression variables that are sector specific and not related to network proximity.

First, we extend the model by adding the trophic level of sectors. In the previous chapters and according to our model of technological change, sectors with a high

trophic level should produce goods for which the price is dropping faster than others. We have also made the connection between the position of an industry in the country production chain and its trophic level. Table 6.6 summarizes the regressions including the trophic level of the adopting and of the producing sectors.

Explanatory Variables	**Probit**	**OLS**
$\ln(d_{ji})$	-0.3823 (0.0057) $[-15.06\%]$	-0.0684 (0.0012) $[-3.40\%]$
$\ln(L_{ij})$	0.1094 $(9.05 \cdot 10^{-4})$ $[30.43\%]$	0.0217 $(1.65 \cdot 10^{-4})$ $[7.76\%]$
$\mathcal{L}_i$	-0.1650 (0.0056) $[-6.67\%]$	-0.0352 (0.0012) $[-1.79\%]$
$\mathcal{L}_j$	-0.5364 (0.0054) $[-22.98\%]$	-0.1136 (0.0011) $[-6.18\%]$
Constant	2.8558 (0.0277)	0.9450 (0.0055)
AUC	0.7459	—
R-squared	—	0.0940

Table 6.6: *Adoption properties including the trophic level of industries.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***. j is the producing industry and i is the adopting industry.

We note that the sign of the trophic level estimates of both the adopting and producing nodes are negative. The trophic level of the producing sector is even more important than the distance between two nodes. This means that a sector with a low trophic level has a higher probability that its good gets adopted as a new input by other industries. We also find a negative relation with the trophic level of the adopting sector though it is not as significant as for the producing sector. One could claim that this result is mainly driven by the fact that we consider the input-output

network as a whole including all the services industries.[9] In appendix A, we redo the regressions considering only manufacturing sectors.[10] The results are summarized in Table 6.11 and we find very similar results. The order of magnitudes of the estimates and their contributions are the same. This provides a consistency check that services are not driving the results and that the mechanism that the goods of industries with low trophic levels are more likely to be adopted is universal.

We now consider some additional variables to capture the difference of productivity across sectors. The total factor productivity (TFP) of industry i at time t is approximated by

$$TFP_i(t) = \frac{Y_i(t)}{VA_i(t)}, \tag{6.7}$$

where Y_i is the total production of industry i and VA_i its value added. It is not a very conventional way to compute the TFP but we cannot use the NBER database used in Carvalho et al. (2014) as it only focuses on Manufacturing industries so it is our only way to capture the productivity of industries. We have checked that the amplitude of the effect we measure is aligned with the results in Carvalho et al. (2014). The change in total factor productivity (ΔTFP) is given by

$$\Delta TFP_i(t) = \frac{TFP_i(t)}{TFP_i(t - \Delta t)} - 1. \tag{6.8}$$

Table 6.7 shows the results of the regression including both the TFP and the change in TFP of the adopting and producing sectors. As for the proximity variables, we use the logarithm of the total factor productivity instead of its level to prevent any outliers from driving the results.

[9] In the work of Carvalho and Voigtländer (2014), they only focus the analysis on manufacturing industries while here we are not making any distinction between manufacturing and non-manufacturing industries.

[10] All the manufacturing industries are listed in Table 6.17 under the Manufacturing classification. A manufacturing sector has its first digit comprised between 2 and 3 according to the 1987 ISIC classification. In this sub-dataset we observe 24581 adoptions which represent 53% of the whole number of adoptions.

Explanatory Variables	Probit		OLS	
	1	2	3	4
$\ln(d_{ji})$	-0.3801^{***} (0.0056) [−14.98%]	-0.3633^{***} (0.0056) [−14.33%]	-0.0762^{***} (0.0012) [−3.79%]	-0.0715^{***} (0.0012) [−3.55%]
$\ln(L_{ij})$	0.1067^{***} $(8.79 \cdot 10^{-4})$ [29.72%]	0.1114^{***} $(8.94 \cdot 10^{-4})$ [30.95%]	0.0226^{***} $(1.68 \cdot 10^{-4})$ [8.09%]	0.0235^{***} $(1.69 \cdot 10^{-4})$ [8.38%]
ΔTFP_i	$6.86 \cdot 10^{-4***}$ $(5.89 \cdot 10^{-5})$ [2.28%]	$5.64 \cdot 10^{-4***}$ $(6.22 \cdot 10^{-5})$ [1.88%]	$1.45 \cdot 10^{-4***}$ $(1.43 \cdot 10^{-5})$ [0.60%]	$1.15 \cdot 10^{-4***}$ $(1.48 \cdot 10^{-5})$ [0..48%]
ΔTFP_j	$2.09 \cdot 10^{-5}$ $(6.90 \cdot 10^{-5})$ [0.07%]	$7.78 \cdot 10^{-4***}$ $(6.59 \cdot 10^{-5})$ [2.47%]	$1.69 \cdot 10^{-6}$ $(1.49 \cdot 10^{-5})$ [0.01%]	$1.56 \cdot 10^{-4***}$ $(1.55 \cdot 10^{-5})$ [0.62%]
$\ln(TFP_i)$		-0.0127^{***} (0.0016) [−1.87%]		-0.0030^{***} $(3.38 \cdot 10^{-4})$ [−0.55%]
$\ln(TFP_j)$		0.0594^{***} (0.0016) [8.90%]		0.0125^{***} $(3.31 \cdot 10^{-4})$ [2.36%]
Constant	1.3848^{***} (0.0224)	1.3245^{***} (0.0226)	0.6653^{***} (0.0047)	0.6485^{***} (0.0047)
AUC	0.6986	0.7042	—	—
R-squared	—	—	0.0593	0.0636

Table 6.7: *Adoption properties including the total factor productivity of industries.* For each functional form, in the first column we only include the change of TFP and in the second column we also include its level. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to a one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. j is the producing industry and i is the adopting industry.

First, we note that the changes in the probability of adoption due to a one standard deviation increase in TFP or in ΔTFP are significantly smaller than with the network proximity variables. We also note that when we only include the change in TFP, the estimate of the producing sector ΔTFP is not significant while when the level of TFP is also included as an explanatory variable it becomes significant and is even more significant than the ΔTFP of the adopting sector. In Table 6.12, we replicate

Table 6.7 but we only include manufacturing industries and we obtain significantly different results. The ΔTFP of the adopting sector is no longer significant. Moreover, we can compare the AUC value of the model including only the proximity variables with the one when we include the additional four explanatory variables related to the total factor productivity of industries. In the first case, it was equal to 0.6979 and with four more variables it only increases to 0.7034. This tells us that in general the model has not improved significantly when controlling for the productivity of industries. We can conclude that it is not a key driver to explain goods adoption in the sectoral input-output network.

Finally, we consider the broad regressions summarized in Table 6.8 that include all the variables already discussed plus a sector size variable.

We first notice that all the variables are significant. Moreover, the new explanatory variable shows that the size of industries is an important parameter both for the producing and the adopting sides. It is not surprising to see that it is even more important for the adopting industry. We have also introduced the growth rate of sectors as another explanatory variable but its effect is small therefore we do not include it in the table. The AUC value for the complete model is equal to 0.7653 and the associated receiver operating characteristic curve is shown in Figure 6.2. To emphasize our contribution in Figure 6.2 we are also showing the ROC curve of the complete model without the influence parameter and also without the trophic level of industries. The model 3 in Figure 6.2 corresponds to the model in Carvalho and Voigtländer (2014) and the associated AUC is equal to 0.6938. The regressions in Table 6.8 highlight the key drivers for new adoptions. As already mentioned, the network proximity is a major variable and the influence proximity parameter has a much larger contribution than the distance proximity parameter. The trophic level variable gives an important contribution to the regressions and it is interesting to relate that to the position of industries in the country production chain. In Table 6.13 we replicate the previous

Explanatory Variables	Probit		OLS	
	1	2	3	4
$\ln(d_{ji})$	-0.3705^{***} (0.0057) [−14.60%]	-0.3313^{***} (0.0059) [−13.07%]	-0.0649^{***} (0.0012) [−3.22%]	-0.0545^{***} (0.0012) [−2.71%]
$\ln(L_{ij})$	0.1128^{***} $(9.20 \cdot 10^{-4})$ [31.32%]	0.0995^{***} $(9.41 \cdot 10^{-4})$ [27.80%]	0.0224^{***} $(1.67 \cdot 10^{-4})$ [8.00%]	0.0188^{***} $(1.67 \cdot 10^{-4})$ [6.74%]
$\mathcal{L}_i$	-0.1747^{***} (0.0056) [−7.06%]	-0.1501^{***} (0.0057) [−6.07%]	-0.0367^{***} (0.0012) [−1.86%]	-0.0295^{***} (0.0011) [−1.50%]
$\mathcal{L}_j$	-0.5197^{***} (0.0055) [−22.28%]	-0.4946^{***} (0.0056) [−21.23%]	-0.1113^{***} (0.0011) [−6.06%]	-0.1020^{***} (0.0011) [−5.55%]
ΔTFP_i	$6.06 \cdot 10^{-4***}$ $(6.38 \cdot 10^{-5})$ [2.02%]	$5.26 \cdot 10^{-4***}$ $(6.35 \cdot 10^{-5})$ [1.75%]	$1.24 \cdot 10^{-4***}$ $(1.45 \cdot 10^{-5})$ [0.52%]	$9.89 \cdot 10^{-5***}$ $(1.43 \cdot 10^{-5})$ [0.41%]
ΔTFP_j	0.0012^{***} $(6.61 \cdot 10^{-5})$ [3.65%]	0.0011^{***} $(6.59 \cdot 10^{-5})$ [3.58%]	$2.48 \cdot 10^{-4***}$ $(1.52 \cdot 10^{-5})$ [0.99%]	$2.38 \cdot 10^{-4***}$ $(1.50 \cdot 10^{-5})$ [0.95%]
$\ln(TFP_i)$	-0.0137^{***} (0.0016) [−2.01%]	-0.0224^{***} (0.0016) [−3.29%]	-0.0023^{***} $(3.32 \cdot 10^{-4})$ [−0.42%]	-0.0048^{***} $(3.28 \cdot 10^{-4})$ [−0.88%]
$\ln(TFP_j)$	0.0386^{***} (0.0016) [5.80%]	0.0300^{***} (0.0016) [4.51%]	0.0099^{***} $(3.26 \cdot 10^{-4})$ [1.86%]	0.0076^{***} $(3.22 \cdot 10^{-4})$ [1.44%]
Sector size i		$2.39 \cdot 10^{-8***}$ $(3.91 \cdot 10^{-10})$ [13.87%]		$7.81 \cdot 10^{-9***}$ $(8.13 \cdot 10^{-11})$ [5.72%]
Sector size j		$7.01 \cdot 10^{-9***}$ $(2.41 \cdot 10^{-10})$ [5.94%]		$2.23 \cdot 10^{-9***}$ $(5.72 \cdot 10^{-11})$ [2.37%]
Constant	2.8062^{***} (0.0281)	2.3262^{***} (0.0293)	0.9299^{***} (0.0055)	0.7918^{***} (0.0057)
AUC	0.7487	0.7653	—	—
R-squared	—	—	0.0972	0.1260

Table 6.8: *Adoption properties with all explanatory variables.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. j is the producing industry and i is the adopting industry.

table but only including the manufacturing industries. Qualitatively we find similar results and both regressions provide similar AUC values.

So far, we have included all the adoptions without any exception. However, in Section 6.2.2 we have observed that some links appear before disappearing and reappearing again. This can introduce a bias as some links can be double counted. In Table 6.15, we redo the regressions but we only keep the adopted links that exist for at least more than 10 years. We perform the analysis both including all sectors and only the manufacturing sectors. We only focus on the proximity variables though we also include a column where we control for all the variables listed in Table 6.8. With this new criterion we observe 19,997 adoptions in total and 9,667 when studying only the manufacturing industries. We find that the influence proximity has an even bigger contribution to explain the probability of adoptions, however the explanatory power of the distance proximity has dropped. Overall, the AUC value of the regressions is higher than before.

Finally, we also control for the size of the adoption. This approach is more arbitrary than the previous one that was controlling for how long a new good will be used in the production. Indeed a good can spread progressively. Thus, the original size of the new link can be small but growing over time. At the opposite, one good can been adopted once in large quantity but it is not informative about whether the adoption was successful or not. However, for the sake of consistency in our comparison with Carvalho and Voigtländer (2014), we redo the analysis and we remove all the adoptions below 1 million dollars. In total, we now have 23,007 adoptions above the threshold and 10,774 when studying the manufacturing industries alone. The regressions are summarized in Table 6.16. We find that the influence proximity has the same contribution as before but now the distance proximity is much more important. Its contribution is still not as important but it has a major contribution. The AUC value of the regression has increased as well.

So far, we have only studied the adoption or not of a new good and we were answering a yes/no question. Now, we look at whether, on top of predicting a new adoption, we can also estimate the amplitude of the money flows. For this purpose, we can only use the OLS regression as the Probit model is not longer appropriate. We decompose the regression analysis in two categories. In the first type of regressions we include all the links and we treat the absent links as a zero money flow. So to prevent any outlier to dominate the regressions we regress $\ln(1+M_{ij})$ against the set of regressors, where M_{ij} is the money flow from j to i, instead of $P(a_{ji}(t)>0|a_{ji}(t-\Delta t)=0)$ to represent the adoption by i of a good produced by j.[11] The second type of regression is then conditional on adoptions. The analysis now becomes: given that we have observed a new adoption from i to j we regress $\ln(1+M_{ij})$ against the set of regressors. The regressions are summarized in Table 6.9. We note that while some variables are important to explain the adoption of a new good, they are no longer significant to explain the amplitude of the money flow conditional on the adoption. This is the case for the level of TFP of the producing industry, $\ln(TFP_j)$. Moreover, we find that some estimates have opposite signs when we studied the yes/no problem of the adoption compared to when we regress the amplitude of the new money flows. For instance, we find that the estimate of the trophic level of adopting industries is negatively related to the existence of new adoptions, meaning that industries with small trophic level are more likely to adopt a new good. But on the other hand, we find that conditional on the adoption, the amplitude of the new money flows is positively correlated with the trophic level of the adopting industry. This suggests that new adoptions by industries with a high trophic level are bigger in size compared to other adoptions. Nonetheless, we find that in all cases the network proximity variables are both significant. The more influence an industry has or the closer the industry is, the

[11] The trick of adding 1 is to prevent any problem when M_{ij} is equal to 0 and it sets the minimum value to 0. We have checked that using 0.01, 0.1, or 10 instead of 1 does not change the results significantly.

Explanatory Variables	All industries		Manufacturing only	
	1	2	3	4
$\ln(d_{ji})$	-0.5136^{***}	-0.7881^{***}	-0.5637^{***}	-1.6513^{***}
	(-58.14)	(-50.10)	(-57.44)	(-64.75)
$\ln(L_{ij})$	0.1450^{***}	0.1248^{***}	0.0925^{***}	0.2352^{***}
	(117.19)	(52.06)	(70.35)	(60.57)
$\mathcal{L}_i$	-0.1487^{***}	0.2049^{***}	-0.1426^{***}	0.1763^{***}
	(-17.74)	(10.54)	(-16.52)	(6.70)
$\mathcal{L}_j$	-0.7299^{***}	-0.3349^{***}	-0.6575^{***}	-0.5829^{***}
	(-92.06)	(-20.58)	(-79.51)	(-25.00)
ΔTFP_i	$8.48 \cdot 10^{-4***}$	$1.81 \cdot 10^{-3***}$	$-1.37 \cdot 10^{-3***}$	$-2.85 \cdot 10^{-3**}$
	(8.10)	(8.77)	(-3.57)	(-2.21)
ΔTFP_j	$1.66 \cdot 10^{-3***}$	$6.15 \cdot 10^{-5}$	$3.26 \cdot 10^{-3***}$	-0.0106^{***}
	(15.17)	(0.19)	(8.67)	(-10.40)
$\ln(TFP_i)$	-0.0219^{***}	0.1022^{***}	$-9.27 \cdot 10^{-3***}$	0.0924^{***}
	(-9.12)	(18.23)	(-3.37)	(10.81)
$\ln(TFP_j)$	0.0546^{***}	$8.31 \cdot 10^{-6}$	0.0324^{***}	-0.0756^{***}
	(23.16)	(0.00)	(11.92)	(-8.93)
Sector size i	$7.02 \cdot 10^{-8***}$	$2.74 \cdot 10^{-8***}$	$7.28 \cdot 10^{-8***}$	$-1.82 \cdot 10^{-9}$
	(117.95)	(39.89)	(58.42)	(-1.01)
Sector size j	$2.27 \cdot 10^{-8***}$	$2.78 \cdot 10^{-8***}$	$3.57 \cdot 10^{-8***}$	$2.40 \cdot 10^{-8***}$
	(54.31)	(40.35)	(32.96)	(11.51)
Constant	5.9539^{***}	10.4850^{***}	5.1662^{***}	15.08^{***}
	(143.48)	(108.8)	(119.26)	(97.00)
R-squared	0.151	0.163	0.097	0.202

Table 6.9: *Analysis of the amplitude of the money flows during adoptions.* In the first two columns we include all industries while in the last two columns we only include Manufacturing industries. Columns 1 and 3 includes all the potential links while regressions 2 and 4 are regressions of the money flows conditional on the fact that the good has been adopted already. In parenthesis, we quote the t-statistic. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. j is the producing industry and i is the adopting industry.

bigger the money flow.

These observations suggest that the mechanism explaining the adoptions of new inputs and the mechanism describing the amplitude of the adoption are different. Yet, they share in common some properties as the network proximity but the trophic

level of industries is not as key as the influence parameter.

6.4 Conclusion

In this chapter, we studied the adoptions of new products in the production chain and for this purpose we study the linkage dynamic of the input-output network. We build our work on the study realised by Carvalho and Voigtländer (2014). While new adoptions are likely to be influenced by some network effects, there are uncertainties about how to capture these effects. In Carvalho and Voigtländer (2014), the authors approximate the proximity between two nodes using the distance given by the shortest path. Instead, we use another approach that takes into account all the possible paths in the network as we believe it is more reflective of the high density of the sectoral input-output network. It turns out that the variable introduced has significantly more explanatory power than the network distance between nodes. Nonetheless, we have highlighted that they are complementary as the distance variable remains significant.

In a second analysis, we show that the position of industries along the production chain is also an important variable when studying the probability of adoption of a new good. We captured this effect by introducing the trophic levels of the producing and the adopting industries in the regression. Indeed, we find that goods that are at the bottom of the production chain are more likely to be adopted than goods that are at the top of the production chain. Moreover, we have found that the total factor productivity or its change do not have any major contribution in the regressions.

In some future work we could extend this analysis to study the adoption at the firm level and check that the key parameters that we have identified remain the major driver. Finally, we can imagine to couple this analysis with a model of growing networks with preferential attachment. In this case, we would not only model the structure of the network but also the weights on the links.

Appendix

6.A Additional figures and tables

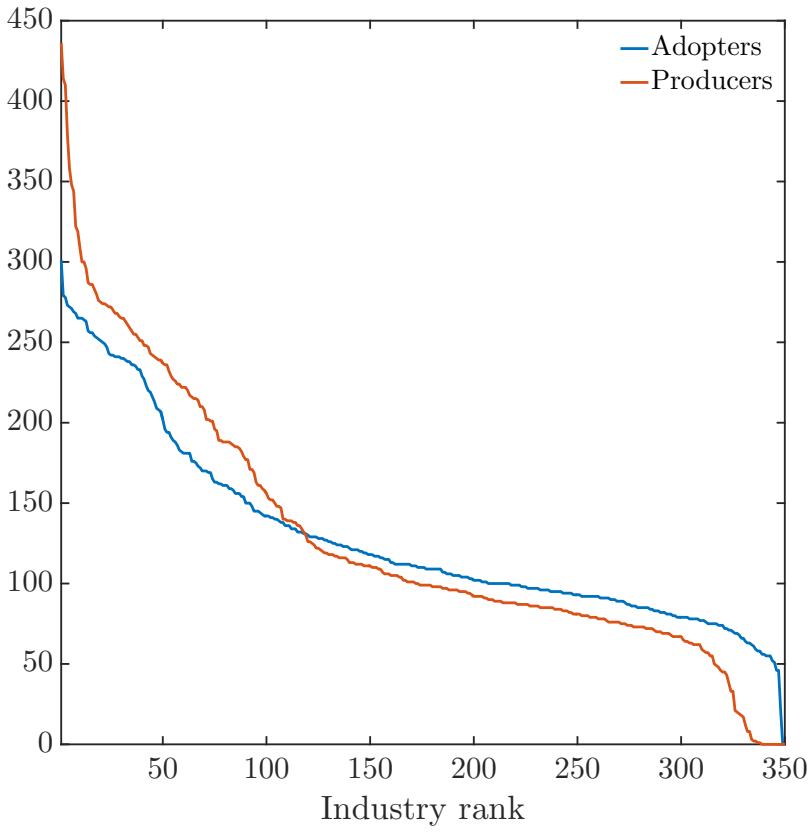


Figure 6.3: *Rank-size distribution of the number of adoptions over the entire sample period for adopters and producers.*

Explanatory Variables	**Probit**			**OLS**		
	1	2	3	4	5	6
$\ln(d_{ji})_{1972}$	-0.1667 (0.0033) [-11.81%]		-0.0848 (0.0043) [-6.02%]	-0.0392 ($7.04 \cdot 10^{-4}$) [-3.49%]		-0.0244 ($8.41 \cdot 10^{-4}$) [-2.18%]
$\ln(L_{ij})_{1972}$		0.0517 ($9.55 \cdot 10^{-4}$) [14.06%]	0.0361 (0.0012) [9.85%]		0.0105 ($1.83 \cdot 10^{-4}$) [3.59%]	0.0070 ($2.19 \cdot 10^{-4}$) [2.40%]
Constant	-0.7686 (0.0101)	-0.7031 (0.0106)	-0.6210 (0.0111)	0.2226 (0.0022)	0.2208 (0.0021)	0.2556 (0.0024)
AUC	0.5863	0.5946	0.6136	—	—	—
R-squared	—	—	—	0.0130	0.0138	0.0173

Table 6.10: *Adoption mechanism using the initial values for the explanatory variables for manufacturing sectors only.* Comparisons of Table 6.5 with Manufacturing sectors only. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***.

Explanatory Variables	**Probit**	**OLS**
$\ln(d_{ji})$	-0.4527 (0.0077) [-16.15%]	-0.0662 (0.0014) [-2.98%]
$\ln(L_{ij})$	0.0939 (0.0012) [24.96%]	0.0137 $(1.82 \cdot 10^{-4})$ [4.65%]
$\mathcal{L}_i$	-0.1650 (0.0072) [-6.54%]	-0.0294 (0.0012) [-1.46%]
$\mathcal{L}_j$	-0.5873 (0.0073) [-24.04%]	-0.1021 (0.0012) [-5.32%]
Constant	2.8530 (0.0366)	0.7743 (0.0060)
AUC	0.7577	—
R-squared	—	0.0704

Table 6.11: *Adoption mechanism including the proximity variables and the trophic level for manufacturing sectors only.* Comparisons of Table 6.6 with Manufacturing sectors only. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***. j is the producing industry and i is the adopting industry.

Explanatory Variables	Probit		OLS	
	1	2	3	4
$\ln(d_{ji})$	-0.4606^{***} (0.0075) [−16.43%]	-0.4353^{***} (0.0076) [−15.54%]	-0.0769^{***} (0.0014) [−3.46%]	-0.0717^{***} (0.0014) [−3.23%]
$\ln(L_{ij})$	0.0936^{***} (0.0012) [24.88%]	0.0972^{***} (0.0012) [25.81%]	0.0149^{***} $(1.84 \cdot 10^{-4})$ [5.05%]	0.0154^{***} $(1.86 \cdot 10^{-4})$ [5.22%]
ΔTFP_i	$2.14 \cdot 10^{-4}$ $(2.91 \cdot 10^{-4})$ [0.20%]	$3.99 \cdot 10^{-4}$ $(3.00 \cdot 10^{-4})$ [0.37%]	$-3.84 \cdot 10^{-5}$ $(5.39 \cdot 10^{-5})$ [−0.04%]	$2.16 \cdot 10^{-5}$ $(5.59 \cdot 10^{-5})$ [0.02%]
ΔTFP_j	0.0017^{***} $(2.57 \cdot 10^{-4})$ [1.56%]	0.0040^{***} $(2.59 \cdot 10^{-4})$ [3.72%]	$2.98 \cdot 10^{-4***}$ $(5.26 \cdot 10^{-5})$ [0.35%]	$7.56 \cdot 10^{-4***}$ $(5.47 \cdot 10^{-5})$ [0.89%]
$\ln(TFP_i)$		0.0046^{**} (0.0022) [0.61%]		0.0012^{***} $(3.92 \cdot 10^{-4})$ [0.21%]
$\ln(TFP_j)$		0.0622^{***} (0.0021) [8.55%]		0.0113^{***} $(3.79 \cdot 10^{-4})$ [1.96%]
Constant	1.3366^{***} (0.0304)	1.2041^{***} (0.0309)	0.5415^{***} (0.0053)	0.5139^{***} (0.0054)
AUC	0.6946	0.7034	—	—
R-squared	—	—	0.0365	0.0402

Table 6.12: *Adoption mechanism including the proximity variables and the total factor productivity for manufacturing sectors only.* Comparisons of Table 6.7 with Manufacturing sectors only. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. j is the producing industry and i is the adopting industry.

Explanatory Variables	Probit		OLS	
	1	2	3	4
$\ln(d_{ji})$	-0.4379^{***} (0.0078) [−15.63%]	-0.4162^{***} (0.0080) [−14.87%]	-0.0625^{***} (0.0014) [−2.81%]	-0.0554^{***} (0.0014) [−2.50%]
$\ln(L_{ij})$	0.0958^{***} (0.0013) [25.44%]	0.0875^{***} (0.0013) [23.30%]	0.0140^{***} $(1.84 \cdot 10^{-4})$ [4.75%]	0.0120^{***} $(1.87 \cdot 10^{-4})$ [4.06%]
$\mathcal{L}_i$	-0.1600^{***} (0.0073) [−6.34%]	-0.1789^{***} (0.0074) [−7.09%]	-0.0289^{***} (0.0012) [−1.44%]	-0.0305^{***} (0.0012) [−1.52%]
$\mathcal{L}_j$	-0.5719^{***} (0.0073) [−23.43%]	-0.5609^{***} (0.0074) [−23.00%]	-0.1012^{***} (0.0012) [−5.27%]	-0.0978^{***} (0.0012) [−5.10%]
ΔTFP_i	$-8.42 \cdot 10^{-5}$ $(3.24 \cdot 10^{-4})$ [−0.08%]	$-4.22 \cdot 10^{-4}$ $(3.36 \cdot 10^{-4})$ [−0.39%]	$-6.26 \cdot 10^{-5}$ $(5.50 \cdot 10^{-5})$ [−0.07%]	$-1.03 \cdot 10^{-4\star}$ $(5.46 \cdot 10^{-5})$ [−0.12%]
ΔTFP_j	0.0045^{***} $(2.70 \cdot 10^{-4})$ [4.27%]	0.0042^{***} $(2.74 \cdot 10^{-4})$ [3.97%]	$8.66 \cdot 10^{-4***}$ $(5.38 \cdot 10^{-5})$ [1.02%]	$8.09 \cdot 10^{-4***}$ $(5.34 \cdot 10^{-5})$ [0.95%]
$\ln(TFP_i)$	$0.0038^{\star}$ (0.0022) [0.51%]	-0.0210^{***} (0.0023) [−2.79%]	0.0017^{***} $(3.85 \cdot 10^{-5})$ [0.28%]	-0.0034^{***} $(3.91 \cdot 10^{-4})$ [0.57%]
$\ln(TFP_j)$	0.0393^{***} (0.0022) [5.41%]	0.0218^{***} (0.0023) [3.00%]	0.0094^{***} $(3.74 \cdot 10^{-5})$ [1.62%]	0.0056^{***} $(3.86 \cdot 10^{-4})$ [0.97%]
Sector size i		$3.76 \cdot 10^{-8***}$ $(9.38 \cdot 10^{-10})$ [10.74%]		$1.02 \cdot 10^{-8***}$ $(1.77 \cdot 10^{-10})$ [3.67%]
Sector size j		$1.60 \cdot 10^{-8***}$ $(7.71 \cdot 10^{-10})$ [5.38%]		$4.58 \cdot 10^{-9***}$ $(1.54 \cdot 10^{-10})$ [1.94%]
Constant	2.7213^{***} (0.0374)	2.5061^{***} (0.0381)	0.7476^{***} (0.0061)	0.6784^{***} (0.0062)
AUC	0.7600	0.7670	—	—
R-squared	—	—	0.0733	0.0896

Table 6.13: *Adoption mechanism including the all variables for manufacturing sectors only.* Comparisons of Table 6.8 with Manufacturing sectors only. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. j is the producing industry and i is the adopting industry.

Explanatory Variables	**Probit**				**OLS**			
$\ln(d_{ji})$	-0.2853 (0.0076) [-10.22%]		-0.4597 (0.0075) [-16.40%]	-0.4747 (0.0077) [-16.93%]	-0.0653 (0.0014) [-2.94%]		-0.0768 (0.0014) [-3.46%]	-0.0817 (0.0014) [-3.68%]
$\ln(L_{ij})$		0.0742 (0.0011) [19.85%]	0.0938 (0.0012) [24.92%]	0.0962 (0.0012) [25.55%]		0.0139 $(1.84 \cdot 10^{-4})$ [4.72%]	0.0150 $(1.84 \cdot 10^{-4})$ [5.07%]	0.0158 $(1.91 \cdot 10^{-4})$ [5.34%]
Constant	-0.3335 (0.0247)	-0.3673 (0.0134)	1.3390 (0.0304)	1.3017 (0.0304)	0.3179 (0.0046)	0.2781 (0.0024)	0.5420 (0.0053)	0.5372 (0.0053)
FYE				$\surd$				$\surd$
AUC	0.6439	0.6101	0.6937	0.6937	—	—	—	—
R-squared	—	—	—	—	0.0092	0.0237	0.0364	0.0373

Table 6.14: *Adoption mechanism including the proximity variables for manufacturing sectors only.* Comparisons of Table 6.3 and Table 6.4 with Manufacturing sectors only. Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***. j is the producing industry and i is the adopting industry.

All sectors

Explanatory Variables	Probit				OLS			
$\ln(d_{ji})$	-0.1164 (0.0070) $[-4.61\%]$		-0.3410 (0.0069) $[-13.45\%]$	-0.3237 (0.0074) $[-12.77\%]$	-0.0178 $(8.48 \cdot 10^{-4})$ $[-0.89\%]$		-0.0304 $(8.35 \cdot 10^{-4})$ $[-1.51\%]$	-0.0254 $(8.37 \cdot 10^{-4})$ $[-1.26\%]$
$\ln(L_{ij})$		0.1051 (0.0010) $[29.30\%]$	0.1218 (0.0011) $[33.67\%]$	0.1023 (0.0012) $[28.53\%]$		0.0137 $(1.15 \cdot 10^{-4})$ $[4.91\%]$	0.0143 $(1.16 \cdot 10^{-4})$ $[5.10\%]$	0.0103 $(1.17 \cdot 10^{-4})$ $[3.68\%]$
Constant	-1.1670 (0.0226)	-0.3594 (0.0116)	0.9109 (0.0279)	1.1676 (0.0373)	0.1194 (0.0028)	0.2294 (0.0015)	0.3343 (0.0032)	0.3169 (0.0039)
Control parameters				$\surd$				$\surd$
AUC	0.6132	0.7034	0.7501	0.8020	—	—	—	—
R-squared	—	—	—	—	0.0014	0.0418	0.0456	0.1140

Manufacturing sectors only

Explanatory Variables	Probit				OLS			
$\ln(d_{ji})$	-0.1684 (0.0101) $[-6.05\%]$		-0.4290 (0.0097) $[-15.32\%]$	-0.4162 (0.0102) $[-14.87\%]$	-0.0212 $(9.08 \cdot 10^{-4})$ $[-0.95\%]$		-0.0284 $(9.02 \cdot 10^{-4})$ $[-1.28\%]$	-0.0250 $(9.21 \cdot 10^{-4})$ $[-1.13\%]$
$\ln(L_{ij})$		0.0941 (0.0015) $[25.00\%]$	0.1208 (0.0017) $[31.76\%]$	0.1119 (0.0018) $[29.52\%]$		0.0090 $(1.19 \cdot 10^{-4})$ $[3.04\%]$	0.0094 $(1.20 \cdot 10^{-4})$ $[3.17\%]$	0.0074 $(1.23 \cdot 10^{-4})$ $[2.51\%]$
Constant	-1.1926 (0.0328)	-0.6483 (0.0169)	1.0242 (0.0402)	1.6694 (0.0508)	0.1103 (0.0030)	0.1531 (0.0015)	0.2506 (0.0035)	0.2686 (0.0041)
Control parameters				$\surd$				$\surd$
AUC	0.6276	0.6638	0.7471	0.7917	—	—	—	—
R-squared	—	—	—	—	0.0023	0.0235	0.0276	0.0579

Table 6.15: *Analysis of the adoption mechanism for links existing at least 10 years.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***.

All sectors

Explanatory Variables	Probit				OLS			
$\ln(d_{ji})$	-0.3169 (0.0067) [-12.51%]		-0.5018 (0.0067) [-19.69%]	-0.4474 (0.0071) [-17.59%]	-0.0555 $(9.01 \cdot 10^{-4})$ [-2.76%]		-0.0681 $(8.89 \cdot 10^{-4})$ [-3.38%]	-0.0511 $(8.93 \cdot 10^{-4})$ [-2.54%]
$\ln(L_{ij})$		0.0909 (0.0010) [25.48%]	0.1143 (0.0011) [31.72%]	0.1046 (0.0012) [29.15%]		0.0132 $(1.24 \cdot 10^{-4})$ [4.71%]	0.0143 $(1.24 \cdot 10^{-4})$ [5.12%]	0.0113 $(1.25 \cdot 10^{-4})$ [4.04%]
Constant	-0.4603 (0.0214)	-0.4320 (0.0113)	1.4177 (0.0267)	1.9314 (0.0359)	0.2507 (0.0030)	0.2317 (0.0016)	0.4666 (0.0034)	0.4656 (0.0042)
Control parameters				$\surd$				$\surd$
AUC	0.6854	0.6614	0.7471	0.8161	—	—	—	—
R-squared	—	—	—	—	0.0115	0.0336	0.0508	0.1140

Manufacturing sectors only

Explanatory Variables	Probit				OLS			
$\ln(d_{ji})$	-0.4700 (0.0095) [-16.76%]		-0.6953 (0.0095) [-24.58%]	-0.6773 (0.0100) [-23.93%]	-0.0632 $(9.49 \cdot 10^{-4})$ [-2.84%]		-0.0700 $(9.44 \cdot 10^{-4})$ [-3.15%]	-0.0607 $(9.67 \cdot 10^{-4})$ [-2.73%]
$\ln(L_{ij})$		0.0790 (0.0014) [21.10%]	0.1200 (0.0017) [31.55%]	0.1179 (0.0018) [31.03%]		0.0079 $(1.26 \cdot 10^{-4})$ [2.69%]	0.0089 $(1.25 \cdot 10^{-4})$ [3.00%]	0.0075 $(1.29 \cdot 10^{-4})$ [2.54%]
Constant	-0.1906 (0.0302)	-0.7591 (0.0166)	1.8993 (0.0389)	2.7872 (0.0498)	0.2523 (0.0031)	0.1447 (0.0016)	0.3854 (0.0036)	0.4166 (0.0043)
Control parameters				$\surd$				$\surd$
AUC	0.7151	0.6401	0.7741	0.8210	—	—	—	—
R-squared	—	—	—	—	0.0185	0.0165	0.0390	0.0643

Table 6.16: *Analysis of the adoption mechanism for new links higher than 1 million dollars.* Standard errors are in parenthesis. The values in square brackets reflect the change in adoption probability due to one standard deviation increase in the explanatory variable. All estimates are statistically significant, p-value< 0.01, so for clarity we do not include any ***.

6.B List of sectors

We list the 350 sectors represented in the input-output table. We have also highlighted the broader categories. The 4 digits number of industries is a classification used the BEA and has a direct connection to the SIC classification; the reference year for the classification is 1977.

Code	4 digits number	Name
		AGRICULTURE, FORESTRY, AND FISHERIES
1	0101	Dairy farm products
2	0102	Poultry & eggs
3	0103	Meat, animals
4	0201	Cotton
5	0202	Food grains
6	0203	Tobacco
7	0204	Fruits
8	0205	Vegetables
9	0206	Oil bearing crops
10	0207	Forest products
11	0300	Forestry & fishery products
12	0400	Agricultural, forestry & fishery services
		MINING
13	0500	Iron & ferroalloy ores mining
14	0601	Copper ore mining
15	0602	Nonferrous metal ores mining, except copper
16	0700	Coal mining
17	0800	Crude petroleum & natural gas
18	0900	Stone & clay mining & quarrying
19	1000	Chemical & fertilizer mineral mining
		CONSTRUCTION
20	1100	New residential
21	1106	New petroleum and natural gas well drilling
22	1202	Maintenance & repair
		MANUFACTURING
23	1301	Complete guided missiles
24	1302	Ammunition, exc. for small arms
25	1303	Tanks & tank components
26	1305	Small arms
27	1306	Small arms ammunition
28	1307	Other ordnance & accessories
29	1401	Meat packing plants
30	1402	Creamery butter

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
31	1403	Cheese, natural & processed
32	1404	Condensed & evaporated milk
33	1405	Ice cream & frozen desserts
34	1406	Fluid milk
35	1407	Canned & cured sea foods
36	1408	Canned specialties
37	1409	Canned fruits & vegetables
38	1410	Dehydrated food products
39	1411	Pickles, sauces, & salad dressing
40	1412	Fresh or frozen packaged fish
41	1413	Frozen fruits & vegetables
42	1414	Flour & other grain mill products
43	1415	Dog, cat, and other pet food
44	1416	Rice milling
45	1417	Wet corn milling
46	1418	Bread, cake, & related products
47	1419	Sugar
48	1420	Confectionery products
49	1421	Malt liquors
50	1422	Bottled & canned soft drinks
51	1423	Flavoring extracts & syrups
52	1424	Cottonseed oils mills
53	1425	Soybean oil mills
54	1426	Vegetable oil mills, n.e.c.
55	1427	Animal & marine fats & oils
56	1428	Roasted coffee
57	1429	Shortening & cooking oils
58	1430	Manufacturing ice
59	1431	Macaroni & spaghetti
60	1432	Food preparations, n.e.c.
61	1501	Cigarettes
62	1502	Tobacco stemming & redrying
63	1601	Broad woven fabric mills & fabric finishing plants
64	1602	Narrow fabric mills
65	1603	Yarn mills & finishing of textiles
66	1604	Thread mills
67	1701	Floor coverings
68	1706	Coated fabrics, not rubberized
69	1707	Tire cord & fabric
70	1709	Cordage & twine
71	1710	Nonwoven fabrics

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
72	1711	Textile goods, n.e.c.
73	1801	Women's hosiery, except socks
74	1802	Knit outerwear mills
75	1803	Knit fabric mills
76	1804	Apparel made from purchased materials
77	1901	Curtains & draperies
78	1902	Housefurnishing, n.e.c.
79	1903	Textile bags
80	2001	Logging camps, & logging contractors
81	2002	Sawmills & planning mills, general
82	2003	Hardwood dimension & flooring
83	2004	Special product sawmills, n.e.c.
84	2005	Millwork
85	2006	Veneer & plywood
86	2007	Structural wood members, n.e.c.
87	2008	Wood preserving
88	2009	Wood pallets and skids
89	2100	Wooden containers
90	2201	Wood household furniture
91	2202	Upholstered household furniture
92	2203	Metal household furniture
93	2204	Mattresses & bedsprings
94	2301	Wood office furniture
95	2302	Metal office furniture
96	2303	Public building furniture
97	2304	Wood partitions & fixtures
98	2305	Metal partitions & fixtures
99	2306	Venetian blinds & shades
100	2307	Furniture & fixtures, n.e.c.
101	2401	Pulp mills
102	2404	Envelopes
103	2405	Sanitary paper products
104	2407	Paper coating & glazing
105	2408	Paper and paperboard mills
106	2500	Paperboard containers & boxes
107	2601	Newspapers
108	2602	Periodicals
109	2603	Book publishing
110	2604	Miscellaneous publishing
111	2605	Commercial printing
112	2606	Manifold business forms

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
113	2607	Greeting card publishing
114	2608	Engraving & plate printing
115	2701	Industrial inorganic & organic chemicals
116	2702	Nitrogenous and phosphatic fertilizers
117	2703	Agricultural chemicals, n.e.c.
118	2704	Gum & wood chemicals
119	2801	Plastics materials & resins
120	2802	Synthetic rubber
121	2803	Cellulosic man-made fibers
122	2804	Organic fibers, noncellulosic
123	2901	Drugs
124	2902	Soap & other detergents
125	2903	Toilet preparations
126	3000	Paints & allied products
127	3101	Petroleum refining & related products
128	3102	Paving mixtures & blocks
129	3103	Asphalt felts & coatings
130	3201	Tires & inner tubes
131	3202	Rubber footwear
132	3203	Reclaimed rubber
133	3204	Miscellaneous plastics products
134	3205	Rubber and plastics hose and belting
135	3206	Gaskets, packing, and sealing devices
136	3300	Leather tanning & finishing
137	3401	Footwear cut stock
138	3402	Shoes, except rubber
139	3403	Leather gloves & mittens
140	3501	Glass & glass products exc. containers
141	3502	Glass containers
142	3601	Cement, hydraulic
143	3602	Brick & structural clay tile
144	3603	Ceramic wall & floor tile
145	3604	Clay refractories
146	3605	Structural clay products, n.e.c.
147	3606	Vitreous plumbing fixtures
148	3607	Vitreous china food utensils
149	3608	Porcelain electrical supplies
150	3609	Pottery products, n.e.c.
151	3610	Concrete block & brick
152	3611	Concrete products, n.e.c.
153	3612	Ready-mixed concrete

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
154	3613	Lime
155	3614	Gypsum products
156	3615	Cut stone & stone products
157	3616	Abrasive products
158	3617	Asbestos products
159	3619	Minerals, ground or treated
160	3620	Mineral wool
161	3621	Nonclay refractories
162	3622	Nonmetallic mineral products, n.e.c.
163	3701	Blast furnaces & steel mills
164	3702	Iron & steel foundries
165	3703	Iron & steel forgings
166	3704	Metal heat treating
167	3801	Primary copper
168	3804	Primary aluminum
169	3805	Primary nonferrous metals, n.e.c.
170	3806	Secondary nonferrous metals
171	3807	Copper rolling & drawing
172	3808	Aluminum rolling & drawing
173	3809	Nonferrous rolling & drawing, n.e.c.
174	3810	Nonferrous wire drawing & insulting
175	3811	Aluminum castings
176	3812	Brass, bronze, & copper castings
177	3813	Nonferrous castings, n.e.c.
178	3814	Nonferrous forgings
179	3901	Metal cans
180	3902	Metal barrels, drums, & pails
181	4001	Metal sanitary ware
182	4002	Plumbing fittings & brass goods
183	4003	Heating equipment, except electric
184	4004	Fabricated structural steel
185	4005	Metal doors, sash, & trim
186	4006	Fabricated plate work (boiler shops)
187	4007	Sheet metal work
188	4008	Architectural metal work
189	4009	Prefabricated metal buildings
190	4101	Screw machine products & bolts, nuts, rivets & washers
191	4102	Automotive stampings
192	4201	Cutlery
193	4202	Hand & edge tools, n.e.c.
194	4203	Hardware, n.e.c.

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
195	4204	Plating & polishing
196	4205	Miscellaneous fabricated wire products
197	4207	Steel springs
198	4208	Pipe, valves, & pipe fittings
199	4210	Metal foil & leaf
200	4211	Fabricated metal products, n.e.c.
201	4301	Steam engines & turbines
202	4302	Internal combustion engines, n.e.c.
203	4400	Farm machinery and equipment
204	4501	Construction machinery
205	4502	Mining machinery
206	4503	Oil field machinery
207	4601	Elevators & moving stairways
208	4602	Conveyors & conveying equipment
209	4603	Hoists, cranes, & monorails
210	4604	Industrial trucks & tractors
211	4701	Machine tools, metal cutting types
212	4702	Machine tools, metal forming types
213	4703	Special dies & tools & machine tool accessories
214	4704	Power driven hand tools
215	4705	Metalworking machinery, n.e.c.
216	4801	Food products machinery
217	4802	Textile machinery
218	4803	Woodworking machinery
219	4804	Paper industries machinery
220	4805	Printing trades machinery
221	4806	Special industry machine, n.e.c.
222	4901	Pumps & compressors
223	4902	Ball & roller bearings
224	4903	Blowers & fans
225	4905	Power transmission equipment
226	4906	Industrial furnaces & ovens
227	4907	General industrial machinery n.e.c.
228	4908	Packaging machinery
229	5001	Carburetors, pistons, rings, and valves
230	5002	Fluid power equipment
231	5003	Scales and balances, except laboratory
232	5004	Industrial and commercial machinery and equipment, n.e.c.
233	5101	Electronic computing equip.
234	5104	Office machines, n.e.c.
235	5201	Automatic merchandising machines

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
236	5202	Commercial laundry equipment
237	5203	Refrigeration machinery
238	5204	Measuring & dispensing pumps
239	5205	Service industry machinery, n.e.c.
240	5302	Transformers
241	5303	Switchgear & switchboard
242	5304	Motors & generators
243	5305	Industrial controls
244	5307	Carbon & graphite products
245	5308	Electrical industrial apparatus, n.e.c.
246	5401	Household cooking equipment
247	5402	Household refrigerators & freezers
248	5403	Household laundry equipment
249	5404	Electric housewares & fans
250	5405	Household vacuum cleaners
251	5407	Household appliances, n.e.c.
252	5501	Electric lamps
253	5502	Lighting fixtures
254	5503	Wiring devices
255	5601	Radio & TV receiving sets
256	5602	Phonograph records
257	5603	Telephone & telegraph apparatus
258	5605	Communication equipment
259	5701	Electron tubes
260	5702	Semiconductors
261	5703	Electronic components, n.e.c.
262	5801	Storage batteries
263	5802	Primary batteries, dry & wet
264	5804	Engine electrical equipment
265	5806	Magnetic and optical recording media
266	5807	Electrical machinery, equipment, and supplies, n.e.c.
267	5901	Truck & bus bodies
268	5902	Truck trailers
269	5903	Motor vehicles
270	6001	Aircraft
271	6002	Aircraft and missile engines & parts
272	6004	Aircraft and missile equipment, n.e.c.
273	6101	Shipbuilding & repairing
274	6102	Boat building & repairing
275	6103	Railroad equipment
276	6105	Motorcycles, bicycles, & parts

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
277	6106	Travel trailers and campers
278	6107	Transportation equipment, n.e.c.
279	6201	Engineering & scientific instruments
280	6202	Mechanical measuring devices
281	6203	Automatic temperature controls
282	6204	Surgical & medical instruments
283	6205	Surgical appliances & supplies
284	6206	Dental equipment & supplies
285	6207	Watches, clocks and parts
286	6208	X-ray apparatus and tubes
287	6209	Electromedical and electrotherapeutic apparatus
288	6210	Laboratory and optical instruments
289	6211	Instruments to measure electricity
290	6302	Ophthalmic goods
291	6303	Photographic equipment & supplies
292	6401	Jewelry, precious
293	6402	Musical instruments & parts
294	6403	Games, toys and childrens vehicles
295	6404	Sporting & athletic goods, n.e.c.
296	6405	Pens & mechanical pencils
297	6407	Buttons
298	6408	Brooms & brushes
299	6409	Hard surface floor covering
300	6410	Morticians goods
301	6411	Signs & advertising display
302	6412	Miscellaneous products, n.e.c.
		TRANSPORTATION, COMMUNICATIONS, AND UTILITIES
303	6501	Railroads & related services
304	6502	Local, suburban & interurban highway passenger transportation
305	6503	Motor freight transportation & warehousing
306	6504	Water transportation
307	6505	Air transportation
308	6506	Pipe line transportation
309	6507	Transportation services
310	6600	Communications, except radio & television
311	6700	Radio & television broadcasting
312	6801	Electric utilities
313	6802	Gas utilities
314	6803	Water & sanitary services
		WHOLESALE AND RETAIL TRADE
315	6901	Wholesale trade

Continued on next page

Table 6.17 – *Continued from previous page*

Code	4 digits number	Name
316	6902	Retail trade
		FINANCE, INSURANCE, AND REAL ESTATE
317	7001	Banking
318	7002	Credit agencies
319	7003	Security & commodity brokers
320	7004	Insurance carriers
321	7005	Insurance agents & brokers
322	7101	Owner-occupied dwellings
323	7102	Real estate
		SERVICES
324	7201	Hotels & lodging places
325	7202	Personal & repair services, except auto repair, barber, & beauty shops
326	7203	Barber & beauty shops
327	7301	Miscellaneous business services
328	7302	Advertising
329	7303	Miscellaneous professional services
330	7400	Eating and drinking places
331	7500	Automobile repair & services
332	7601	Motion pictures
333	7602	Amusement & recreation services
334	7701	Doctors & dentists
335	7702	Hospitals
336	7703	Other medical & health services
337	7704	Educational services
338	7705	Nonprofit organizations
339	7706	Job training and related services
340	7707	Child day care services
341	7708	Residential care
342	7709	Social services, n.e.c.
343	7801	Post office
344	7802	Federal electric utilities
345	7805	Other Federal Government enterprises
346	7901	Local government passenger transit
347	7902	State & Local electric utilities
348	7903	Other state & local government enterprises
349	8200	Government industry
350	8400	Household industry

Table 6.17: *List of sectors in the input-output network from the BEA.* The 4 digits number classification is related to the SIC and the year of reference is 1977.

Chapter 7

Common Asset Holdings and Stock Return Comovement

7.1 Introduction

A growing literature highlights the effect of investors' common asset holdings on both price and liquidity dynamics. For example, Koch et al. (2016) find that commonality in stock liquidity is likely driven by correlated trading among a given stock's investors. The authors show that stocks with high mutual fund ownership have liquidity comovements that are about twice as large as those for stocks with low mutual fund ownership. Regarding the impact on prices, Antón and Polk (2014) showed that stocks sharing many common investors tend to comove more strongly with each other in the future than otherwise similar stocks, even after controlling for various stock-pair specific characteristics. Similar to Koch et al. (2016), Antón and Polk (2014) explain their finding based on correlated trading strategies among mutual funds. Interestingly, Greenwood and Thesmar (2011) argue that if the investors of mutual funds have correlated trading needs, the stocks that are held by mutual funds can comove even without any portfolio overlap of the funds themselves. For the set of U.S. mutual funds, however, a significant portfolio overlap has been documented (Fricke, 2016), suggesting that common asset holdings are indeed of relevance in the study of market dynamics.

Note that Antón and Polk (2014) use data from the CRSP Mutual Fund database for the period 1980 - 2008, and restrict their analysis to relatively large stocks only.[1] While this sample selection approach is pervasive in the literature, it is well known that institutional preferences have changed over the last decades. For example, Falkenstein (1996) and Gompers and Metrick (2001) found that institutional investors tend to prefer holding large, liquid stocks. Later on, Bennett et al. (2003) compare institutional investments for the periods 1983-1990 and 1990-1997 and find that investors shifted their preferences towards smaller, riskier securities over this relatively short period. Indeed, we confirm that institutional ownership has increased rather dramatically over the last two decades for most stocks (see Table 7.1 and Figure 7.1), naturally leading us to ask whether the main results of Antón and Polk (2014) also hold when including smaller stocks to the analysis.[2] This is particularly interesting since nowadays small stocks' level of institutional ownership is comparable to that of large stocks in the 1990s. Hence, we would expect that small stocks' return comovements should be increasingly affected by common asset ownership of institutional investors. This is indeed what we find in the data.

In this chapter, we therefore perform a similar analysis as Antón and Polk (2014) for stocks from different size categories, including very small stocks. We use data for the period 1990 - 2014 for the broad set of U.S. institutional investors, and our main findings can be summarized as follows: first, we document a strong increase in both institutional ownership and common asset holdings over the sample period, particu-

[1] More precisely, they state that "*[w]e restrict our analysis to common stocks (share codes 10 and 11) from NYSE, Amex, and NASDAQ whose market capitalizations are above the NYSE median market cap (i.e., "big" stocks). We choose these screening criteria because common ownership by active managers is not pervasive - small stocks, especially in the beginning of the sample, have little institutional ownership. Limiting the data in this way also keeps the sample relatively homogeneous and ensures that the patterns we find are not just due to small or microcap stocks.*" (See Antón and Polk (2014, p. 1104))

[2] Koch et al. (2016) find that the positive relationship between the mutual fund liquidity beta and mutual fund ownership is strongest for the largest and most liquid stocks. The relationship becomes much weaker for smaller stocks and, in fact, insignificant for the smallest and most illiquid stocks, which are not predominantly held or traded by mutual funds.

larly so for the smallest stocks. Interestingly, the most illiquid stocks (Amihud-ratio) do not display a similar growth in institutional ownership, suggesting that liquidity remains an important characteristic for professional asset managers. Second, we show that raw stock return correlations have also increased strongly over time, while the increase is much more modest when looking at factor-model residual correlations. Third, we confirm that common ownership is significantly related to future return correlations for relatively large stocks, but the relationship appears to become less important over time. For relatively small stocks the relationship is often insignificant, but the effect appears to become more important over time. Lastly, we find no evidence that institutional ownership Granger-causes return correlations for the subset of small stocks. These findings are important because they suggest that smaller stocks, despite being more widely held by professional asset managers, are still relatively illiquid and thus are not generally liquidated in stress periods.

This chapter mainly adds to the literature on common asset holdings. To the best of our knowledge, this work is the first to explore the role of common ownership on stock return correlations for relatively small stocks. As such, it also contributes to the literature on institutional preferences by looking at the relationship with stock liquidity (Falkenstein (1996); Gompers and Metrick (2001); Bennett et al. (2003)). In a way, one might argue that, by being more widely held by a larger number of investors, small stocks should have become more liquid. Our findings suggest that the smallest stocks are not necessarily the most illiquid - in fact, we find that the level of institutional ownership has sharply increased for the smallest stocks, but much less so for the most illiquid stocks. It also adds to the growing literature that explores industry linkages via the input-output network. By adding input-output linkages between firms from different industries to our set of explanatory variables in the Antón-Polk regressions, we control more explicitly for fundamental factors that are likely to affect correlations between different stocks. Our main finding is that this

effect is largely significant, robust over time and it does not depend of the size of the stocks.

This chapter is structured as follows: First, we introduce the dataset and describe our main research methodology. Then we present the main empirical results. We also study the (Granger-)causality between institutional ownership, return correlation, and liquidity. Finally, we conclude and elaborate on interesting avenues of future work.

7.2 Data and Summary Statistics

In the following, we briefly introduce the dataset and provide some summary statistics that motivate our analysis.

7.2.1 Data

We combine data from different sources. First, the institutional holdings data come from Thomson Reuters (13F filings) and are updated quarterly. Under Securities Exchange Act Section 3(a)(9) and Section 13(f)(5)(A) all money managers/investment companies with assets under management exceeding $100 million have to submit detailed information on their holdings (number of shares and market valuation) to the SEC on a quarterly basis. The set of institutions includes banks, insurance companies, investment companies, independent investment advisors, and other types of institutions (such as pension funds and university endowments). The reported holdings are at the security level (CUSIPs) and generally constitute equities.[3] We complement the holdings data with the merged CRSP-Compustat files, providing us with daily stock market data, and quarterly accounting data. Lastly, we also have access to the

[3]Institutions are required to report all equity positions greater than 10,000 shares or $200,000 in market value. The data are not necessarily complete for two additional reasons: first, institutions can request exception from filing their reports in order to prevent disclosure of their trading strategies. Second, institutions only report their longterm positions.

CRSP Mutual Fund database from 2003 onwards and we will use the mutual fund data mainly in order to check the robustness of our results. Our regression analyses below will be performed on monthly data.

In line with the literature, we focus on common stocks (share codes 10 and 11) traded on NASDAQ, NYSE, or AMEX. Our final sample covers the period between January 1990 and December 2014 (300 months) and 13,934 unique stocks for which we have information in the merged CRSP-Compustat-Thomson dataset.

7.2.2 Dynamics of Institutional Ownership

Year	Stocks	Pairs	Institutions	InstOwn	FCAP
1990	4,755	11,302,635	954	0.267	0.052
1995	5,942	17,650,711	1,223	0.309	0.058
2000	5,913	17,478,828	1,791	0.333	0.081
2005	4,505	10,145,260	2,263	0.515	0.174
2010	3,791	7,183,945	2,639	0.581	0.244
2014	3,478	6,046,503	3,106	0.560	0.199
Mean	4,839	12,206,711	1,972	0.437	0.132
Median	4,717	11,122,686	1,926	0.425	0.114
SD	998	4,984,672	670	0.124	0.075
Min	3,381	5,713,890	954	0.267	0.038
Max	6,659	22,167,811	3,180	0.615	0.257

Table 7.1: *Summary Statistics on stocks and institutional investors.* The first part reports the number of stocks, pairs of stocks, number of institutions, the (cross-sectional) average institutional ownership ($InstOwn$), and our main measure of common investors ($FCAP$). A pair of stocks is two stocks that are held by at least one common investor. We show statistics for specific dates, and the second part summarizes the statistics over the entire sample period.

Table 7.1 provides some basic summary statistics of our dataset. For example, we see that the number of financial institutions has tripled over the sample period, but on the other hand, the number of stocks has been decreasing ever since bursting of the dot-com bubble. At the end of our sample period, there were roughly 3,100

reporting institutions and 3,500 stocks (compared with 954 institutions and 4,755 stocks in 1990). As a result, the number of stock pairs is quite large and changes over time: there is a factor 4 between the minimum and maximum number of stock pairs.[4] We also calculate the level of institutional ownership for each stock ($InstOwn$), which we define as

$$InstOwn_{j,t} = \frac{\sum_i S_{i,j,t}}{ShareOut_{j,t}}, \tag{7.1}$$

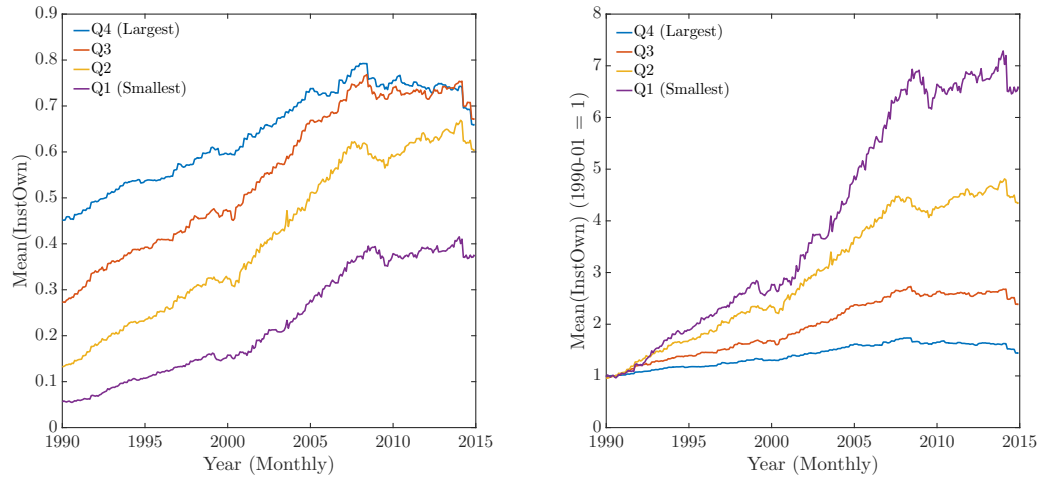
where $S_{i,j,t}$ is the number of shares of stock j held by institution i at time t, and $ShareOut_{j,t}$ is the number of outstanding shares of stock j. In Table 7.1, we show the cross-sectional average value of $InstOwn$ for different years. We find that it increases during the first half of the dataset and then settles around 0.55 afterwards. In order to explore, to what extent the average $InstOwn$ is representative for different sets of stocks, the top panels of Figure 7.1 show the average $InstOwn$ for different size quartiles, based on market capitalization ('Q1' are the smallest stocks, and 'Q4' the largest stocks). Not surprisingly, the largest stocks have the highest level of institutional ownership, and the smallest stocks have the lowest level of institutional ownership. In line with the findings of Bennett et al. (2003), however, we find that the level of institutional ownership in small stocks has increased quite rapidly between 2000 and 2005. Hence, institutional preferences indeed appear to have changed over time.[5]

The top right panel of Figure 7.1 shows evidence of this rapid growth by indexing the $InstOwn$ values in 1990 to 1; with this normalization it becomes clear that the level of institutional ownership for the smallest stocks has increased by a factor of 7 over our sample period. For the largest stocks, this relative increase is much weaker.

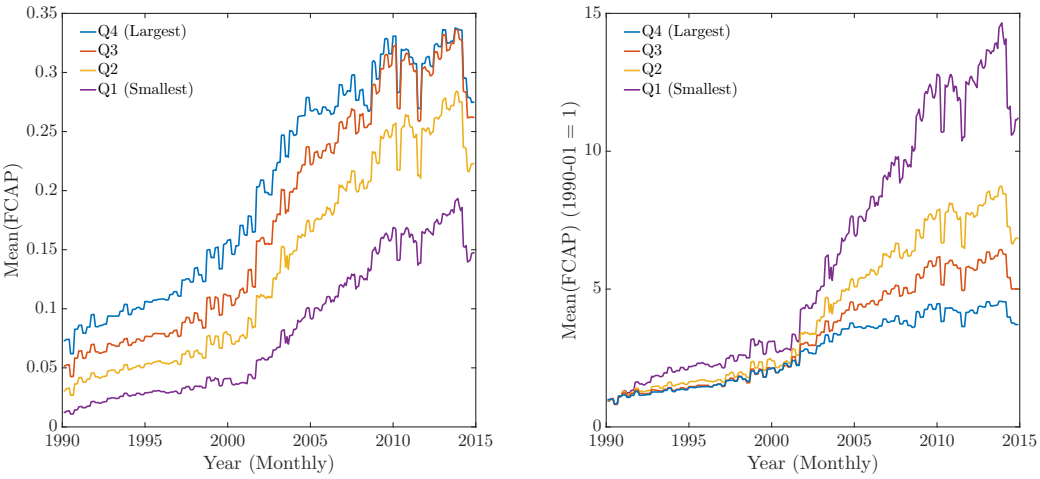
Finally, we define common institutional ownership between stocks j and k, $FCAP_{j,k}$,

[4] A pair of stocks is two stocks that are held by at least one common investor. The maximum number of stock pairs is $\frac{n(n-1)}{2}$, where n is the number of stocks.

[5] Early papers (Gompers and Metrick, 2001; Falkenstein, 1996) show that institutions used to prefer large and very liquid stocks.



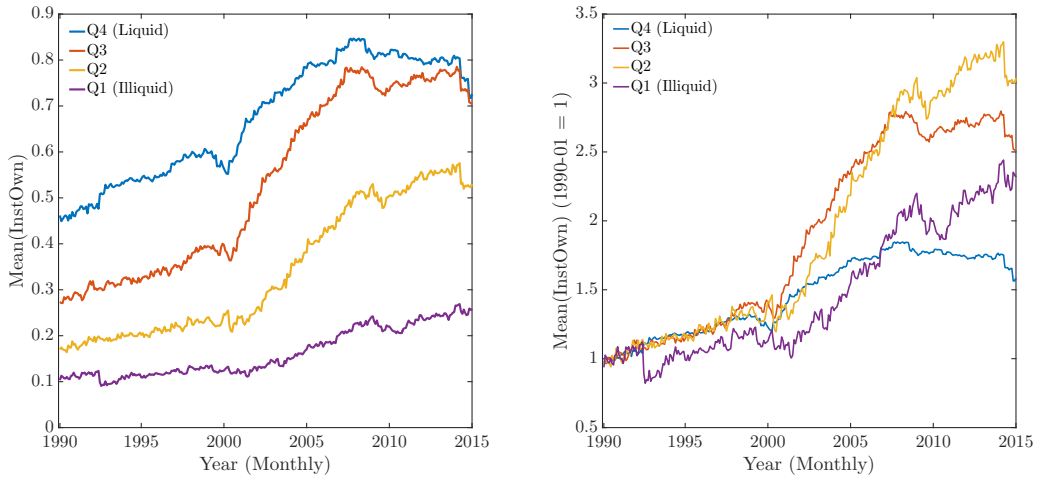
(a) *InstOwn* for different size quartiles (market capitalization).



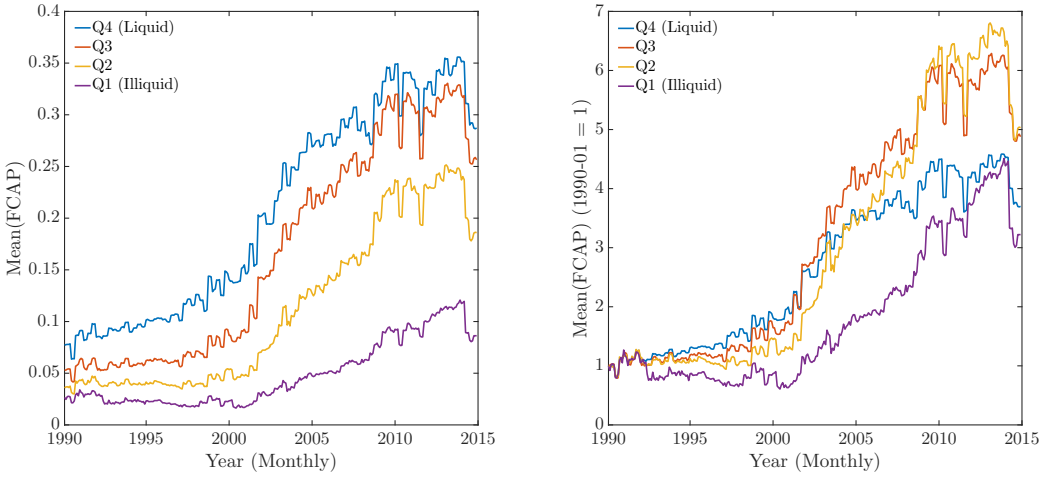
(b) *FCAP* for different size quartiles (market capitalization).

Figure 7.1: *Average InstOwn and FCAP by market capitalization quartiles over time.* Top: $InstOwn_j$ is defined as the sum of shares of stock j held by institutional investors relative to the number of outstanding shares. Bottom: $FCAP_{j,k}$ is defined as the relative share of these stocks that are held by investors that hold both of these stocks in their portfolios. The right panels normalize the absolute values for each category such that the initial values in 1990-01 are equal to 1.

as the total value of stocks held by all common funds of the two stocks, divided by



(a) Average $InstOwn$ by illiquidity (Amihud-ratio) quartiles over time.



(b) Average $FCAP$ by illiquidity (Amihud-ratio) quartiles over time.

Figure 7.2: *Average InstOwn and FCAP by illiquidity quartiles over time.* Top: $InstOwn_j$ is defined as the sum of shares of stock j held by institutional investors relative to the number of outstanding shares. Bottom: $FCAP_{j,k}$ is defined as the relative share of these stocks that are held by investors that hold both of these stocks in their portfolios. The right panels normalize the absolute values for each category such that the initial values in 1990-01 are equal to 1.

the total market capitalization of the two stocks

$$FCAP_{j,k,t} = \frac{\sum_{i=1}^{F}(S_{i,j,t}P_{j,t} + S_{i,k,t}P_{k,t})}{ShareOut_{j,t}P_{j,t} + ShareOut_{k,t}P_{k,t}}, \tag{7.2}$$

where F is the total number of common funds, and $P_{j,t}$ is the market price of stock j at date t. This will be the main variable of interest in our empirical application below. Interestingly, Table 7.1 shows that the average $FCAP$ appears to increase quite dramatically over our sample period, with values around 0.05 in 1990 and values closer to 0.20 at the end of the sample period. We also explore to what extent these averages are representative for the typical stock: the bottom left panel of Figure 7.1 shows the typical $FCAP$ values for different size quartiles (again based on market capitalization). In line with the results for $InstOwn$, we find that common asset ownership appears to be most important for the largest stocks. Interestingly, the average $FCAP$ for the smallest stocks (Q1) at the end of the sample period exceeds that of the largest stocks (Q4) at the beginning of the sample period. In other words, nowadays the smallest stocks have more common investors than large stocks in the early 1990s. Finally, the bottom right panel of Figure 7.1 normalizes the absolute values in the left panel, illustrating an even more dramatic growth in $FCAP$ for the smallest stocks. Overall, these results confirm that institutional ownership, and thus common asset holdings, has become much more important for relatively small stocks. From the analysis of Antón and Polk (2014) we would expect that small stocks should therefore become much more correlated with other stocks over time, and thus $FCAP$ should be an important explanatory variable for future stock return correlations.

Lastly, Figure 7.2 is structured similar to Figure 7.1, but sorts stocks based on their liquidity (using the Amihud-ratio, (Amihud, 2002)). Not surprisingly, more liquid stocks are generally held by more institutional investors. Most of the growth in institutional ownership, however, was concentrated in the relatively liquid stocks. The change in institutional ownership in both the most liquid and the least liquid stocks has been much less impressive. The result for the latter set of stocks thus suggests that the smallest stocks are not necessarily the most illiquid (see Figure 7.1).

7.2.3 Stock Return Correlations

We are interested in whether future stock return correlations can be explained via common stock ownership ($FCAP$). The next section will be dedicated to explaining our methodology in more detail. Before doing so, let us briefly document the dynamics of our dependent variable, namely stock return correlations.

In this regard, Figure 7.3 shows the average within-category return correlations of stocks from different size quartiles.[6] The left panel shows the results for raw returns, and the right panel shows the same results for the correlations based on 4-factor residuals, noted ϵ,

$$(r_{j,t} - r_t^f) = \alpha_j + \beta_{j,M}(r_t^M - r_t^f) + \beta_{j,SMB} SMB_t + \beta_{j,HML} HML_t + \beta_{j,MOM} MOM_t + \epsilon_{j,t}, \tag{7.3}$$

where $(r_t^M - r_t^f)$ denotes the excess return of the market portfolio over the risk-free rate, SMB is the return difference between small and large market capitalization stocks, HML is the return difference between high and low book-to-market ratio stocks, and MOM is the return difference between stocks with high and low past returns.[7] In both cases, we smoothen the estimates using a 12-month moving average.[8]

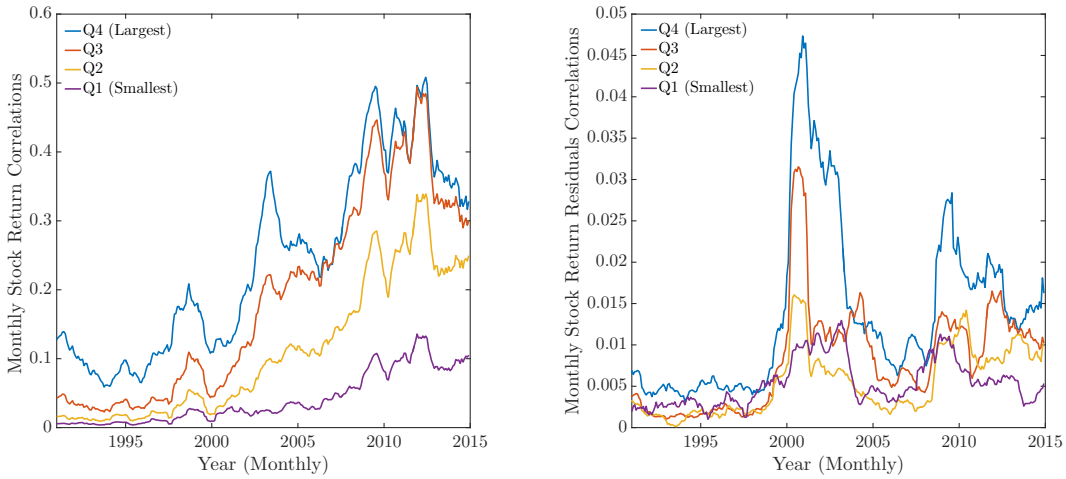
From the raw correlations (left panel) it becomes clear that larger stocks are more correlated among each other compared with smaller stocks. Moreover, we also see that return correlations have increased over our sample period, irrespective of the size of the stocks.[9] On the other hand, when we look at 4-factor residual correlations (right panel), the level of the estimated correlations is much smaller in comparison. As for the raw correlations, the level of comovement among the largest stocks is

[6] In Section 7.4.2, we also explore between-category correlations in more detail.

[7] Data source: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library`.

[8] Figure 7.6 in the Appendix shows a similar Figure using 3-factor residuals (excluding the Momentum-factor).

[9] For example, a report by J.P. Morgan from 2010 ("Why We Have a Correlation Bubble", `https://www.cboe.com/institutional/jpmderivativesthemescorrelation.pdf`) argues that high-frequency trading strategies are the most likely source of this increase in correlations.



(a) Correlations of raw returns.

(b) Correlations of 4-factor residuals.

Figure 7.3: *12-month moving average of the monthly stock return correlations within stocks of similar size.* Left panel: average correlations based on raw returns. Right panel: average correlations based on 4-factor residuals.

the largest in general. However, the smallest stocks are not necessarily the least correlated. Furthermore, while the values appear to have increased over time as well for all categories, there is a much weaker time trend in the typical residual correlations. Lastly, we observe some peaks during stress periods for which the correlations can be very large (Pollet and Wilson, 2010) - in fact, for the largest stocks the values during the bursting of the dot-com bubble are almost 10 times as large as the pre-crisis values.

7.3 Methodology

Let us explain the methodology to test our hypotheses in detail. We follow the approach of Antón and Polk (2014) as closely as possible and are mainly interested in the following monthly cross-sectional regressions

$$\rho_{i,j,t+1} = \alpha_t + \beta_{1,t} FCAP^*_{i,j,t} + \sum_{k=2}^{n} \beta_{k,t} CONTROL_{i,j,k,t} + \epsilon_{i,j,t+1}, \tag{7.4}$$

where $\rho_{i,j,t+1}$ is the daily realized stock return correlations within month $t+1$, α_t is the intercept, and $\beta_{k,t}$ is the estimate of the variable k and $CONTROL_{i,j,k,t}$ is a set of control variables. Hence, we want to explain future stock return comovement based on some fundamental factors and common stock ownership. We use the Fama-MacBeth methodology in everything that follows, and thus show time-averages of the above parameter estimates below. Given the trends in the return correlations documented in the previous section, we use the 4-factor residual correlations in everything that follows.

Furthermore, we normalize all explanatory variables to have zero mean and unit standard deviation in order to make the coefficients comparable over time. In the regression tables, all normalized rank-transformed variables are identified with a star $(*)$.

7.3.1 Control Variables

Our set of control variables is as follows: first, we expect stocks from similar industries to comove more strongly with each other. Here we compute the number of consecutive $NAICS$ digits ($NUMNAICS$), beginning with the first digit, that are equal for a given pair. Note that this is a very crude measure of fundamental linkages between firms from different industries, and we suspect that this measure is not sufficient to capture the complex interaction between firms. Ideally we would want to include input-output relationships between different firms, but clearly such data are not easily available. Therefore, we approximate fundamental linkages based on the input-output tables (US Bureau of Economic Analysis, 2015), where we map each stock to one of the 376 industries listed in the USA input-output network. We expect that firms from industries with strong input-output relationships should comove more strongly with each other. Because the input-output network is a directed network for any pair of industries we have two money flows: one from industry a to industry b, and another

one from b to a. In the following, we denote IO_1 as the larger value, and IO_2 as the smaller value of the two. Since we have no prior about the functional form that is to be expected, we experimented with different combinations of IO_1 and IO_2. In our regressions below we include the product of these two values, namely $IO_1 \times IO_2$, but we will also control for other combinations.[10] Note that this measure takes complex production chain effects into account (Acemoglu et al., 2012) that are unlikely to be captured by the $NUMNAICS$ control variable.

We also control for other characteristics like $SAMESIZE$ and $SAMEBM$ that are the negative of the absolute difference in percentile ranking for the size and the book-to-market ratio. We also include the absolute difference in financial leverage ratio, noted $DIFFLEV$, and the absolute value of the difference in the two stocks' log share price, $DIFFPRICE$. We also incorporate the past correlation in the two stocks' abnormal trading volume, denoted as $VOLCORR$. Finally, we also include three dummy variables that control for whether the two firms are located in the same states, denoted as $DSTATE$, for stock pairs that are part of the $S\&P500$ index, denoted as $DINDEX$, for stocks that are listed on the same stock exchange, denoted as $DLISTING$.

In the complete regressions, we will also include different functional forms for some of the control variables. Moreover, we will also include the size of each stock, $SIZE_1$ and $SIZE_2$, and the book-to-market ratio of each stock, BM_1 and BM_2. The label 1 indicates the variable associated to the largest of the two stocks. As mentioned above, we also include different functional forms for the input-output variables.

7.3.2 Differences to Antón and Polk

Let us briefly summarize the main features that distinguish our analysis from Antón and Polk (2014): first, our main analysis is not restricted to the subset of mutual

[10] We also included additional functional forms of these measures, namely IO_1, IO_2, $(IO_1)^2 \times IO_2$, $IO_1 \times (IO_2)^2$, and $(IO_1)^2 \times (IO_2)^2$.

funds, but includes the broader set of institutional investors. Second, in line with the results from the previous section, we are particularly interested in the results for relatively small stocks for which institutional ownership has increased dramatically over our sample period. Third, the sample of Antón and Polk ends in 2008, just around the time when the global financial crisis of 2008-09 occurred. Our sample includes data up until 2014, thus covering the most recent crisis period which might have an impact on the relationship. Fourth, we add an extra explanatory variable that is linked to the production chain, and thus should be able to capture network effects between firms from different industries. Finally, while we use the same 4-factor model, we also analysed the results for the 3-factor model (excluding the Momentum factor of Carhart, 1997). The results can be found in the Appendix.[11]

7.4 Empirical Results

7.4.1 Baseline Specification

The main regression results are summarized in Table 7.2.[12] The two panels show the results without and with the input-output variables, respectively. As mentioned above, these regressions include all stocks, not only the relatively large ones. In general, we find that the input-output variables are all significant and positive, even when including various other control variables. Thus, the production chain has a large impact on the stock return correlations. With regards to the main variable of interest, we find that $FCAP$ is always positively significant when including all institutions and all stocks, thus confirming the main results of Antón and Polk.

[11] Somewhat surprisingly, the results do depend on whether we use a 3-factor or 4-factor model. This is largely due to the fact that very small stocks appear to load heavily on the Momentum-factor. Thus, residual correlations can be significantly different with or without this factor. In line with Antón and Polk, we thus stick to the 4-factor model in the main text.

[12] The complete Tables with all explanatory variables can be found in the Appendix.

Panel A - Without input-output control variables

Explanatory variables	(1)	(2)	(3)	(4)
Constant	0.00281	−0.00188	−0.00609	−0.00557
	(32.21)	(−8.55)	(−13.59)	(−12.17)
$FCAP^*$	0.00078	0.00033	0.00080	0.00063
	(7.14)	(2.73)	(6.83)	(5.02)
$SAMESIZE^*$		−0.00071	−0.00110	−0.00187
		(−0.86)	(−1.33)	(−0.67)
$NUMNAICS^*$		0.00253	0.00244	0.00242
		(28.51)	(27.49)	(27.20)
Nonlinear size controls	No	Yes	Yes	Yes
Pair characteristic controls	No	No	Yes	Yes
Nonlinear style controls	No	No	No	Yes

Panel B - With input-output control variables

Explanatory variables	(1)	(2)	(3)	(4)	(5)
Constant	0.00139	0.00123	−0.00300	−0.00699	−0.00654
	(11.33)	(10.01)	(−12.30)	(−15.35)	(−13.88)
$FCAP^*$		0.00088	0.00046	0.00090	0.00075
		(8.16)	(3.94)	(7.75)	(6.06)
$IO_1^* \times IO_2^*$	0.00298	0.00300	0.00226	0.00217	0.00221
	(14.52)	(14.63)	(10.82)	(10.38)	(10.50)
$SAMESIZE^*$			−0.00053	−0.00094	−0.00252
			(−0.61)	(−1.10)	(−0.94)
$NUMNAICS^*$			0.00229	0.00222	0.00219
			(25.21)	(24.38)	(24.00)
Nonlinear IO controls	Yes	Yes	Yes	Yes	Yes
Nonlinear size controls	No	No	Yes	Yes	Yes
Pair characteristic controls	No	No	No	Yes	Yes
Nonlinear style controls	No	No	No	No	Yes

Table 7.2: *Summary of the Fama-MacBeth regressions with and without the input-output control variables.* Panel A shows the regression results for different specifications, excluding the input-output variables. For example, column (1) shows the results when using $FCAP$ as the sole explanatory variable. The other columns include further control variables. Panel B shows similar results when including the input-output controls, with column (5) being the most complete specification. In parenthesis, we indicate the t-statistic associated to each estimate. Table 7.6 in the Appendix shows the complete regression with a complete list of control variables. Note: $*$ indicates rank-transformed variables (zero mean, unit standard deviation).

7.4.2 Subsample Analysis

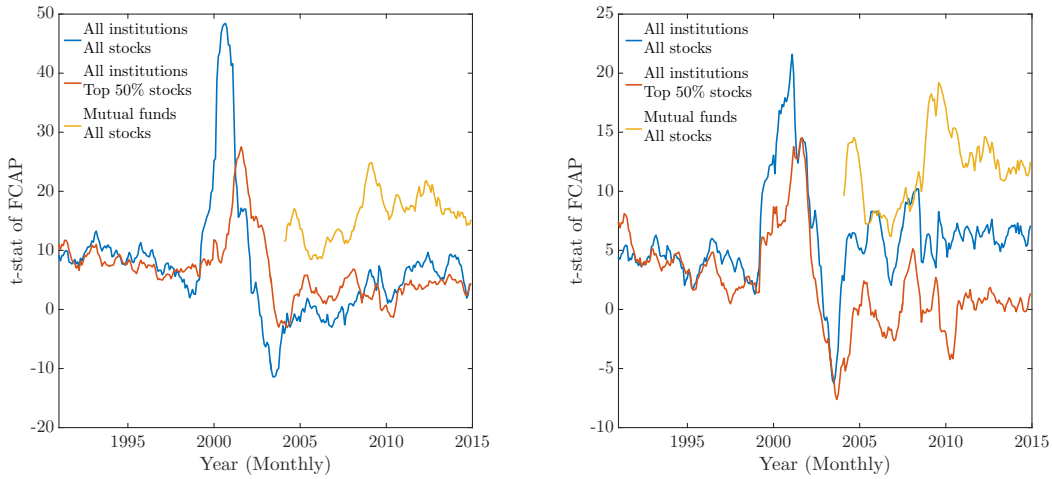
In order to explore to what extent the above results are stable over time, Table 7.3 splits the sample into three subsamples. It turns out that the parameter on FCAP is always positively significant in the three subsamples and appears to be very stable over time (the parameter on FCAP is around 0.00075 in all three subsamples). However, Figure 7.4 illustrates that these averages hide some interesting dynamics. For example, the blue lines show the 12-month moving average of the t-statistic associated with $FCAP$ for our main specification (the left panel corresponds to column (1) in panel A of Table 7.2; the right panel corresponds to column (5) in panel B of Table 7.2). Clearly, the results are not constant over time, since there are very large fluctuations, especially so during and after crisis periods: for example, during the dot-com bubble and the 9/11 attacks in 2001, the t-statistic associated to $FCAP$ becomes very large and positive, and then switches to become strongly negative. On the other hand, it is interesting to notice that the effect of the 2008-09 global financial crisis is minor in comparison. Given these strong cyclicalities is it even more remarkable how stable the parameter estimates in Table 7.3 are.

As a next step, Figure 7.4 also shows the 12-month moving average t-statistics when running the regressions using (a) only the sample of large stocks (above-median market capitalization), and (b) the set of mutual funds. Regarding (a), the red lines show the t-statistics using the relatively large stocks only; in this case, $FCAP$ is positive and statistically significant, confirming that the results of Antón and Polk (2014) also appear to hold for the larger set of institutional investors. Interestingly, however, the t-statistics appear to be smaller relatively to the baseline specification (blue line) at least for the full model. Lastly, we also show the results based on the CRSP Mutual Fund data, including all stocks (yellow lines). The t-statistic is significantly higher compared to the two previous cases, suggesting that mutual funds are comparably active relative to other asset managers and thus have a stronger

impact on the return correlations. This finding is in line with the existing literature, where institutional investors are generally seen as rather inactive long-term investors. However, our results suggest that the common asset holdings of this broader set of professional asset managers also tends to have an impact on future return correlations.

Explanatory variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
	First subsample (1990-99)				
Constant	0.00173	0.00091	−0.00063	−0.00419	−0.00213
	(16.66)	(6.24)	(−2.69)	(−7.54)	(−4.16)
$IO_1^* \times IO_2^*$		0.00175	0.00139	0.00137	0.00140
		(7.17)	(5.56)	(5.45)	(5.60)
$FCAP^*$	0.00121	0.00123	0.00054	0.00074	0.00077
	(9.20)	(9.35)	(3.69)	(4.81)	(4.79)
	Second subsample (2000-09)				
Constant	0.00422	0.00235	−0.00365	−0.00850	−0.00888
	(52.93)	(20.07)	(−16.12)	(−20.52)	(−19.91)
$IO_1^* \times IO_2^*$		0.00349	0.00253	0.00239	0.00235
		(18.77)	(13.47)	(12.74)	(12.46)
$FCAP^*$	0.00043	0.00057	0.00047	0.00101	0.00072
	(5.52)	(7.16)	(4.79)	(10.05)	(7.11)
	Third subsample (2010-14)				
Constant	0.00212	−0.00038	−0.00638	−0.00953	−0.01052
	(21.59)	(−2.64)	(−23.72)	(−20.48)	(−20.96)
$IO_1^* \times IO_2^*$		0.00450	0.00344	0.00332	0.00351
		(21.13)	(15.94)	(15.41)	(16.20)
$FCAP^*$	0.00064	0.00079	0.00029	0.00101	0.00075
	(6.31)	(7.81)	(2.71)	(8.95)	(6.47)
Nonlinear IO controls	No	Yes	Yes	Yes	Yes
Nonlinear size controls	No	No	Yes	Yes	Yes
Pair characteristic controls	No	No	No	Yes	Yes
Nonlinear style controls	No	No	No	No	Yes

Table 7.3: *Summary of the regressions for three subsamples.* The periods covered are 1990-99, 2000-09, and 2010-14, respectively. For each period, we perform the same regressions as in Table 7.6. For the sake of clarity, we only report the estimates of $FCAP$ and $IO_1 \times IO_2$. In parenthesis, we indicate the t-statistic associated to each estimate. Note: $*$ indicates rank-transformed variables (zero mean, unit standard deviation).



(a) Only $FCAP$.

(b) Full Model.

Figure 7.4: *12-months moving average of the t-statistic associated to $FCAP$ for each cross-section.* Here we show the values for three different cases: first, the baseline analysis which includes all institutions and all stocks (blue lines). Second, including only relatively large stocks with above-median market capitalization (red lines). Third, using the CRSP Mutual Fund data including all stocks (yellow lines). The left panel shows the t-statistics when $FCAP$ is the only explanatory variable (corresponding to column (1) in panel A of Table 7.2). The right panel shows the results for the full model including all control variables (corresponding to column (5) in panel B of Table 7.2).

7.4.3 A Closer Look at Different Size Quartiles

The analysis in the previous subsection suggest that small stocks do not necessarily behave dramatically different from large stocks. The next step is to re-run our regressions for the different size quartiles - as before, we separate stocks based on their market capitalization (Q1 denotes the smallest stocks, and Q4 the largest stocks, respectively) to study the dynamics of each category separately.

Therefore, the next step is to run the same regressions from Section 7.4.1 for stocks of different size quantiles. More precisely, we will run separate regressions as specification (5) panel B of Table 7.2 for each pair of size quartiles. For example, 'Q1-Q1' corresponds to the within-category correlations of the smallest stocks, and 'Q1-Q4' corresponds to the between-category correlations between the smallest and

largest stocks, respectively. This analysis helps us to fully disentangle the impact of small stocks on the results in the previous section.

Average parameter estimates for each pair of size categories

	Q1-Q1	Q1-Q2	Q1-Q3	Q1-Q4	Q2-Q2	Q2-Q3	Q2-Q4	Q3-Q3	Q3-Q4	Q4-Q4
Full sample	0.00075 [0.00381]	0.00062 [0.00259]	−0.00014 [0.00285]	0.00007 [0.00312]	0.00007 [0.00288]	−0.00014 [0.00285]	0.00004 [0.00266]	0.00034 [0.00428]	0.00074 [0.00352]	0.00127 [0.00441]
First subsample (1990-1999)	0.00025 [0.00418]	0.00065 [0.00196]	0.00001 [0.00197]	0.00030 [0.00206]	0.00015 [0.00177]	0.00022 [0.00154]	0.00036 [0.00162]	0.00093 [0.00206]	0.00077 [0.00201]	0.00108 [0.00240]
Second subsample (2000-2009)	0.00030 [0.00385]	0.00028 [0.00303]	0.00299 [0.00481]	0.00080 [0.00660]	0.00029 [0.00380]	0.00205 [0.00290]	0.00056 [0.00455]	0.00513 [0.00359]	0.00418 [0.00373]	0.00703 [0.00761]
Third subsample (2010-2014)	0.00094 [0.00352]	0.00062 [0.00294]	−0.00086 [0.00276]	−0.00024 [0.00309]	−0.00018 [0.00337]	−0.00119 [0.00314]	−0.00067 [0.00287]	−0.00144 [0.00473]	0.00032 [0.00411]	0.00114 [0.00451]
Trend analysis										
Intercept	0.00016 (0.36)	0.00027 (0.91)	0.00066 (2.02)	0.00022 (0.61)	0.00021 (0.62)	0.00086 (2.69)	0.00023 (0.76)	0.00169 (3.51)	0.00086 (2.10)	0.00189 (3.70)
Trend	$3.03 \cdot 10^{-6}$ (1.19)	$1.36 \cdot 10^{-6}$ (0.79)	$-5.32 \cdot 10^{-6}$ (−2.82)	$-1.42 \cdot 10^{-6}$ (−0.68)	$-2.17 \cdot 10^{-6}$ (−1.13)	$-8.56 \cdot 10^{-6}$ (−4.63)	$-2.64 \cdot 10^{-6}$ (−1.49)	$-1.16 \cdot 10^{-5}$ (−4.17)	$1.50 \cdot 10^{-7}$ (−0.06)	$-1.61 \cdot 10^{-6}$ (−0.54)
R-squared	0.0047	0.0021	0.0261	0.0015	0.0043	0.0672	0.0074	0.0553	0.0000	0.0010

Average t-statistics for each pair of size categories

	Q1-Q1	Q1-Q2	Q1-Q3	Q1-Q4	Q2-Q2	Q2-Q3	Q2-Q4	Q3-Q3	Q3-Q4	Q4-Q4
Full sample	0.31 [1.37]	0.42 [1.81]	−0.11 [2.36]	0.05 [2.59]	0.05 [1.90]	−0.17 [3.12]	0.06 [3.02]	0.34 [3.84]	1.27 [4.56]	1.52 [4.69]
First subsample (1990-1999)	0.07 [1.27]	0.36 [1.37]	0.01 [1.77]	0.31 [2.10]	0.14 [1.38]	0.32 [2.19]	0.55 [2.65]	1.37 [2.61]	1.70 [3.92]	1.96 [4.10]
Second subsample (2000-2009)	0.11 [1.51]	0.21 [2.20]	2.64 [4.34]	0.68 [5.86]	0.23 [2.74]	2.48 [3.85]	0.67 [5.31]	5.53 [4.10]	6.51 [5.82]	8.60 [8.83]
Third subsample (2010-2014)	0.40 [1.41]	0.47 [2.05]	−0.59 [2.10]	−0.17 [2.22]	−0.12 [2.10]	−1.16 [3.08]	−0.56 [2.74]	−1.14 [3.58]	0.38 [4.23]	0.84 [3.30]
Trend analysis										
Intercept	0.0071 (0.04)	0.1588 (0.75)	0.5105 (1.88)	0.2632 (0.88)	0.1400 (0.64)	1.1026 (3.15)	0.5942 (1.71)	−0.0121 (4.91)	2.6027 (4.98)	4.0534 (7.66)
Trend	0.0015 (1.63)	0.0010 (0.86)	−0.0038 (−2.43)	−0.0015 (−0.84)	−0.0014 (−1.09)	−0.0093 (−4.58)	−0.0045 (−2.22)	−0.0121 (−4.87)	−0.0083 (−2.75)	−0.0130 (−4.26)
R-squared	0.0088	0.0025	0.0195	0.0024	0.0040	0.0661	0.0163	0.0739	0.0249	0.0577

Table 7.4: *Summary of the relationship between stock return correlation and $FCAP$ for different combinations of size categories.* The regressions performed here include all the variables listed in column (5) of Table 7.6 in the Appendix. The first part shows the parameter estimates associated to $FCAP$ when comparing the stock return correlation of stocks in category i with stocks in category j and the second part shows the t-statistic associated to $FCAP$. We also indicate the average estimates and t-statistic within different sub-periods and we study the evolution of each coefficient over time. In parenthesis, we indicate the t-statistic associated to each estimate and in brackets, the standard deviation.

The main results are summarized in Table 7.4. For the sake of brevity, we only report the average parameter estimate and the average t-statistic associated with $FCAP$ (corresponding to column (5) in Panel B of Table 7.2) over the entire sample and for the three subsamples. Lastly, we also explore whether there are any significant

time-trends in the corresponding estimates and t-statistics ('Trend analysis'). For this purpose we run the following simple regressions

$$\text{parameter}_{c,t}^{FCAP} = a_{1,c} + b_{1,c} \times t + \epsilon_{1,t}, \tag{7.5}$$

$$\text{t-statistic}_{c,t}^{FCAP} = a_{2,c} + b_{2,c} \times t + \epsilon_{2,t}, \tag{7.6}$$

separately for each pair of categories (here simply denoted as 'c').

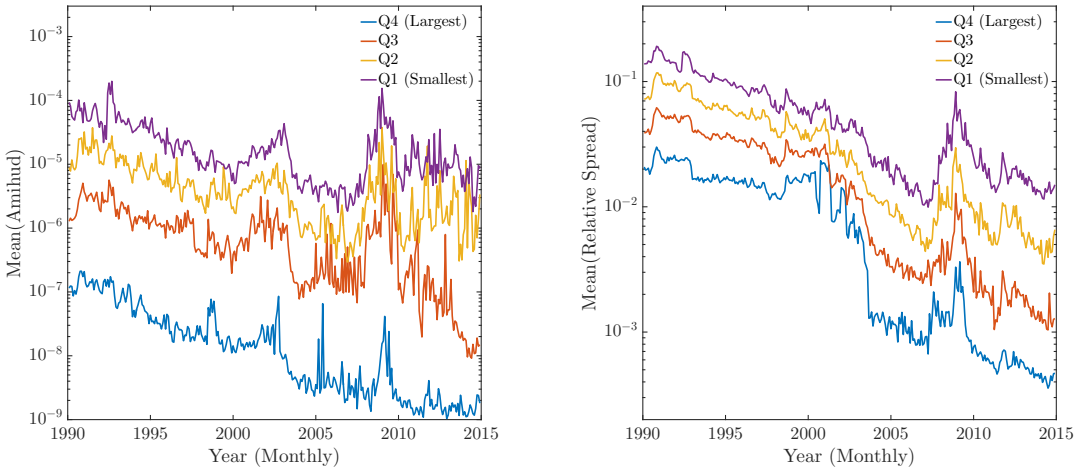
Table 7.4 shows that for the within-category correlations for the largest stocks ('Q4-Q4'), $FCAP$ has a significantly positive impact, however the strength of the relationship is much weaker compared to Antón and Polk (2014). Furthermore, the different subsamples show that the sign of the relationship can change - in this case we see that the typical t-statistics are positive only for the first and second subsample, but negative for the third subsample. Interestingly, we also find a significantly negative trend in the estimates and the t-statistics for the Q4-Q4 case - hence, $FCAP$ appears to become less and less important for explaining future correlations between the set of the largest stocks.

When looking at the results for the relatively small stocks (those involving Q1 and/or Q2), we find positive but insignificant t-statistics in most of the cases. While the values tend to be small, we find a significantly positive trend for the t-statistics involving the smallest stocks (Q1-Q1). Hence, FCAP tends to become more important for explaining the return correlations among the very small stocks over time.

7.4.4 Discussion

In summary, these findings show that despite the fact that institutional ownership (and thus $FCAP$) has been strongly increasing for the small stocks over our sample period, we find no evidence that this coincided with a significant increase in return correlations. However, given the positive sign of the relationship between $FCAP$

and 4-factor residual correlations, it appears that pairs of relatively small stocks tend to comove more strongly with each other when they share many common investors. While this effect is small, it appears to become more important over time. Antón and Polk (2014) explain their results based on correlated trading strategies, so an obvious question is what is the source of the positive but insignificant relationship for the smaller stocks.



(a) Amihud illiquidity measure

(b) Quoted relative spread

Figure 7.5: *Evolution of the liquidity of stocks relative to their size.* We compare two measurements for the liquidity (a) the Amihud illiquidity measure and (b) the mean relative spread.

Clearly, liquidity is a likely explanation for this finding: smaller stocks are much less liquid compared to large stocks. In this regard, Figure 7.5 shows the time dynamics of two illiquidity measures for stocks from different size quartiles: the left panel shows the Amihud-ratio (Amihud, 2002), and the right panel shows the relative quoted bid-ask spread. (Note that the y-axes are shown on logarithmic scale.) First, we see that size is an important factor for liquidity, and larger stocks are generally more liquid. Second, regarding the time dynamics we see that stocks from all size categories have become much more liquid over the last 25 years. However, it turns out that these dynamics are not necessarily uniform across the different categories:

for the largest stocks, we see that the typical Amihud-ratio (relative spread) has decreased by two (one and a half) orders of magnitude, while the values for the smallest stocks have decreased roughly by one order of magnitude in both cases. In other words, nowadays large stocks are even more liquid relative to small stocks than they were 25 years ago.

Coupling these findings with the $FCAP$ dynamics and that institutions hold a larger set of illiquid stocks (see Figure 7.2), it appears reasonable that stressed asset managers are likely to liquidate the most liquid assets first, but refrain from selling small and illiquid stocks. Existing evidence is in line with this interpretation. For example, Nyborg and Östberg (2014) find that, in response to funding shocks, most trading occurs in the set of the most liquid stocks. Relatively illiquid stocks, however, show very little trading activity in these periods. In the next section we provide additional evidence in favor of this explanation.

7.5 Institutional Ownership, Liquidity, and Correlations

The final step is to analyse the interplay between institutional ownership, stock return correlations, and market liquidity in more detail. Earlier we found that institutional ownership of the smallest stocks has increased significantly over our sample period, and its explanatory power for predicting stock return correlations has increased as well. Hence, we might suspect a causal relationship between these two variables in the sense that higher institutional ownership increases return correlations. Similarly, one might expect that higher institutional ownership improves stocks' liquidity in the sense that there are more potential buyers of a given stock. The effect, however, might also work the other way in the sense that higher liquidity makes stocks more attractive to institutional investors.

Ultimately, the above discussion is about causal effects: does higher institutional

ownership lead to higher liquidity? Or does higher liquidity lead to higher institutional ownership? Similarly, does higher institutional ownership increase stocks' return correlations?

In order to answer these questions, ideally we would want using a natural experiment of some kind to properly establish causal effects. However, given that the reported portfolio holdings of institutional investors are at the quarterly level, this seems overly ambitious. Hence, we restrict ourselves to establish Granger-causal relationships in the following (Granger, 1969).

The idea is simple: we wish to estimate the following tri-variate vector autoregressive (VAR) model

$$\begin{pmatrix} InstOwn_{j,t} \\ \text{Amihud}_{j,t} \\ \text{Correlation}_{j,t} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{pmatrix} \times \begin{pmatrix} InstOwn_{j,t-1} \\ \text{Amihud}_{j,t-1} \\ \text{Correlation}_{j,t-1} \end{pmatrix} + \begin{pmatrix} e_{j,t} \\ f_{j,t} \\ g_{j,t} \end{pmatrix}, \quad (7.7)$$

where the as and bs are the parameters to be estimated, and e, f, and g are independent error terms.[13] Granger-causality is ultimately concerned with the significance of the cross-terms. Broadly speaking, three results can occur: (1) no Granger causality; (2) one-directional Granger causality; and (3) bi- or even tri-directional Granger causality. For instance, if $b_{1,2} = 0$ and $b_{2,1} = 0$ then there is no Granger causality between $InstOwn$ and Amihud. Now, if $b_{1,2} \neq 0$ and $b_{2,1} = 0$ or if $b_{1,2} = 0$ and $b_{2,1} \neq 0$ then there is one-directional Granger causality (from Amihud to $InstOwn$ or from $InstOwn$ to Amihud).

In principle, we can perform the above analysis separately for each stock. It is not clear, however, how we should aggregate the information as the number of stocks is generally on the order of a few thousand. Rather, we will categorize stocks into market capitalisation quartiles as before, and take averages of the corresponding

[13] We include additional lags in the regressions, but focus on the case with just one lag, VAR(1), here for illustrative purposes.

Quartiles	Direction	Number of lags																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Q1 (small stocks)	InstOwn→ Correlation	0.386	0.708	0.786	0.518	0.531	0.631	0.665	0.578	0.724	0.770	0.694	0.773	0.582	0.585	0.549	0.634	0.619	0.559
	Amihud → Correlation	0.078	0.251	0.231	0.444	0.571	0.531	0.580	0.666	0.533	0.639	0.685	0.705	0.597	0.631	0.642	0.724	0.659	0.435
	Correlation → InstOwn	0.509	0.676	0.747	0.863	0.689	0.746	0.823	0.865	0.907	0.807	0.783	0.833	0.823	0.658	0.657	0.653	0.673	0.745
	Amihud → InstOwn	0.352	0.382	0.336	0.314	0.196	0.240	0.394	0.382	0.292	0.293	0.359	0.363	0.460	0.657	0.610	0.593	0.654	0.650
	Correlation → Amihud	0.546	0.247	0.268	0.237	0.163	0.317	0.140	0.091	0.123	0.132	0.123	0.212	0.045	0.051	0.085	0.060	0.121	0.195
	InstOwn → Amihud	0.294	0.911	0.980	0.690	0.562	0.544	0.609	0.692	0.568	0.726	0.796	0.819	0.863	0.919	0.952	0.944	0.961	0.968
Q2	InstOwn→ Correlation	0.206	0.284	0.297	0.362	0.314	0.406	0.041	0.036	0.007	0.008	0.012	0.016	0.025	0.019	0.018	0.007	0.007	0.008
	Amihud → Correlation	0.053	0.162	0.083	0.204	0.319	0.304	0.265	0.350	0.279	0.361	0.466	0.444	0.472	0.404	0.419	0.579	0.720	0.762
	Correlation → InstOwn	0.039	0.047	0.237	0.340	0.403	0.215	0.140	0.157	0.195	0.080	0.116	0.140	0.182	0.184	0.193	0.259	0.285	0.384
	Amihud → InstOwn	0.589	0.742	0.913	0.961	0.901	0.922	0.882	0.903	0.553	0.524	0.514	0.637	0.769	0.802	0.732	0.735	0.779	0.816
	Correlation → Amihud	0.267	0.805	0.673	0.596	0.695	0.584	0.249	0.357	0.507	0.613	0.571	0.682	0.797	0.480	0.367	0.446	0.512	0.646
	InstOwn → Amihud	0.302	0.269	0.423	0.577	0.649	0.156	0.166	0.179	0.144	0.208	0.189	0.249	0.326	0.292	0.271	0.134	0.181	0.253
Q3	InstOwn→ Correlation	0.864	0.750	0.547	0.474	0.411	0.531	0.594	0.669	0.507	0.645	0.299	0.302	0.459	0.438	0.470	0.552	0.565	0.482
	Amihud → Correlation	0.009	0.046	0.125	0.189	0.374	0.511	0.632	0.670	0.494	0.630	0.697	0.797	0.847	0.875	0.916	0.943	0.961	0.969
	Correlation → InstOwn	0.058	0.119	0.014	0.035	0.077	0.052	0.082	0.124	0.026	0.031	0.032	0.065	0.096	0.158	0.171	0.192	0.239	0.139
	Amihud → InstOwn	0.137	0.284	0.272	0.397	0.304	0.349	0.352	0.346	0.406	0.347	0.153	0.228	0.264	0.257	0.379	0.445	0.477	0.604
	Correlation → Amihud	0.144	0.454	0.803	0.862	0.684	0.502	0.524	0.545	0.519	0.472	0.506	0.573	0.375	0.476	0.534	0.599	0.644	0.650
	InstOwn → Amihud	0.150	0.128	0.190	0.225	0.171	0.323	0.263	0.292	0.189	0.006	0.007	0.003	0.001	0.001	0.001	0.003	0.005	0.007
Q4 (large stocks)	InstOwn→ Correlation	0.061	0.079	0.031	0.052	0.091	0.151	0.051	0.074	0.090	0.116	0.086	0.111	0.114	0.083	0.108	0.117	0.159	0.200
	Amihud → Correlation	0.013	0.165	0.349	0.498	0.656	0.786	0.885	0.874	0.904	0.934	0.947	0.971	0.982	0.967	0.986	0.909	0.929	0.927
	Correlation → InstOwn	0.737	0.805	0.954	0.637	0.595	0.652	0.767	0.857	0.422	0.466	0.582	0.615	0.651	0.657	0.516	0.569	0.664	0.725
	Amihud → InstOwn	0.167	0.189	0.412	0.601	0.715	0.706	0.802	0.825	0.589	0.655	0.753	0.837	0.888	0.928	0.965	0.960	0.965	0.980
	Correlation → Amihud	0.959	0.990	0.995	0.897	0.951	0.960	0.933	0.933	0.934	0.955	0.954	0.851	0.883	0.928	0.935	0.952	0.951	0.974
	InstOwn → Amihud	0.359	0.466	0.680	0.793	0.725	0.834	0.935	0.955	0.958	0.969	0.978	0.931	0.952	0.947	0.912	0.923	0.963	0.957

Table 7.5: *Granger causality analysis between institutional ownership, correlations of stock return residuals and Amihud.* The arrows in the second column indicate the direction of the Granger causality. The values indicate for each lag and each test the probability to reject the null hypothesis of no causality. For each test, the shade cells represent the best model according to the likelihood-ratio test. We use the first order difference of the institutional ownership because of the non-stationarity. Amihud is the Amihud (2002) illiquidity-measure.

measures ($InstOwn$, Amihud, and Correlations) for the analysis. Hence, the results shown below are for the typical stock in each size category.

In order to account for the potential non-stationarity of $InstOwn$ we take the first-difference in everything that follows. Correlations are, as before, the 4-factor residual correlations and are the average correlations within each category.[14]

We use a symmetric number of lags meaning that the three variables have the same number of lags in the regressions. To find the best number of lags, we perform a statistical test: we first set the maximum number of lags equal to 18 months. The model with 18 lags is viewed as the unrestricted model. On the other hand, a model with fewer lags is considered as a restricted version of the previous model. Then, we use the likelihood ratio test[15] to check whether the restrictions have significantly degraded the fit of the model. The best model is then the one with the minimal number of terms without any significant loss in the explanatory power of the unrestricted model.[16]

Table 7.5 summarizes the main results. For each quartile, the best number of lags is indicated by the cells in grey, which are the specifications of interest. The Table shows the p-values for each of the potential Granger-causal relationships. We first note that for Q1, Q3, and Q4 the optimal number of lags is around 12 months, while for Q2 it is 18 months. We find that $InstOwn$ Granger-causes Correlations only for stocks in Q2. In other words, for the smallest stocks, higher institutional ownership does not increase those stocks' correlations with other stocks. Interestingly, we find that for the smallest stocks (Q1) Correlations Granger-cause Amihud but we do not

[14]Table 7.8 in the Appendix shows the results when using return correlations instead. Because these correlations appear to be non-stationary (see Figure 7.3), we use the first-difference in this case. The qualitative results are very similar to the ones presented here for stocks in Q4. For Q1, Q2, and Q3 there are some differences.

[15]The likelihood ratio statistic is the chi-squared distributed test statistic and is given by $LR = (T - c)(\log|\Sigma_r| - \log|\Sigma_u|)$ with T the number of observations, c is a degrees of freedom correction factor, and $|\Sigma_r|$, $|\Sigma_u|$ denote the determinant of the error covariance matrices from the restricted and unrestricted models (LeSage, 1999).

[16]To accept the restricted model, the threshold for the marginal probability is set to 1%. Increasing the marginal probability to 5% does not change the results.

find any additional significant relationship for these stocks. In particular, we find no evidence that higher $InstOwn$ improves these stocks' liquidity. For relatively large stocks (Q3 and Q4), the results are not uniform either: for example, for Q3 we find that Correlations Granger-cause $InstOwn$, and $InstOwn$ Granger-causes Amihud. For the largest stocks, similar to the stocks in Q2, we find that $InstOwn$ Granger-causes Correlations, but this relationship is not significant at the 10% level.

Overall, these results are in line with the findings in previous sections: the liquidity of small stocks has not increased dramatically despite the rise in institutional ownership in these stocks. On the other hand, we find that institutional ownership Granger-causes correlations only for stocks in Q2 and is borderline significant for the largest stocks.

7.6 Conclusion

In this chapter, we explored to what extent common asset holdings are useful in predicting future return correlations for stocks of different sizes. In particular, we analysed the results for very small stocks for which a shift in institutional preferences has been documented.

Our main results were as follows: first, we document a strong increase in both institutional ownership and common asset holdings over the sample period, particularly so for the smallest stocks. Second, raw stock return correlations have also increased strongly over time, while the increase is much more modest for factor model residual correlations. Third, we confirm that common ownership is significantly related to future return correlations for relatively large stocks, but the relationship appears to become less important over time. For relatively small stocks the relationship is often insignificant, but the effect appears to become more important over time. Lastly, we find no evidence that institutional ownership Granger-causes return correlations for

the subset of small stocks.

Overall, these findings suggest that the relationship between common asset holdings and return correlations is not necessarily uniform for stocks of different size

Appendix

7.A Tables

Explanatory variables	(1)	(2)	(3)	(4)	(5)
Constant	0.00281	0.00123	−0.00300	−0.00699	−0.00654
	(32.21)	(10.01)	(−12.30)	(−15.35)	(−13.88)
$IO_1^* \times IO_2^*$		0.00300	0.00226	0.00217	0.00221
		(18.31)	(13.87)	(13.22)	(12.09)
$FCAP^*$	0.00078	0.00088	0.00046	0.00090	0.00075
	(7.14)	(8.16)	(3.94)	(7.75)	(6.06)
$SAMESIZE^*$			−0.00053	−0.00094	−0.00252
			(−0.61)	(−1.10)	(−0.94)
$SAMEBM^*$			0.00147	0.00153	0.00319
			(1.90)	(1.96)	(1.38)
$NUMNAICS^*$			0.00229	0.00222	0.00219
			(25.21)	(24.38)	(24.00)
$SIZE_1^*$			−0.00452	−0.00290	−0.00283
			(−28.95)	(−16.02)	(−15.36)
$SIZE_2^*$			0.00263	0.00105	0.00107
			(15.87)	(5.92)	(6.05)
$SIZE_1^* \times SIZE_2^*$			−0.00415	−0.00474	−0.00480
			(−14.00)	(−15.58)	(−15.84)
$DIFFLEV^*$				−0.00050	−0.00079
				(−5.14)	(−7.56)
$DIFFPRICE^*$				−0.00188	−0.00185
				(−17.90)	(−17.50)
$DINDEX$				0.00207	0.00198
				(5.53)	(5.59)
$DSTATE$				0.00412	0.00410
				(10.80)	(10.81)
$DLISTING$				0.00230	0.00230
				(12.19)	(12.26)
$VOLCORR^*$				0.00375	0.00379
				(10.46)	(10.59)
IO_1^*		0.00033	0.00085	0.00082	0.00082
		(1.98)	(4.71)	(4.54)	(4.49)
IO_2^*		0.00037	0.00004	0.00005	0.00004

Continued on next page

Table 7.6 – *Continued from previous page*

Explanatory variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
		(2.21)	(0.00)	(0.10)	(0.05)
$IO_1^{*2} \times IO_2^*$		0.00119	0.00091	0.00085	0.00085
		(6.71)	(5.07)	(4.71)	(4.72)
$IO_1^* \times IO_2^{*2}$		0.00118	0.00078	0.00069	0.00070
		(8.34)	(5.45)	(4.79)	(4.85)
$IO_1^{*2} \times IO_2^{*2}$		0.00051	0.00038	0.00033	0.00033
		(6.95)	(5.09)	(4.50)	(4.47)
$SAMESIZE^{*2}$			0.00085	0.00065	0.00203
			(8.27)	(6.70)	(1.06)
$SAMESIZE^{*3}$			−0.00049	−0.00030	0.00118
			(−3.98)	(−2.51)	(1.17)
$SIZE_1^{*2}$			0.00407	0.00514	0.00514
			(20.66)	(22.36)	(22.49)
$SIZE_2^{*2}$			0.00291	0.00340	0.00341
			(14.05)	(16.01)	(16.08)
$SIZE_1^{*2} \times SIZE_2^{*2}$			−0.00071	−0.00081	−0.00080
			(−5.44)	(−6.06)	(−6.03)
$SIZE_1^* \times SIZE_2^{*2}$			−0.00005	−0.00044	−0.00046
			(−0.70)	(2.00)	(−2.12)
$SIZE_1^{*2} \times SIZE_2^*$			0.00112	0.00168	0.00169
			(4.82)	(6.51)	(6.61)
$SAMEBM^{*2}$					−0.00144
					(−0.59)
$SAMEBM^{*3}$					−0.00152
					(−1.31)
BM_1^*					0.02150
					(−5.59)
BM_2^*					−0.00529
					(9.26)
$BM_1^* \times BM_2^*$					−0.08882
					(−2.03)
BM_1^{*2}					0.24179
					(2.52)
BM_2^{*2}					−0.08384
					(−1.12)
$BM_1^* \times BM_2^{*2}$					−1.31108
					(0.51)
$BM_1^{*2} \times BM_2^*$					1.30231
					(−0.71)
$BM_1^{*2} \times BM_2^{*2}$					0.00001
					(2.92)
R^2	0.00004	0.00022	0.00102	0.00126	0.00140

Table 7.6: *Complete regression table with the input-output variables.* In parenthesis, we indicate the t-statistic associated to each estimate. Note: $*$ indicates rank-transformed variables (zero mean, unit standard deviation).

Explanatory variables	(1)	(2)	(3)	(4)
$Constant$	0.00281	−0.00188	−0.00609	−0.00557
	(32.21)	(−8.55)	(−13.59)	(−12.17)
$FCAP^*$	0.00078	0.00033	0.00080	0.00063
	(7.14)	(2.73)	(6.83)	(5.02)
$SAMESIZE^*$		−0.00071	−0.00110	−0.00187
		(−0.86)	(−1.33)	(−0.67)
$SAMEBM^*$		0.00183	0.00184	0.00272
		(2.40)	(2.40)	(1.20)
$NUMNAICS^*$		0.00253	0.00244	0.00242
		(28.51)	(27.49)	(27.20)
$SIZE_1^*$		−0.00445	−0.00275	−0.00262
		(−28.56)	(−15.27)	(−14.32)
$SIZE_2^*$		0.00270	0.00103	0.00113
		(16.36)	(5.83)	(6.37)
$SIZE_1^* x SIZE_2^*$		−0.00413	−0.00476	−0.00483
		(−13.95)	(−15.65)	(−15.93)
$DIFFLEV^*$			−0.00046	−0.00077
			(−4.85)	(−7.40)
$DIFFPRICE^*$			−0.00199	−0.00194
			(−18.94)	(−18.30)
$DINDEX$			0.00217	0.00212
			(6.05)	(6.00)
$DSTATE$			0.00403	0.00405
			(10.55)	(10.65)
$DLISTING$			0.00241	0.00242
			(12.79)	(12.95)
$VOLCORR^*$			0.00378	0.00382
			(10.54)	(10.66)
$SAMESIZE^{*2}$		0.00088	0.00068	0.00172
		(8.55)	(6.90)	(0.98)
$SAMESIZE^{*3}$		−0.00051	−0.00032	0.00065
		(−4.19)	(−2.64)	(0.80)
$SIZE_1^{*2}$		0.00409	0.00520	0.00523
		(20.77)	(22.67)	(22.85)
$SIZE_2^{*2}$		0.00292	0.00344	0.00344
		(14.09)	(16.16)	(16.17)
$SIZE_1^{*2} x SIZE_2^{*2}$		−0.00071	−0.00081	−0.00081
		(−5.47)	(−6.08)	(−6.04)
$SIZE_1^* x SIZE_2^{*2}$		−0.00005	−0.00047	−0.00048
		(−0.68)	(−2.08)	(−2.18)
$SIZE_1^{*2} x SIZE_2^*$		0.00111	0.00171	0.00172
		(4.81)	(6.63)	(6.76)
$SAMEBM^{*2}$				−0.00108
				(−0.46)
$SAMEBM^{*3}$				−0.00100
				(−0.95)
BM_1^*				0.02283
				(−4.76)
BM_2^*				−0.00502
				(10.90)

Continued on next page

Table 7.7 – *Continued from previous page*

Explanatory variables	**(1)**	**(2)**	**(3)**	**(4)**
$BM_1^* x BM_2^*$				-0.09458
				(-2.23)
BM_1^{*2}				0.25105
				(2.79)
BM_2^{*2}				-0.08540
				(-1.27)
$BM_1^* x BM_2^{*2}$				-1.34393
				(0.31)
$BM_1^{*2} x BM_2^*$				1.33493
				(-0.50)
$BM_1^{*2} x BM_2^{*2}$				0.00001
				(2.53)
R^2	0.00004	0.00089	0.00113	0.00128

Table 7.7: *Complete regression table without input-output control variables.* In parenthesis, we indicate the t-statistic associated to each estimate. Note: $*$ indicates rank-transformed variables (zero mean, unit standard deviation).

Quartiles	Direction	Number of lags																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Q1 (small stocks)	InstOwn→ Correlation	0.485	0.897	0.925	0.212	0.194	0.101	0.057	0.086	0.008	0.003	0.004	0.002	0.002	0.003	0.006	0.002	0.005	0.009
	Amihud → Correlation	0.278	0.517	0.641	0.642	0.817	0.846	0.812	0.865	0.782	0.789	0.893	0.914	0.896	0.910	0.840	0.726	0.740	0.763
	Correlation → InstOwn	0.111	0.217	0.439	0.460	0.577	0.530	0.626	0.726	0.744	0.694	0.783	0.127	0.237	0.293	0.330	0.283	0.274	0.370
	Amihud → InstOwn	0.393	0.448	0.406	0.377	0.286	0.310	0.414	0.354	0.280	0.415	0.508	0.410	0.523	0.603	0.676	0.723	0.830	0.840
	Correlation → Amihud	0.166	0.233	0.381	0.190	0.164	0.143	0.242	0.134	0.117	0.173	0.186	0.234	0.254	0.199	0.242	0.324	0.455	0.265
	InstOwn → Amihud	0.367	0.956	0.994	0.760	0.544	0.502	0.644	0.737	0.475	0.628	0.707	0.772	0.746	0.760	0.803	0.810	0.855	0.872
Q2	InstOwn→ Correlation	0.380	0.728	0.854	0.918	0.910	0.966	0.929	0.953	0.901	0.737	0.574	0.609	0.735	0.776	0.856	0.590	0.587	0.460
	Amihud → Correlation	0.127	0.232	0.185	0.242	0.386	0.373	0.472	0.597	0.451	0.444	0.143	0.121	0.056	0.072	0.091	0.087	0.131	0.179
	Correlation → InstOwn	0.320	0.613	0.744	0.662	0.837	0.736	0.302	0.059	0.025	0.047	0.069	0.006	0.010	0.015	0.030	0.048	0.075	0.045
	Amihud → InstOwn	0.892	0.980	0.963	0.980	0.949	0.932	0.919	0.934	0.626	0.697	0.743	0.841	0.898	0.897	0.921	0.911	0.942	0.956
	Correlation → Amihud	0.259	0.003	0.005	0.009	0.012	0.016	0.043	0.067	0.067	0.162	0.081	0.124	0.132	0.159	0.178	0.165	0.076	0.130
	InstOwn → Amihud	0.380	0.244	0.448	0.600	0.700	0.147	0.098	0.070	0.058	0.082	0.076	0.103	0.138	0.205	0.232	0.084	0.107	0.174
Q3	InstOwn→ Correlation	0.594	0.863	0.653	0.586	0.698	0.615	0.011	0.024	0.011	0.007	0.024	0.013	0.021	0.010	0.019	0.009	0.006	0.012
	Amihud → Correlation	0.498	0.359	0.155	0.191	0.109	0.098	0.128	0.122	0.130	0.040	0.072	0.019	0.016	0.022	0.033	0.036	0.038	0.069
	Correlation → InstOwn	0.658	0.928	0.972	0.977	0.975	0.879	0.772	0.195	0.267	0.209	0.200	0.115	0.162	0.116	0.059	0.086	0.072	0.067
	Amihud → InstOwn	0.067	0.170	0.283	0.419	0.275	0.318	0.383	0.243	0.301	0.195	0.106	0.198	0.280	0.158	0.273	0.330	0.368	0.385
	Correlation → Amihud	0.908	0.724	0.979	0.978	0.553	0.182	0.161	0.191	0.229	0.430	0.296	0.231	0.265	0.418	0.430	0.481	0.522	0.620
	InstOwn → Amihud	0.219	0.159	0.217	0.274	0.234	0.314	0.255	0.267	0.213	0.025	0.021	0.016	0.003	0.003	0.008	0.014	0.017	0.026
Q4 (large stocks)	InstOwn→ Correlation	0.168	0.323	0.556	0.280	0.273	0.321	0.108	0.104	0.114	0.112	0.090	0.162	0.155	0.076	0.106	0.136	0.186	0.143
	Amihud → Correlation	0.453	0.653	0.738	0.653	0.874	0.932	0.912	0.921	0.959	0.973	0.867	0.885	0.904	0.905	0.913	0.927	0.935	0.919
	Correlation → InstOwn	0.493	0.539	0.566	0.731	0.723	0.490	0.526	0.217	0.200	0.246	0.318	0.388	0.464	0.530	0.276	0.154	0.134	0.183
	Amihud → InstOwn	0.119	0.191	0.401	0.529	0.682	0.779	0.859	0.826	0.723	0.813	0.845	0.921	0.917	0.965	0.984	0.966	0.964	0.978
	Correlation → Amihud	0.459	0.406	0.249	0.356	0.502	0.695	0.917	0.943	0.980	0.987	0.977	0.988	0.936	0.962	0.941	0.953	0.928	0.960
	InstOwn → Amihud	0.395	0.448	0.667	0.745	0.779	0.874	0.970	0.952	0.966	0.985	0.995	0.979	0.986	0.984	0.938	0.960	0.955	0.957

Table 7.8: *Granger causality analysis between institutional ownership, correlations of the raw stock return and Amihud.* The arrows in the second column indicate the direction of the Granger causality. The values indicate for each lag and each test the probability to reject the null hypothesis of no causality. For each test, the shade cells represent the best model according to the likelihood-ratio test. We use the first order difference of the institutional ownership and of the raw stock return correlations because of the non-stationarity. Amihud is the Amihud (2002) illiquidity-measure.

7.B Results derived using the 3-factor model instead of the 4-factor model

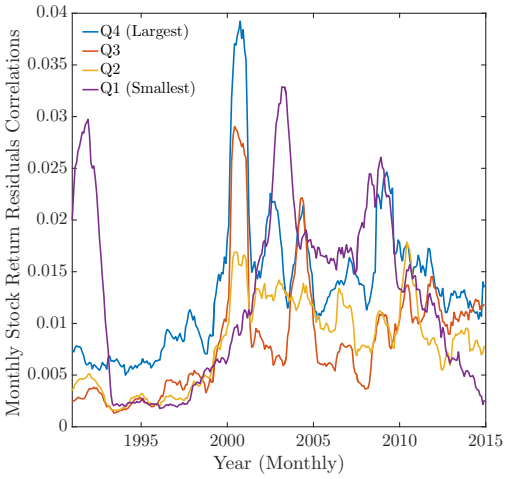
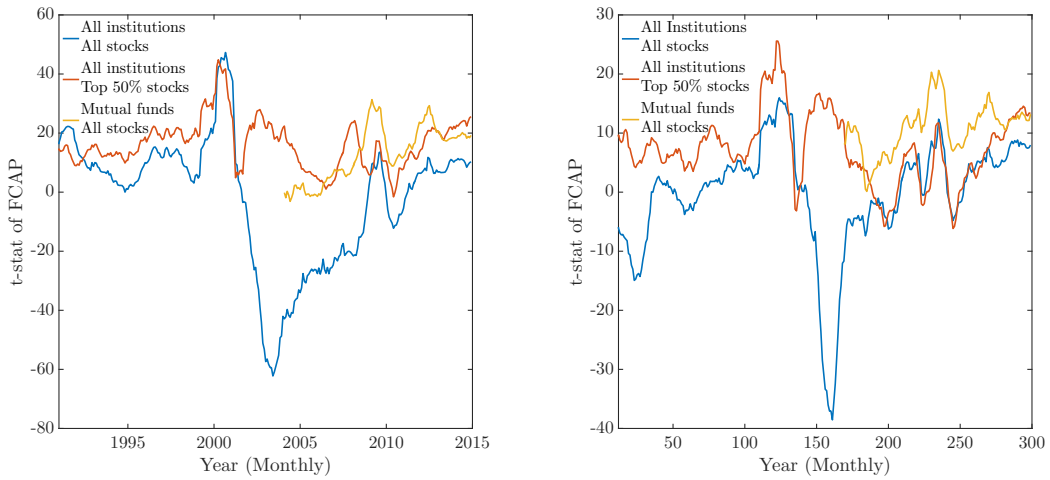


Figure 7.6: *12-months moving average of the monthly stock return residuals correlations within stocks of similar size based on the 3-factor model.*



(a) Only $FCAP$. (b) Full Model.

Figure 7.7: *12-months moving average of the evolution of the t-statistic associated to $FCAP$ for each cross-section based on the 3-factor model.* Here we show the values for three different cases: first, the baseline analysis which includes all institutions and all stocks (blue lines). Second, including only relatively large stocks with above-median market capitalization (red lines). Third, using the CRSP Mutual Fund data including all stocks (yellow lines). The left panel shows the t-statistics when $FCAP$ is the only explanatory variable (corresponding to column (1) in panel A of Table 7.2). The right panel shows the results for the full model including all control variables (corresponding to column (5) in panel B of Table 7.2).

Panel A - Without input-output control variables

Explanatory variables	(1)	(2)	(3)	(4)
Constant	0.00506	−0.00054	−0.00490	−2.26320
	(72.78)	(−0.94)	(−12.26)	(−11.19)
$FCAP^*$	−0.00016	−0.00021	0.00013	−0.00011
	(−1.42)	(−1.18)	(1.43)	(−1.42)
$SAMESIZE^*$		−0.00121	−0.00136	−0.00098
		(−2.32)	(−2.47)	(−1.60)
$NUMNAICS^*$		0.00246	0.00240	0.00236
		(32.55)	(31.72)	(31.11)
Nonlinear size controls	No	Yes	Yes	Yes
Pair characteristic controls	No	No	Yes	Yes
Nonlinear style controls	No	No	No	Yes

Panel B - With input-output control variables

Explanatory variables	(1)	(2)	(3)	(4)	(5)
Constant	0.00336	0.00337	−0.00179	−0.00590	7.11102
	(33.92)	(34.00)	(−6.69)	(−14.62)	(−13.27)
$FCAP^*$		−0.00005	−0.00008	0.00024	−0.00001
		(0.12)	(−1.00)	(2.60)	(−0.26)
$IO_1^* \times IO_2^*$	0.00313	0.00311	0.00237	0.00227	0.00210
	(18.46)	(18.31)	(13.87)	(13.22)	(12.07)
$SAMESIZE^*$			−0.00102	−0.00120	−0.00227
			(−1.90)	(−2.10)	(−1.14)
$NUMNAICS^*$			0.00219	0.00214	0.00210
			(28.32)	(27.66)	(27.15)
Nonlinear IO controls	Yes	Yes	Yes	Yes	Yes
Nonlinear size controls	No	No	Yes	Yes	Yes
Pair characteristic controls	No	No	No	Yes	Yes
Nonlinear style controls	No	No	No	No	Yes

Table 7.9: *Summary of the 3-factor Fama-MacBeth regressions with and without the input-output control variables.* Panel A shows the regression results for different specifications, excluding the input-output variables. For example, column (1) shows the results when using $FCAP$ as the sole explanatory variable. The other columns include further control variables. Panel B shows similar results when including the input-output controls, with column (5) being the most complete specification. In parenthesis, we indicate the t-statistic associated to each estimate. Note: $*$ indicates rank-transformed variables (zero mean, unit standard deviation).

Average estimates for each pair of size categories

	Q1-Q1	Q1-Q2	Q1-Q3	Q1-Q4	Q2-Q2	Q2-Q3	Q2-Q4	Q3-Q3	Q3-Q4	Q4-Q4
Full sample	0.00004 [0.00265]	−0.00008 [0.00246]	−0.00046 [0.00246]	−0.00081 [0.00240]	−0.00023 [0.00308]	−0.00017 [0.00315]	−0.00047 [0.00278]	0.00011 [0.00456]	0.00023 [0.00378]	0.00037 [0.00441]
First subsample (1990-1999)	−0.00026 [0.00161]	−0.00017 [0.00103]	−0.00029 [0.00123]	−0.00035 [0.00139]	−0.00025 [0.00090]	0.00001 [0.00081]	−0.00023 [0.00095]	0.00025 [0.00107]	0.00034 [0.00106]	0.00087 [0.00205]
Second subsample (2000-2009)	0.00029 [0.00320]	−0.00031 [0.00284]	−0.00083 [0.00319]	−0.00102 [0.00271]	−0.00072 [0.00305]	−0.00100 [0.00330]	−0.00146 [0.00333]	−0.00058 [0.00494]	−0.00048 [0.00421]	0.00023 [0.00522]
Third subsample (2010-2014)	0.00041 [0.00281]	0.00049 [0.00289]	−0.00046 [0.00247]	−0.00149 [0.00278]	0.00077 [0.00420]	0.00054 [0.00413]	−0.00060 [0.00305]	0.00019 [0.00599]	0.00081 [0.00470]	0.00005 [0.00520]
Trend analysis										
Intercept	−0.00107 (−3.54)	−0.00093 (−3.27)	−0.00074 (−2.61)	−0.00045 (−1.62)	−0.00059 (−1.64)	−0.00027 (−0.74)	−0.00048 (−1.48)	0.00026 (0.48)	0.00000 (0.00)	0.00081 (1.57)
Trend	$5.93 \cdot 10^{-6}$ (3.39)	$3.95 \cdot 10^{-6}$ (2.42)	$-1.35 \cdot 10^{-6}$ (−0.82)	$-4.34 \cdot 10^{-6}$ (−2.73)	$1.88 \cdot 10^{-6}$ (0.91)	$-2.86 \cdot 10^{-6}$ (−1.36)	$-1.60 \cdot 10^{-6}$ (−0.86)	$-3.37 \cdot 10^{-6}$ (−1.10)	$0.60 \cdot 10^{-6}$ (0.24)	$-2.68 \cdot 10^{-6}$ (−0.91)
R-squared	0.0373	0.0193	0.0023	0.0245	0.0028	0.0062	0.0025	0.0041	0.0002	0.0028

Average t-statistics for each pair of size categories

	Q1-Q1	Q1-Q2	Q1-Q3	Q1-Q4	Q2-Q2	Q2-Q3	Q2-Q4	Q3-Q3	Q3-Q4	Q4-Q4
Full sample	0.05 [2.69]	−0.28 [3.67]	−0.80 [3.80]	−1.42 [3.46]	−0.42 [3.02]	−0.37 [4.19]	−1.00 [3.91]	0.21 [4.38]	0.45 [5.16]	0.49 [4.52]
First subsample (1990-1999)	−0.20 [1.73]	−0.47 [2.78]	−0.67 [3.60]	−0.97 [3.68]	−0.63 [2.64]	0.12 [3.50]	−0.74 [3.85]	0.72 [3.60]	1.54 [4.81]	1.66 [5.06]
Second subsample (2000-2009)	0.20 [3.68]	−0.45 [4.54]	−1.30 [4.20]	−1.97 [3.41]	−0.88 [3.42]	−1.61 [4.42]	−2.14 [3.87]	−0.91 [5.04]	−1.38 [5.89]	−0.40 [4.71]
Third subsample (2010-2014)	0.30 [1.73]	0.53 [2.59]	−0.42 [1.80]	−1.13 [2.25]	0.53 [2.72]	0.71 [3.72]	−0.39 [2.70]	0.27 [3.35]	1.41 [2.98]	0.40 [2.37]
Trend analysis										
Intercept	−2.1203 (−5.73)	−2.4535 (−5.49)	−2.8506 (−5.88)	−2.4470 (−5.23)	−1.0349 (−3.01)	−0.6618 (−1.39)	−0.7827 (−1.72)	0.9421 (1.95)	1.4478 (2.41)	2.3141 (4.20)
Trend	0.0106 (4.97)	0.0103 (4.00)	0.0079 (2.83)	0.0048 (1.77)	0.0038 (1.89)	−0.0016 (−0.60)	−0.0012 (−0.44)	−0.0062 (−2.24)	−0.0073 (−2.11)	−0.0104 (−3.27)
R-squared	0.0768	0.0511	0.0262	0.0105	0.0119	0.0012	0.0007	0.0166	0.0148	0.0348

Table 7.10: *Summary of the relationship between stock return correlation and $FCAP$ for different combinations of size categories based on the 3-factor model.* The regressions performed here include all the variables listed in column 5 in Table 7.6 in the Appendix. The first part shows the estimates associated to $FCAP$ when comparing the stock return correlation of stocks in category i with stocks in category j and the second part shows the t-statistic associated to $FCAP$. We also indicate the average estimates and t-statistic within different sub-periods and we study the evolution of each coefficient over time. In parenthesis, we indicate the t-statistic associated to each estimate and in brackets, the standard deviation.

Quartiles	Direction	Number of lags																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Q1 (small stocks)	InstOwn→ Correlation	0.944	0.115	0.226	0.144	0.199	0.286	0.384	0.374	0.390	0.399	0.281	0.243	0.031	0.026	0.029	0.022	0.045	0.050
	Amihud → Correlation	0.461	0.719	0.579	0.771	0.715	0.784	0.866	0.904	0.942	0.959	0.760	0.690	0.381	0.455	0.637	0.619	0.671	0.643
	Correlation → InstOwn	0.427	0.619	0.844	0.491	0.600	0.772	0.853	0.606	0.651	0.718	0.763	0.543	0.589	0.536	0.628	0.771	0.660	0.686
	Amihud → InstOwn	0.364	0.399	0.396	0.311	0.243	0.344	0.493	0.408	0.231	0.266	0.266	0.196	0.276	0.416	0.437	0.500	0.605	0.609
	Correlation → Amihud	0.184	0.228	0.039	0.082	0.028	0.056	0.062	0.034	0.086	0.127	0.125	0.094	0.124	0.113	0.112	0.103	0.039	0.029
	InstOwn → Amihud	0.331	0.931	0.974	0.666	0.543	0.570	0.679	0.815	0.796	0.894	0.928	0.946	0.948	0.979	0.982	0.977	0.979	0.972
Q2	InstOwn→ Correlation	0.313	0.684	0.857	0.732	0.839	0.418	0.428	0.508	0.198	0.193	0.185	0.276	0.194	0.216	0.194	0.060	0.082	0.048
	Amihud → Correlation	0.008	0.061	0.061	0.154	0.251	0.409	0.460	0.537	0.477	0.456	0.495	0.743	0.513	0.181	0.239	0.289	0.210	0.141
	Correlation → InstOwn	0.766	0.570	0.455	0.458	0.616	0.115	0.115	0.157	0.061	0.067	0.087	0.119	0.085	0.067	0.095	0.122	0.122	0.159
	Amihud → InstOwn	0.993	0.975	0.933	0.970	0.901	0.866	0.786	0.824	0.361	0.378	0.353	0.426	0.500	0.617	0.636	0.651	0.769	0.786
	Correlation → Amihud	0.097	0.727	0.880	0.896	0.928	0.773	0.863	0.939	0.964	0.977	0.980	0.986	0.987	0.964	0.969	0.984	0.974	0.970
	InstOwn → Amihud	0.314	0.282	0.466	0.641	0.753	0.170	0.184	0.142	0.103	0.139	0.114	0.145	0.201	0.196	0.252	0.092	0.159	0.197
Q3	InstOwn→ Correlation	0.882	0.618	0.748	0.673	0.502	0.563	0.667	0.478	0.219	0.272	0.112	0.107	0.209	0.234	0.260	0.289	0.370	0.349
	Amihud → Correlation	0.029	0.060	0.126	0.142	0.272	0.424	0.523	0.532	0.338	0.383	0.514	0.564	0.624	0.672	0.733	0.822	0.873	0.891
	Correlation → InstOwn	0.097	0.218	0.004	0.010	0.025	0.020	0.024	0.044	0.028	0.038	0.066	0.078	0.108	0.159	0.163	0.227	0.282	0.154
	Amihud → InstOwn	0.154	0.299	0.217	0.357	0.310	0.297	0.254	0.274	0.308	0.224	0.128	0.189	0.219	0.197	0.304	0.332	0.381	0.481
	Correlation → Amihud	0.031	0.198	0.491	0.491	0.584	0.549	0.516	0.602	0.668	0.720	0.731	0.758	0.514	0.612	0.658	0.715	0.775	0.731
	InstOwn → Amihud	0.136	0.109	0.153	0.223	0.191	0.325	0.238	0.223	0.177	0.009	0.009	0.005	0.001	0.001	0.003	0.005	0.010	0.014
Q4 (large stocks)	InstOwn→ Correlation	0.294	0.050	0.066	0.117	0.165	0.235	0.138	0.161	0.102	0.112	0.086	0.100	0.128	0.096	0.093	0.061	0.056	0.073
	Amihud → Correlation	0.003	0.065	0.158	0.268	0.406	0.568	0.636	0.448	0.565	0.487	0.576	0.662	0.702	0.720	0.829	0.532	0.484	0.528
	Correlation → InstOwn	0.874	0.984	0.968	0.947	0.584	0.669	0.646	0.750	0.495	0.533	0.656	0.523	0.614	0.618	0.617	0.709	0.774	0.829
	Amihud → InstOwn	0.153	0.197	0.398	0.589	0.653	0.628	0.672	0.741	0.483	0.541	0.658	0.702	0.766	0.814	0.892	0.920	0.934	0.961
	Correlation → Amihud	0.927	0.989	0.994	0.862	0.914	0.909	0.916	0.926	0.877	0.879	0.892	0.906	0.919	0.940	0.958	0.957	0.987	0.989
	InstOwn → Amihud	0.359	0.463	0.678	0.783	0.697	0.792	0.912	0.934	0.921	0.932	0.947	0.906	0.929	0.910	0.863	0.878	0.905	0.885

Table 7.11: *Granger causality analysis between institutional ownership, correlations of stock return residuals based on the 3-factor model and Amihud.* The arrows in the second column indicate the direction of the Granger causality. The values indicate for each lag and each test the probability to reject the null hypothesis of no causality. For each test, the shade cells represent the best model according to the likelihood-ratio test. We use the first order difference of the institutional ownership because of the non-stationarity. Amihud is the Amihud (2002) illiquidity-measure.

Chapter 8

Conclusion

In this thesis, I have focused my research on input-output networks and its application to understand economic growth. For this, I have developed a theoretical model that makes the connection between the two and then I have performed several empirical analysis to validate the pertinence of the model.

The model developed is about technological progress. We assume that to produce the same quantity of output an industry will need less inputs. This simple approach allows us to derive some important analytical results. We use the input-output network to study the propagation of the technical change and find that, according to the position of industries in the production chain, they are not improving at the same rate. Instead, we find that overall economic growth is directly related to a quantity that we called the trophic depth, by analogy with ecology, that captures the depth of the economic network and appears to be a multiplier of technical progress. We have proved the important property that the trophic depth is independent of the level of aggregation. Therefore, we can test our theory even with highly aggregated networks. We also make additional predictions about the evolution of prices that is related to the position of industries in the production chain network.

Then, I test the model using the input-output tables of 40 countries at the level

of 35 industries. We validate the four predictions made in the model despite the need of some interpretations to match the data with the model. We also compare in more details the production chain structures of China and the USA and find that while they are the two biggest economies in the world, they are radically different when we study their trophic structure. The main finding is that the trophic depth is a strong determinant of economic growth. Moreover, we find that its value has some predictive power for future growth. Our observation remains valid when controlling for other variables used in neoclassical growth theory. Finally, we compare our finding with other variables developed in the complexity economics literature and we show that the trophic depth performs better for the subset of countries in the dataset.

In another part, we find that along the development path the trophic depth changes. We have been able to estimate the trophic depth for the USA over more than two centuries and find that it increases during the first stage of development before it decreases. Relating this observation with the finding of our model, we provide a potential explanation for the acceleration and deceleration of growth along the economic development. We use a model selection approach to better quantify the non-linear dynamic relation between trophic level and level of GDP per capita. It confirms our initial statement for the USA. Using this new framework, we perform some growth forecasting analysis and we find that despite the simplicity of our approach, the forecasting errors we make are aligned with those made by the IMF. We use this results and the fact that it is a strong determinant of economic growth to claim that the trophic depth should be used when assessing the development of an economy aside of other indicators.

Now that we have identified the role of the input-output networks in economic growth theory, we focus on the driver of its dynamics. For this purpose, we used a more disaggregated input-output network for the USA and studied its evolution over 25 years. We build on our expertise to add some additional variables that have not

been identified originally in the literature as key variables. We find that they are useful to explain the linkage dynamic of the input-output network and we believe they are less sensitive to network aggregation compared to other variables.

Finally, we show that input-output network can explain some important features in fields outside macroeconomics. We study the role of investors' common asset holdings on market dynamics. We extend previous studies to a larger set of stocks and investors to acknowledge for recent shifts in institutional preferences. We find that the effect of institutional preferences on stock return comovement for the largest stocks is decreasing while it is increasing for the smallest stocks. However, we do not find that any causal relation between institutional ownership and stock return comovement for the largest and the smallest stocks. We use the input-output network to take into account the complex production chain effects.

One of the main limitations of the work presented here is the data available. The data used here on input-output networks is the best that we could find that has enough countries. Nonetheless, the dataset only contains a small subset of countries. While in terms of GDP shares they capture most of the world GDP, they are not well representative of the spatial or wealth repartition of countries in the world. For instance, we do not have any African countries in our dataset. Moreover, the time scale of the dataset is relatively small compared to the overall time scale of economic development. This work would largely benefit from the use of a larger dataset both in term of number of countries but also the time covered.

One potential extension of this work would be to integrate investment in the model. Many studies have identified that saving is a key factor for economic growth. When we looked at the determinants of economic growth we have noticed that it has a strong explanatory power. For this purpose, we would need to rewrite the first equations of the model to take into account that the money flowing towards one node is larger than the money flowing out. The difference will then be the investment.

Within this framework, the balance sheets of sectors are becoming more complex and cannot be represented into a simple network as the money flows will not be preserved in the network but can accumulate. Instead, we could imagine that the economy is now a multiplex network: a series of networks that are coupled. Entities exchanged in these networks do not have to be the same. I believe this is the future as it can capture the entire specificity of the economic ecosystem into a simple framework where networks can be treated individually or in a whole with coupling relations.

Bibliography

Acemoglu, D., D. Autor, D. Dorn, G. H. Hanson, and B. Price (2016). Import Competition and the Great U.S. Employment Sag of the 2000s. *Journal of Labor Economics 34*(S1), 141–198.

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

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Microeconomic origins of macroeconomic tail risks. *National Bureau of Economic Research* (20865), 29.

Adams, S. M., B. L. Kimmel, and G. R. Ploskey (1983). Sources of Organic Matter for Reservoir Fish Production: a Trophic-Dynamics Analysis. *Canadian Journal of Fisheries and Aquatic Science 40*, 1480–1495.

Aghion, P. and P. Howitt (1992). A Model of Growth Through Creative Destruction. *Econometrica 60*(2), 323–351.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Antón, M. and C. Polk (2014). Connected Stocks. *The Journal of Finance 69*(3), 1099–1127.

Aroche-Reyes, F. (2003). A qualitative inpur-output method to find basic economic structures. *Regional Science 82*(4), 581–590.

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies 29*(3), 155.

Arthur, W. B. (2009). *The nature of technology : what it is and how it evolves*. Free Press.

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

Baqaee, D. R. (2015). Labor Intensity in an Interconnected Economy.

Barrot, J.-N. and J. Sauvagnat (2014). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. pp. 66.

Bartelme, D. and Y. Gorodnichenko (2015). Linkages and Economic Development. *National Bureau of Economic Research* (21251), 71.

Bennett, J. A., R. W. Sias, and L. T. Starks (2003). Greener Pastures and the Impact of Dynamic Institutional Preferences. *Review of Financial Studies 16*(4), 1203–1238.

Bliss, C. I. (1934). The Method of Probits. *Science 79*(2037), 38–39.

Blöchl, F., F. J. Theis, and F. Vega-Redondo (2011). Vertex centralities in input-output networks reveal the structure of mordern economies. *Physical Review E 83*(4), 046127.

Boehm, C., A. Flaaen, and N. Pandalai (2014). Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tohoku Earthquake. pp. 69.

Bolt, J. and J. L. Van Zanden (2014). The Maddison Project: collaborative research on historical national accounts. *The Economic History Review 67*(3), 627–651.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance 52*(1), 57–82.

Carvalho, V. M. (2008). *Aggregate Fluctuations and the Network Structure of Intersectoral Trade*. Ph. D. thesis, University of Chicago.

Carvalho, V. M. (2014). From Micro to Macro via Production Networks. *Journal of Economic Perspectives 28*(4), 23–48.

Carvalho, V. M., M. Nirei, and Y. U. Saito (2014). Supply Chain Disruptions: Evidence from the Great East Japan Earthquake.

Carvalho, V. M. and N. Voigtländer (2014). Input Diffusion and the Evolution of Production Networks. *National Bureau of Economic Research* (20025), 53.

Chenery, H. B. and T. Watanabe (1958). International Comparisons of the Structure of Production. *Econometrica 26*(4), 487–521.

Cobb, C. W. and P. H. Douglas (1928). A theory of Production. *The American Economic Review 18*, 139–165.

Coelli, T. J., D. S. P. Rao, C. J. O'Donnell, and G. E. Battese (2005). *An Introduction to Efficiency and Productivity Analysis* (2nd ed.). Springer.

Cristelli, M., A. Tacchella, and L. Pietronero (2015). The Heterogeneous Dynamics of Economic Complexity. *PLOS ONE 10*(2), 15.

Domar, E. (1946). Capital Expansion, Rate of Growth, and Employment. *Econometrica 14*(2), 137–147.

Fadinger, H., C. Ghiglino, and M. Teteryatnikova (2015). Income Differences and Input-Output Structure. pp. 61.

Falkenstein, E. G. (1996). Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings. *The Journal of Finance 51*(1), 111–135.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *The American Economic Review 105*(10), 3150–3182.

Foerster, A. T., P.-D. G. Sarte, and M. W. Watson (2011). Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. *Journal of Political Economy 119*(1), 1–38.

Frambach, R. T. and N. Schillewaert (2002). Organizational Innovation Adoption: A Multi-Level Framework of Determinants and Opportunities for Future Research. *Journal of Business Research 55*(2), 163–176.

Frankel, M. (1962). The Production Function in Allocation and Growth: A Synthesis. *The American Economic Review 52*(5), 996–1022.

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.

Fricke, D. (2016). Are Specialists 'Special'? *SSRN* (2785537), 25.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica 79*(3), 733–772.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis* (2nd ed.).

Gompers, P. A. and A. Metrick (2001). Institutional Investors and Equity Prices. *The Quarterly Journal of Economics 116*(1), 229–259.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica 37*(3), 424.

Greenwood, R. and D. Thesmar (2011). Stock price fragility. *Journal of Financial Economics 102*(3), 471–490.

Hanley, J. A. and B. J. McNeil (1983). A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. *Radiology 148*(3), 839–843.

Harrod, R. F. (1939). An essay in Dynamic Theory. *The Economic Journal 49*(193), 14–33.

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, A. Simoes, and M. Yildirim (2011). *The Atlas of Economic Complexity*. Puritan Press.

Hausmann, R., L. Pritchett, and D. Rodrik (2004). Growth Accelerations. *National Bureau of Economic Research* (10566), 24.

Hawkins, D. and H. A. Simon (1949). Note: Some Conditions of Macroeconomic Stability. *Econometrica 17*(3), 245–248.

Horvath, M. (2000). Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics 15*, 69–106.

Hulten, C. R. (1978). Growth Accounting with Intermediate Inputs. *The Review of Economic Studies 45*(3), 511–518.

Inada, K.-i. (1963). On a Two-Sector Model of Economic Growth: Comments and a Generalization. *Review of Economic Studies 30*(2), 119–127.

International Monetary Fund (2009). World Economic Outlook: Sustaining the Recovery. pp. 226.

International Monetary Fund (2016). World Economic Outlook: Too Slow for Too Long. pp. 230.

Jackson, M. O. and B. W. Rogers (2007). Meeting Strangers and Friends of Friends: How Random Are Social Networks? *The American Economic Review 97*(3), 890–915.

Johnson, S., V. Domínguez-García, L. Donetti, and M. A. Muñoz (2014). Trophic coherence determines food-web stability. *Proceedings of the National Academy of Sciences of the United States of America 111*(50), 17923–8.

Jones, C. I. (1995). R&D - Based Models of Economic Growth. *Journal of Political Economy 103*(4), 759–784.

Jones, C. I. (2013). Misallocation, Economic Growth, and Input-Output Economics. In *Advances in Economics and Econometrics*. Westview Press.

Katz, L. (1953). A New Status Index Derived from Sociometric Analysis. *Psychometrika 18*(1), 39–43.

Koch, A., S. Ruenzi, and L. Starks (2016). Commonality in Liquidity: A Demand-Side Explanation. *Review of Financial Studies 29*(8), 1943–1974.

Krüger, J. J. (2008). Productivity and Structural Change: A Review of the Literature. *Journal of Economic Surveys 22*(2), 330–363.

Kurtz, H. D. and N. Salvadori (1995). *Theory of Production*. Cambridge University Press.

Kuznets, S. (1973). Modern Economic Growth: Findings and Reflections. *The American Economic Review 63*(3), 247–258.

Lenzen, M., K. Kanemoto, D. Moran, and A. Geschke (2012). Mapping the Structure of the World Economy. *Environmental Science and Technology 46*(15), 8374–8381.

Leontief, W. (1936). Quantitative Input-Output Relations in the Economic System of the United States. *Review of Economics and Statistics 18*, 105–125.

Leontief, W. (1941). *The Structure of American Economy 1919-1939: An Empirical Application of Equilibrium Analysis*. New York: Oxford University Press.

LeSage, J. P. (1999). Applied Econometrics using MATLAB.

Levine, S. (1980). Several Measures of Trophic Structure Applicable to Complex Food Webs. *Journal of Theoretical Biology 83*, 195–207.

Long, J. B. and C. I. Plosser (1983). Real Business Cycles. *Journal of Political Economy 91*(1), 39–69.

Lü, L., L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley (2015). Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences of the United States of America 112*(8), 2325–2330.

Lucas, R. E. (1988). On the Mechanism of Economic Development. *Journal of Monetary Economics 22*, 3–42.

Malthus, T. R. (1798). *An Essay on the Principle of Population*. J. Johnson, London.

Marshall, A. (1890). *Principles of Economics*.

McMillan, M. S. and D. Rodrik (2011). Globalization, Structural Change and Productivity Growth. *National Bureau of Economic Research* (17143), 54.

McNerney, J. (2012). *Applications of Statistical Physics to Technology Price Evolution*. Ph. D. thesis, Boston University.

McNerney, J., B. D. Fath, and G. Silverberg (2013). Network structure of inter-industry flows. *Physica A 392*(24), 6427–6441.

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

Neumann, T. V. (1945). A Model of General Economic Equilibrium. *The Review of Economic Studies 13*(1), 1–9.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

Nyborg, K. G. and P. Östberg (2014). Money and liquidity in financial markets. *Journal of Financial Economics 112*(1), 30–52.

Officer, L. H. and S. H. Williamson (2013). Annual Inflation Rates in the United States, 1775-2014, and United Kingdom, 1265-2014. Technical report, Measuring-Worth.

Officer, L. H. and S. H. Williamson (2015). Annual Wages in the United States, 1774-Present. Technical report, MeasuringWorth.

Page, L., S. Brin, R. Motwani, and T. Winograd (1999). The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab.

Palgrave Macmillan (2013). *International Historical Statistics*. Edited by Palgrave Macmillan Ltd.

Pasinetti, L. l. (1981). *Structural Change and Economic Growth: a Theoretical Essay on the Dynamics of the Wealth of Nations*. Cambridge University Press.

Pasinetti, L. l. (1993). *Structural Economic Dynamics - A theory of the economic consequences of human learning*. Cambridge University Press.

Pollet, J. M. and M. Wilson (2010). Average correlation and stock market returns. *Journal of Financial Economics 96*(3), 364–380.

Potron, M. (1912). Possibilité et détermination du juste prix et du juste salaire. *Le mouvement social 73*(4), 289–316.

Pritchett, L. (2000). Understanding patterns of economic growth : searching for hills among plateaus, mountains, and plains. *The World Bank Economic Review 14*(2), 221–250.

Quesnay, F. (1758). *Tableau Economique*.

Ranganathan, S., V. Spaiser, R. P. Mann, and D. J. T. Sumpter (2014). Bayesian Dynamical Systems Modelling in the Social Sciences. *PLOS ONE 9*(1).

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). The Free Press.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy 98*(5).

Sala-i Martin, X. (1997). I Just Ran Two Million Regressions. *The American Economic Review 87*(2), 178–83.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. New York: Harper and Brothers.

Schwab, K. and X. Sala-i Martin (2014). The Global Competitiveness Report. Technical report, World Economic Forum.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics 70*(1), 65–94.

Sraffa, P. (1960). *Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory*. Cambridge University Press.

Stein, A. A. (1982). Coordination and collaboration: regimes in an anarchic world. *International Organization 36*(2), 299–324.

Swan, T. W. (1956). Economic Growth and Capital Accumulation. *Economic Record 32*(2), 334–361.

ten Raa, T. (2010). *Input-Output Economics: Theory and Applications: Featuring Asian Economies*. World Scientific Publishing.

Theil, H. (1961). *Economic forecasts and policy*. Amsterdam, North-Holland Pub. Co.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. de Vries (2015). An Illustrated User Guide to the World Input-Output Database: the Case of Global Automotive Production. *Review of International Economics*.

US Bureau of Economic Analysis (2015). Benchmark Input-Output Accounts.

US Bureau of Labor Statistics (2015). Nonfarm Business Sector: Labor Share. Technical report, retrieved from Federal Reserve Bank of St. Louis.

Walras, L. (1874). *Elements of Pure Economics*.

Williams, R. J. and N. D. Martinez (2004). Limits to Trophic Levels and Omnivore in Complex Food Webs: Theory and Data. *The American Naturalist 163*(3), 458–468.

World Bank (2015). World Bank Open Data.