Spatial Interaction and Economic Growth

Zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von
M.Sc. in Economics Tim Deeken

Institut für Volkswirtschaftslehre (ECON)
Lehrstuhl für Wirtschaftspolitik

Referentin: Prof. Dr. Ingrid Ott
Korreferent: Dr. Marten Hillebrand

Mündliche Prüfung: 6. November 2015 in Karlsruhe

Abstract

Spatial interaction is a central characteristic of all economic activity at different levels of aggregation. This thesis investigates its role in a variety of contexts and pursues several goals. First, it aims at providing new insights into the role of integration in the agglomeration-growth nexus. Embedded in a discussion on size and scale, the concept of integration as a spatial institution is developed. It is argued that integration is a multidimensional concept that pins down the impact of institutions to a spatial dimension.

One dimension of integration is the strength of knowledge spillovers. A deeper understanding of these constitutes the second goal. In a recent influential contribution to the cross-country growth literature, the importance of technological interdependence working via spatial externalities between countries has been demonstrated within an integrated theoretic and empirical framework. This thesis supports the hypothesis of interdependence, but challenges a series of empiric results by demonstrating that they hinge crucially on the particular version of the Penn World Table that is used to build the data set.

In addition, spatial econometric methods are applied to provide new estimates on the strength of knowledge spillovers between countries. Furthermore, this thesis contributes to the literature by modeling interdependence between countries via their bilateral genetic distance instead of the default measure in the literature, geographic distance. Also in this case, it is shown that the estimation results are highly sensitive to data revisions. This issue is of particular importance in providing policy advice.

The final goal is to investigate knowledge spillovers within the setting of a single country. By taking a further recent extension of the model from the previous analysis, a gap in the literature is filled by applying this model to the US states. It is shown that the hypothesis of technological interdependence between US states receives support. However, the estimation results from the Spatial Durbin Model fail to indicate either a direct or indirect impact of R&D investments on per worker income in the US states.

Table of Contents

List of Figures

List of Tables

Introduction

The topic of this thesis is “Spatial Interaction and Economic Growth”. It thus addresses two key facets in human society. Starting with economic growth, it denotes the increase in the market value of the goods and services that are produced in e.g. a country over a period of time. Differences in the rates of economic growth across countries affect their respective levels of per capita income over time. For the largest part of human history, differences in per capita income levels have been comparatively small, and only during the last two centuries have income differences widened (see, for example, Galor, 2005). The determination of this evolution merits detailed study. Not simply from a purely materialistic perspective of “the more goods, the better”, but also from a broader point of view. Per capita income levels correlate with characteristics that fundamentally affect people’s lives, like health (Weil, 2014).[1] Economic growth is thus clearly a topic of central importance.

Before elaborating on interaction, a second key aspect in human society, and how it is related to economic growth, space will be introduced into the discussion. Figure 1 serves as a starting point for this. It depicts Earth at night. As can be seen, some areas like Europe, Japan or the east of the United States are brightly illuminated. In contrast, large areas of Russia, China or on the South American and African continents exhibit noticeably less light at night. Illumination at night correlates with economic activity.[2] Hence, the figure clearly demonstrates that economic activity has a spatial dimension. In fact, the World Bank notes in its 2009 World Development Report that location, understood as the place of work, is the “best predictor of income in the world today” (World Bank, 2009, 1).

An important factor in this result are a location’s specific environmental characteristics (or first-nature geography), i.e. navigable rivers, access to the sea, the presence of natural resources, or the disease burden (see, for example, Gallup et al., 1999). These are, however, not the only determinants of the spatial distribution of economic activity. Second-nature geography also plays an important role. At this point, interaction enters the picture. Interaction, as related to economics, comprises several categories, among these are trade

[1] Higher incomes per capita are, for instance, negatively associated with infant mortality and positively with higher life expectancy (Pritchett and Summers, 1996). In the respective country with the highest life expectancy (for females), this variable has grown by the remarkable rate of 3 months per year since 1840 (Oeppen and Vaupel, 2002).

[2] See e.g. Henderson et al. (2012) or Michalopoulos and Papaioannou (2013).

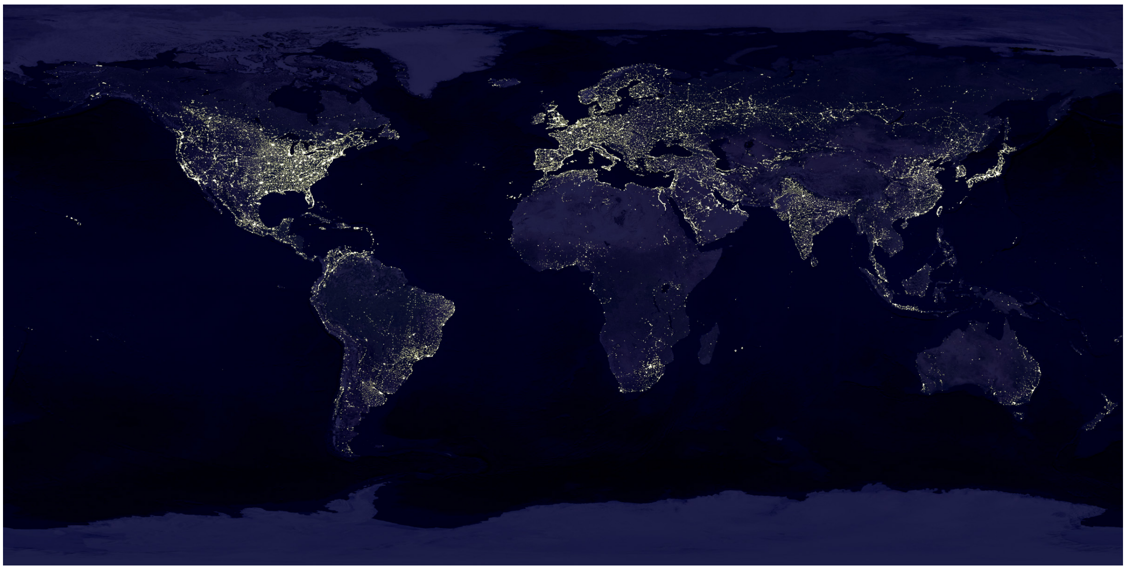


Figure 1: Earth at Night (Source: NASA (2012). Data courtesy Marc Imhoff of NASA GSFC and Christopher Elvidge of NOAA NGDC. Image by Craig Mayhew and Robert Simmon, NASA GSFC. `http://visibleearth.nasa.gov/view.php?id=55167` (accessed: 11 August, 2015)).

in goods and services, but also migration and, for example, foreign direct investment. Every single one of these interactions is possibly associated with an exchange in ideas, knowledge, or technology. Moreover, these dimensions, by their very definition, all include a role for space: They have an impact on the spatial distribution of economic activity and vice versa. For example, migrants crossing a border and working in the new country have an impact on its economic activity as some of them might set up a new firm which then possibly attracts new workers. More generally, agglomeration and dispersion forces shape the economic landscape. People and firms concentrate in certain locations in order to take advantage of large markets (*sharing*), thick labor markets (*matching*) or to benefit from knowledge spillovers (*learning*).[3] Opposing these agglomeration forces, dispersion forces like pollution, crime or high land rents also need to be taken into account. The distribution of economic activity across space can then be understood as a result of balancing these forces in a spatial equilibrium. Necessarily, a certain degree of integration like relatively free movement of labor, capital, goods or services between the locations needs to exist in order for interaction to take place between them. These considerations describe the background for Chapter 1 in this thesis.[4]

Chapter 1 successively provides key empirical facts on the evolution of agglomeration, economic growth and integration over the course of recent economic history and provides a theoretical economic underpinning for each of these. It then presents a model by Baldwin

[3]Compare, for example, Marshall (1890) and Krugman (1998) for these agglomeration economies. The words in parentheses denote the labels by which the concepts are often summarized in the literature. See, e.g. Puga (2010, 210).

[4]Chapter 1 has been published in similar form as Deeken and Ott (2014a) and as Working Paper No. 59 in the Working Paper Series in Economics at KIT (Deeken and Ott, 2014b).

and Forslid (2000), which connects all three characteristics. The advantage of this model is that it combines two models from different strands of literature. This is made feasible as they share the same foundation in the Dixit-Stiglitz approach to modeling monopolistic competition (Dixit and Stiglitz, 1977). In the combined model, the first one, from the field of new economic geography, is the two-region core-periphery model by Krugman (1991), in which transport costs for goods traded between regions play a role so that the location of economic activity matters.[5] In addition, it is based on the assumption of increasing returns to scale on a firm level. Therefore, firms strive to concentrate production in a single location.[6] Agglomeration and dispersion forces then endogenously determine the spatial distribution of economic activity for an exogenously given value of transport costs. The endogenous growth model of increasing variety by Romer (1990) constitutes the second part of the combined model. A crucial assumption in this model is that investment leads to knowledge spillovers, which in turn determine growth. Importantly, when combining these two models, Baldwin and Forslid (2000) assume that only a fraction of the knowledge generated in one region spills over into the other region. Later chapters will introduce the concept of an interaction matrix which allows, although in a different modeling framework, to capture interaction between more than two regions. Nonetheless, the general setup described here allows for a more nuanced consideration of integration compared to traditional new economic geography models as it includes a second exogenous policy parameter besides transport costs and that is the one governing the strength of knowledge spillovers between regions, i.e. the cost of sharing information.

Scale is another important component in Chapter 1. The global structure of areas in which economic activity is spatially concentrated across continents, is repeated at lower levels of aggregation as well, i.e. when zooming in on a section of the map. Examples include the differences between the northwest and southeast of Europe on the continental scale or between Madrid and the rest of Spain on a national scale. Agglomeration and dispersion forces determine the location of economic activity at all scales. The precise design of integration, understood as an institution with a spatial dimension, matters for taking advantage of the benefits of agglomeration. Hence, policy implications concerning integration which foster interaction are presented at the end of Chapter 1.[7]

[5] As Gallup et al. (1999, 185) point out, the importance of transport was well understood by Adam Smith, as can be seen from the following passage from his work they provide: "There are in Africa none of those great inlets, such as the Baltic and Adriatic seas in Europe, the Mediterranean and Euxine seas in both Europe and Asia, and the gulphs of Arabia, Persia, India, Bengal, and Siam, in Asia, to carry maritime commerce into the interior parts of that great continent" (Smith, 1904, I.3.8.).

[6] The Spatial Impossibility Theorem by Starrett (1978) matters in this regard. It states that if space is homogeneous, increasing returns and indivisibilities are absent, then in a competitive equilibrium with positive transport costs, the characteristic feature of the economy will be backyard capitalism (see also Ottaviano and Thisse (2004, 2571-2573)).

[7] The benefits of integration were also clearly understood by Smith. Gallup et al. (1999) were not explicitly concerned with integration when excerpting Smith's work. However, reading on in the original

Chapter 2[8] addresses both theoretically and empirically the issue of how interaction between countries via knowledge spillovers influences per capita income levels. The chosen context is the growth model developed by Ertur and Koch (2007), which explicitly incorporates technological interdependence between countries in an integrated theoretical and empirical framework. This interdependence is assumed to operate through spatial externalities, i.e. knowledge spillovers that cross borders.

The question of how to capture the strength of interaction between countries is a main part of this chapter. It should be kept in mind that whereas it is relatively straightforward to model the interaction between two countries (as in Chapter 1 for the transport costs or the strength of knowledge spillovers), this problem becomes more challenging when the number of countries is increased. As Ertur and Koch (2011) point out, this is the case, since with just two countries only direct effects between these two exist. Adding an additional country, the possible interaction effects become more complex. In this case, countries can not only interact directly with each other, but also indirectly so that feedback effects are now present. This increase in complexity that results in the transition from modeling interdependence between two countries to interdependence between three (or more) countries is the so-called "three-ness effect" (compare Ertur and Koch (2011, 218) with reference to Behrens and Thisse (2007, 461)).

Introducing an interaction matrix which specifies the strength of the connections between countries into a model and thereby creating a system that includes feedback effects between observations is a way of addressing this effect. Such a matrix includes an entry (or weight) for every country pair under consideration reflecting how close the countries are to each other and thus serves as a proxy for the strength of knowledge spillovers between these two countries. Ertur and Koch (2007) base the interaction terms between countries on their respective geographical distance to each other. This is standard practice in the literature, due to empirical evidence that an increase in the geographical distance between originating and receiving country has a negative effect on the amount of knowledge spillovers in a wide variety of industry sectors (Keller, 2002).[9]

Chapter 2, however, diverts from this standard measure and contributes to the cross-country growth literature by picking up a suggestion by Ertur and Koch (2011) to base the interaction matrix on the genetic distance between countries. Genetic distance can be seen as a summary statistic that picks up a divergence across populations in char-

work, Smith notes at the end of the section that the "navigation of the Danube is of very little use to the different states of Bavaria, Austria and Hungary, in comparison of what it would be if any of them possessed the whole of its course till it falls into the Black Sea" (Smith, 1904, I.3.8.).

[8] Chapter 2 has been published in similar form as Working Paper No. 74 in the Working Paper Series in Economics at KIT (Deeken, 2015a).

[9] This is similar to a rise in the costs of trading information discussed in Chapter 1.

acteristics like habits, beliefs, norms or conventions that are slowly changing over time, and Spolaore and Wacziarg (2009) point out that this divergence creates barriers to the diffusion of development or technology, even after accounting for differences in e.g. language. Adaptation of innovations might be impeded by non-codifiable cultural differences between societies (Spolaore and Wacziarg, 2009, 513). Chapter 2 investigates whether the empirical results for geographic distance carry over to the alternative measure of genetic distance.

Naturally, other distance measures, like technological or institutional distances between countries come to mind that could be used instead of genetic distance to construct an interaction matrix. However, the model by Ertur and Koch (2007) leads to a reduced form whose theoretical implications are investigated with tools developed in the spatial econometric literature, and at this step the exogeneity of the interaction matrix becomes crucial. The objective is to determine within the context of the model how income per capita in a given country is influenced by changes in the variables, like the investment rate in physical capital, in the countries it is connected to. The reduced form includes a multiplicative term between the interaction matrix and the matrix of regressors. If now the weights in the interaction matrix were not exogenous, as would be the case for technological distance, then some form of interdependence between the interaction matrix and the regressors might exist (LeSage and Pace, 2014). Furthermore, the weights might change with changes in the regressors, complicating the interpretation of the model as the impact on the dependent variable from changes in the weights is hard to disentangle from the impact that results from changes in the regressors (ibid.). These problems are absent when the weights are exogenous as in the case of geographic and genetic distance.

Chapter 2 contributes to the literature also in another way. It investigates in a specific sense the important issue of robustness of empirical results. Even though data availability is a problem in many empirical studies, for the sample of countries investigated by Ertur and Koch (2007) the opposite is the case. Multiple versions of the same data set, the Penn World Table, are available and the default approach by researchers often is to use the most recent version. Rarely is it checked, if the empirical results are consistent across the various versions. Chapter 2 fills a gap in the literature by conducting extensive robustness checks for the influential Ertur and Koch model concerning this aspect. It is important to highlight that robustness is not a foregone conclusion. Issues for the empirical results in other studies have been detected (see Johnson et al. (2013)), with consequences for the conclusions drawn. In the current complex world, this issue is particularly pertinent when it comes to providing policy advice based on empirical analyses. Sound data is a requirement for this, and policy makers need to be aware if, for instance, implications hinge on the specific version of the data set that is used

Chapter 3[10] is linked to the modeling framework from the previous chapter, but shifts the focus in two ways. First, it makes the transition from exogenous to endogenous growth with technological interdependence by presenting a model by Ertur and Koch (2011) in detail. Second, it moves the aggregation level for the units of analysis from countries as in Ertur and Koch (2011) to regions within in a state. In particular, the empirical analysis focusses on the states in the United States. To the best of my knowledge, this model has not been empirically investigated in the literature for the US states before.

The United States is an important choice though as on a global level it accounts for a high proportion of investment in research and development (37% in 2001 according to the National Science Board 2014, 4-17) and foreign countries strongly benefit from spillovers due to research conducted in the United States (Eaton and Kortum, 1996). This chapter analyzes, if the individual US states also benefit from these knowledge spillovers.

The issue of how to specify interdependence that has been mentioned in the context of the analysis between countries in Chapter 2 continues to be valid in this chapter. However, at the level of US states no measures of genetic distance are available so that distances based on the exogenous characteristic of geography are used instead. In general, geographic distances are suggested to capture spillovers related to trade and foreign direct investment (Klenow and Rodríguez-Clare, 2005, 842),[11] but also differences in institutions (Ertur and Koch, 2007, 1036). The frictions associated with knowledge spillovers between states in a common institutional framework should thus be lower than globally across countries where this is not the case to the same extent. This provides an additional motivation for investigating the model's implications for the states in US.

The concluding chapter, summarizes the thesis' main results and provides an outlook concerning future research on the topic of spatial interaction and economic growth. Emphasis is put on the issue of sensitivity of empirical results to data quality.

[10] Chapter 3 has been published in similar form as Working Paper No. 75 in the Working Paper Series in Economics at KIT (Deeken, 2015b).

[11] Ertur and Koch (2011, 236) also point to this article for this statement. However, they refer to page numbers that are outside the range contained in Klenow and Rodríguez-Clare (2005). The reference given above points to the appropriate page number.

Chapter 1: Integration as a Spatial Institution: Implications for Agglomeration and Growth[12]

1.1 Introduction to Chapter 1

The 2009 World Development Report by the World Bank opens with the statement "Production concentrates in big cities, leading provinces, and wealthy nations. Half the world's production fits onto 1.5 percent of its land" (World Bank, 2009, xiii). This immense concentration of economic activity has its counterpart in the fact that urban areas currently account for more than 50% of the global population (United Nations, 2012). Economic activity and people are thus unlikely to be randomly distributed across space. Together these facts indicate the existence of benefits from concentration. These benefits are well known and date back to Marshall (1890) and his description of external economies. More specifically, these can be broken down to market size effects operating through forward and backward linkages, thick labor markets, and pure external economies like, for example, knowledge spillovers (Krugman, 1998). However, if only these agglomeration forces were at work, the implication was that the whole world would end up in one gigantic agglomeration. In effect, these forces that are conducive to agglomeration are balanced by a variety of dispersion forces like the existence of land rents or pure external diseconomies (e.g. pollution or crime). The resulting economic landscape is the outcome of the tension between these opposing forces.[13]

In addition to this concentration of economic activity, recent economic history over approximately the past 70 years is furthermore characterized by the ongoing process of globalization in its various forms as well as by an unprecedented increase in per capita income levels. Agglomeration and growth are connected via integration (globalization) to each other. In this chapter, key empirical facts on each aspect are presented and a selection of important models from the theoretical literature concerning these is commented

[12] Chapter 1 has been published in similar form as Deeken and Ott (2014a) and as Working Paper No. 59 in the Working Paper Series in Economics at KIT (Deeken and Ott, 2014b).

[13] Expressed differently, these forces are second-nature determinants for location decisions in contrast to first-nature or exogenous determinants like natural resources, climate or natural harbors.

on. A model by Baldwin and Forslid (2000) that combines an endogenous growth model of the Romer (1990) variety with a new economic geography model along the lines of Krugman (1991) is looked at in more detail, but without delving deep into the formal analysis. This combined model incorporates all three characteristics of interest and also allows for a more sophisticated consideration of the impact of integration on agglomeration than is the case in traditional new economic geography models. The reason is that it not only considers the costs of trading goods, but also accounts for the costs of trading information, a lowering of which tends to weaken the agglomeration forces. A detailed understanding of the concept of economic integration is particularly relevant, as it is the "way to get both the immediate benefits of the concentration of production and the long-term benefits of a convergence in living standards" (World Bank, 2009, 1).

The issue of size and scale economies, and the role integration plays herein is important in these considerations. For instance, the implication of many endogenous growth models that larger economies exhibit higher growth rates is not necessarily borne out empirically (Jones, 1995b). Another aspect is that scale economies might not be relevant at all levels of aggregation (i.e. city, region, nation). In particular, it is argued that future research should more precisely focus on integration as a dynamic concept that does not only affect agglomeration and growth, but which is itself the endogenous outcome of various interdependencies and which complements the institutional settings of the territories that are linked to each other.

The chapter is organized as follows. Section 1.2 presents stylized facts and theoretical issues on agglomeration, growth, and globalization. In Section 1.3 a model by Baldwin and Forslid (2000) is described that links these aspects. Section 1.4 deals with the issue of scale, size, and density while Section 1.5 derives policy implications from the insights of the previously presented arguments. Section 1.6 offers some concluding remarks and perspectives for future research.

1.2 Building Blocks

1.2.1 Spatial Concentration

The earliest urban structures date back to the time of the Neolithic Revolution (Bairoch, 1988). Prior to the Industrial Revolution, however, "the urban way of life had for thousands of years been the exception, it now became the rule" (Bairoch, 1988, 213). This tendency towards increasing urbanization continues today and reflects an ongoing global pattern as can be seen in Figure 1.1, which illustrates a positive trend that is expected to last during the next several decades.[14] Though the trend is global, there are marked

[14]The coarse aggregation scheme is chosen to ensure comparability of urbanization rates and GDP per capita throughout the subsequent section. Since there is no perfect overlap in data availability

differences between the various 'global regions'. Western Europe, the United States and Australia have crossed the threshold of more than half of their respective population living in urban areas before 1950 and now have urbanization rates between 75% and 90%. Africa and Asia on the other hand had urbanization rates of ca. 15% in 1950 and are expected to cross the value of 50% in 2035 and 2020, respectively. Latin America has followed yet another path. It reached an urbanization rate of 50% in the early 1960s, had a higher rate than Western Europe in 2000, and currently has a rate of nearly 80%.

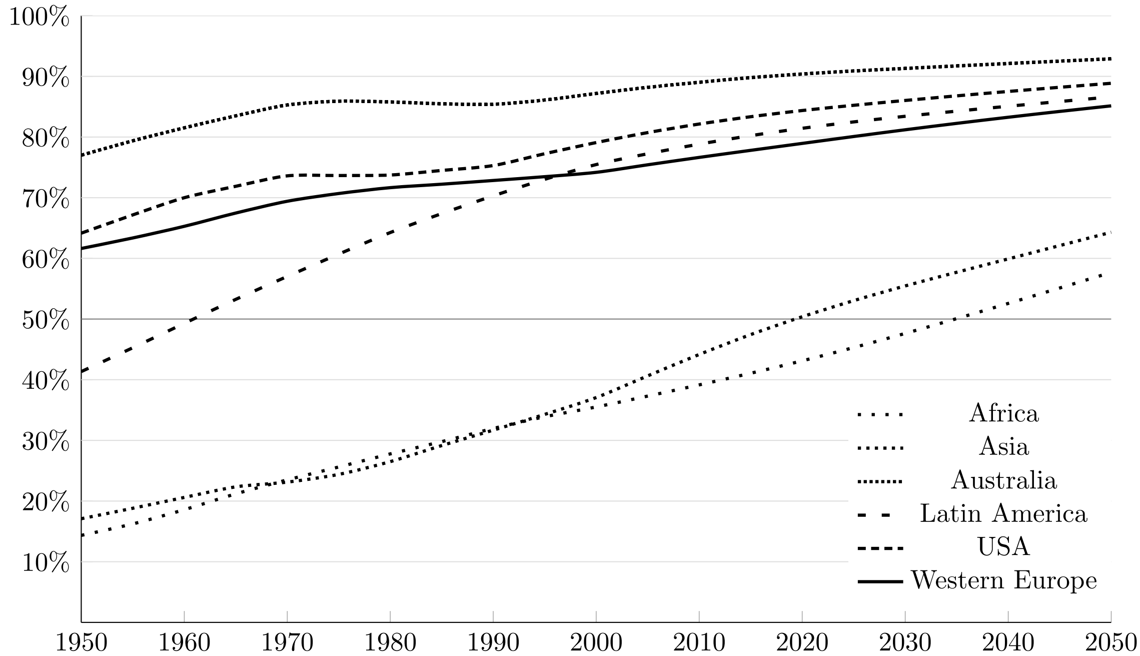


Figure 1.1: Urban Population by Major Geographical Area.

Note: The data are from United Nations (2012). See Appendix A for a list of geographical entities included in the aggregates.

These numbers for the urbanization rate provide little information, however, on whether people are living in large cities of several million people or if they are living in comparatively small cities of a few hundred thousand inhabitants. Table 1.1 remedies this situation to some extent by distinguishing the absolute number of urban residents according to five city-size classes.[15] What can then be inferred from combining the data in Figure 1.1 and

as regards urbanization rates (Figure 1.1) and growth (Figure 1.2), the notion of 'global regions' from the United Nations World Urbanization Prospect is adopted and applied to the New Maddison Project Database (Bolt and van Zanden, 2013). Details on this can be found in Appendix A. In Figure 1.1, the concept of 'urban population' is based on the definitions of the respective national statistical agencies and thus may vary across regions. For an alternative agglomeration measure that aims at enabling cross-country comparability see Uchida and Nelson (2010). Note that the population forecasts (from 2012 onwards (United Nations, 2012)) depend upon national census data that are also only comparable with restrictions.

[15]The entries in Table 1.1 are absolute numbers of urban residents. Therefore, the data offers only limited information about the number of cities in each group. The 630 million people, who are forecasted

Table 1.1 is that not only are more and more people living in cities, but they are also increasingly living in large cities. In 1970, for instance, approximately 14% of the world's population lived in cities of more than 500,000 inhabitants, whereas the corresponding figure was approximately 26% in 2011 and will rise to approximately 33% in 2025.[16]

The share of the urban population living in cities with less than 500,000 inhabitants is forecasted to fall by 20 percentage points from 62% in 1970 to 42% in 2025, but the share of urban residents in cities with more than 10 million inhabitants is expected to increase from 3% to 14% over the same period. Combining these aspects with the empirical evidence on city-size distributions implies the emergence and even intensification of core-periphery structures at various levels of spatial scale.

Table 1.1: Evolution and Forecast of Total Population in Millions According to City-size Classes.

	< 0.5m	0.5-1m	1-5m	5-10m	> 10m
1970	833	128	244	109	39
1990	1333	206	456	142	145
2011	1849	365	776	283	359
2025	1966	516	1129	402	630

Note: The data are from United Nations (2012).

In other words, the trend of ongoing concentration may be observed at different levels of spatial aggregation. Brakman et al. (2009, 13) neatly summarize this phenomenon by stating: "It appears that the highly uneven distribution of economic activity across space has a fractal dimension – that is, it repeats itself at different levels of aggregation." Put differently, ongoing concentration might be observed across several spatial scales whereupon from a global perspective the resulting core-periphery structure remains unchanged.

These facts on urbanization are strong indicators for the existence of 'local' scale effects. The resulting spatial pattern is the outcome of the location decisions of firms and households, and the underlying economic reasoning may be summarized as follows: Individuals are indifferent as regards relocation if benefits and costs are equalized. In other words, for any degree of aggregation a spatial equilibrium is reached whenever forces attracting people (so-called agglomeration forces) and those pushing off people (dispersion forces)

to live in cities of over 10 million residents in 2025 might, for instance, be distributed more or less evenly over some 60 cities or on the other hand be concentrated in a few gigantic cities of 50 million residents and many comparatively small ones with "only" approximately 10 million residents. Any other combination that distributes 630 million residents over a number of cities with a minimum size of 10 million is also possible. More details on the city-size distribution can be found in (Gabaix and Ioannides, 2004).

[16] The shares are calculated from the data in United Nations (2012, Figure II and Table A.5).

are balanced.[17] From a slightly different perspective, a location may be attractive for people or firms due to characteristics that are external to these actors.

The associated agglomeration (or external) economies that act as attractors for firms and individuals to a certain place date back to the seminal work of Marshall (1890), who identified three main sources of spatial concentration processes:[18] (i) *Market-size effects:* Local concentration leads to large local markets and vice versa. Via 'forward' or 'cost linkages' a local concentration benefits both consumers, who profit from more varieties and lower prices, as well as firms, since the local production of intermediate goods reduces the costs of downstream producers. On the other hand, firms gain from producing in a large market with good access to customers ('backward' or 'demand linkages'). (ii) *Thick labor markets* ease the matching problem between supply and demand of (today frequently specialized) labor. (iii) *(Pure) external economies* that are the more distinct the more densely populated a region is, since proximity allows for more frequent interaction and increased information spillovers. Infrastructure fixed costs are also spread over more heads. These forces mostly cover *specialization advantages* as the considered actors can benefit from scale economies and the related cost decreases, which usually are meant to apply within single sectors.[19]

Equally important for the attractiveness of a certain location, though not specifically mentioned in the context of Marshall's triad, are: (iv) *Selection effects* that occur in highly competitive markets where only efficiently working firms are able to survive. This effect is reinforced, since efficiency acts as an attraction point for internationally mobile capital and/or frequently highly qualified labor. (v) *Diversity:* Especially in light of the aforementioned increase of large urban agglomerations (Table 1.1), it is quite reasonable that the economic centers are not only characterized by specialization in a single but in several different fields at the same time. Groups of interconnected companies together with the supporting institutions, which are all located close to each other, are frequently said to form a cluster (Porter, 1990). In particular, if several clusters are co-located, it is not just specialization, but in contrast the opportunity to interact with various – also heterogeneous – actors that attracts new firms and people. In this connection, Jacobs

[17] This argument holds with the caveat that in formal theoretical models it is only valid for interior equilibria, but not for equilibria in which complete agglomeration occurs.

[18] In this context, Alfred Marshall shaped the notion of the 'industrial districts' – a formulation that must be understood in light of the 19th century's economic conditions. The basic mechanisms, however, are still valid today only that aside from the industrial sector an important share of value creation is realized within the service sector. Marshall's concept has been picked up by Arrow (1962) and Romer (1986), who introduced it in theoretical models. As a consequence, agglomeration economies that are related to spillovers and (industry) specialization are mostly denoted as Marshall-Arrow-Romer (MAR) externalities; see also Krugman (1998) for a compact overview.

[19] Considering current production conditions, this includes the service sector and the joint use of research infrastructure.

(1969) was the first to point to the positive effect of the co-location of diverse actors.[20] The resulting productivity and consumption economies are associated with the diversity and intensity of economic activity. In addition, large agglomerations are also more efficient in solving the matching problem that may arise in the factor markets thus linking the reasoning to the aforementioned thick labor markets and adapting them to a more dynamic environment that also might account for structural change. It is therefore reasonable to assume that as a consequence of 'density', the corresponding region's productivity increases and thereby drives local productivity above average.

But if only agglomeration forces were at work, the spatial equilibrium would result in a unique agglomeration of economic activity thus contradicting the empirical facts summarized in Figure 1.1 and Table 1.1. In fact, there exist opposing dispersion forces that are mainly based on the following sources: (i) *Immobile factors* such as land, natural resources, but also workers, imply that industries have to some extent go to where factors and their owners are located. At the same time, institutional arrangements may also foster or hamper the mobility of factors. Two descriptive examples are the four freedoms within the EU or the necessity of obtaining work permits for non-residents in the US. (ii) *Land rents* increase as a result of concentration both for producing firms, but also for private individuals, who seek to minimize commuting costs. In the extreme, increasing land rents may even lead to the relocation of economic activity not only from the core to peripheral regions, but also as regards the location of different operating areas of firms within urbanized areas (functional specialization, Duranton and Puga, 2000). (iii) *(Pure) external diseconomies of scale* such as congestion, pollution, and crime also hamper concentration.

As argued before, a spatial equilibrium is reached when both, agglomeration and dispersion forces, are balanced. Ongoing urbanization in this context implies the emergence and reinforcement of core-periphery structures in economies with increasing population size. In Section 1.4 the identified core-periphery structures are related more precisely to scale effects and size in growing economies.

1.2.2 Growth and Prosperity

Aside from the spatial concentration of population and production at certain locations, another major trend since the era of industrialization is ongoing economic growth as measured by gross domestic product (GDP) per capita. Figure 1.2 depicts GDP per capita based on 1990 international dollars[21] for the period 1850-2010 and strikingly highlights the

[20]The associated external effects are nowadays called Jacobs or urbanization externalities (thus being distinguished from the previously mentioned MAR or specialization/localization externalities).

[21]In this unit 1 international dollar has the same purchasing power as $1 US had in 1990, and the GDP per capita values in other currencies are adjusted by purchasing power parities. Keeping in mind the difficulties of estimating these time series (see e.g. Bolt and van Zanden (2013)), it is nonetheless to a certain degree possible to compare the values across time and countries.

world-wide growth story for the same major global regions as in the preceding section.[22] Around the world, centuries of Malthusian stagnation in which per capita income was near the subsistence level precede the Industrial Revolution until economies entered a period of sustained economic growth.[23] The positive trend might already be discerned in the 19th century. However, the effect of exponential growth becomes distinctly visible after World War II, though there are clear differences between the various illustrated global regions.

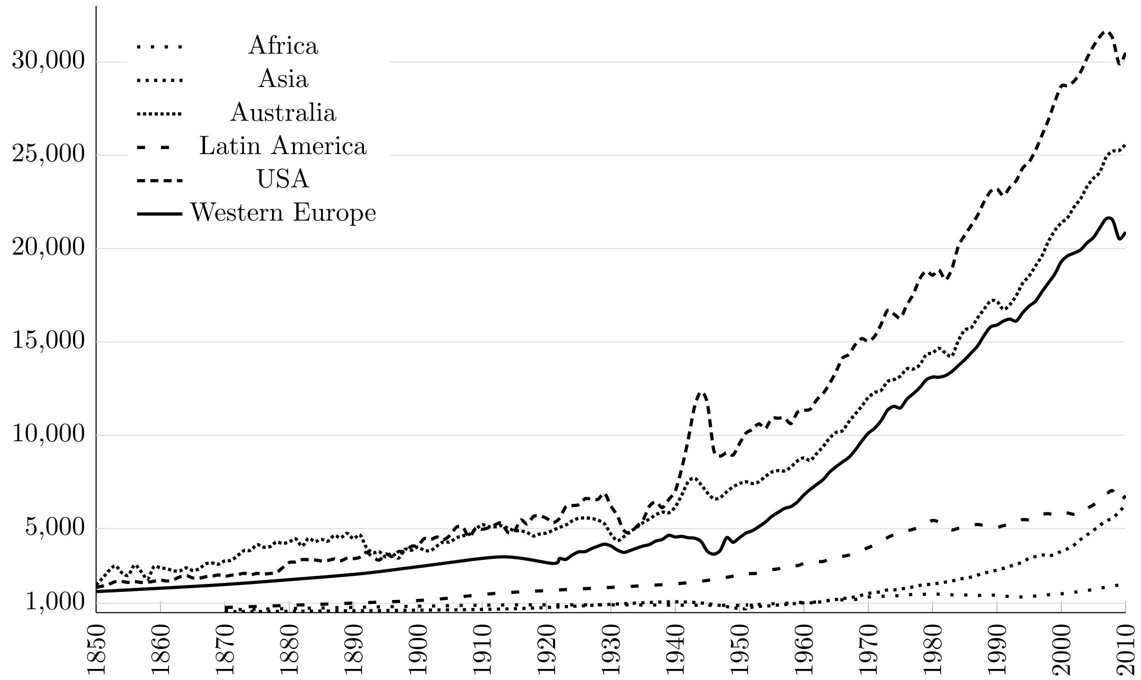


Figure 1.2: GDP per Capita in 1990 International Dollars.

Note: The data are from the New Maddison Project Database (Bolt and van Zanden, 2013). See Appendix A for a list of geographical entities included in the aggregates.

The dominating positive trend for the global regions tends to hide drastic differences in income at the less aggregated level of individual countries. At the top, the USA has a GDP per capita of $30,491 in 2010 while the poorest country in the Maddison sample in that year is the Democratic Republic of the Congo with a GDP per capita of $260. Another noteworthy characteristic of the data is the possibility that a laggard overtakes the leader in the GDP per capita ranking ('leapfrogging'). This is exemplified by the USA becoming richer than the previous leader Australia in 1899.[24]

[22]The New Maddison Project Database provides information on the economic performance of countries and geographic regions from the year 1 onwards (at least for a subset of countries). The selected regions were chosen to allow for an immediate comparison to Figure 1.1.

[23]See, for instance, Galor (2005, 2011) and the references therein for empirical evidence and a theoretical model explaining this transition to sustained growth.

[24]For additional empirical evidence on the changing leaders and laggards for the period 1-2003, compare e.g. Figure 9 in Brakman and van Marrewijk (2008).

The pattern of ongoing growth in Figure 1.2 has been accompanied by additional economic regularities. Kaldor (1957, 591) stated for capitalist economies that there are "remarkable historical constancies revealed by recent empirical investigations." His observations led to the nowadays well recognized 'stylized facts of growth' (Kaldor, 1961).[25] At the same time, these observations set the frame for a research agenda to develop a consistent, comprehensive theory, which led to the *neoclassical growth model* (Solow, 1956; Swan, 1956). Based on the empirical observation of ongoing economic growth – often interpreted as increasing prosperity that in turn allows for a life above subsistence level – economic theory in general tries to gain a deeper understanding of the driving forces of this process. The major goal is to identify the underlying determinants, to comprehend their interaction, and, given any indications for forces hampering this process (e.g. market failures), to derive appropriate policy recommendations in order to maximize overall welfare. Growth thereby refers to the evolution of the GDP per capita, mostly analyzed at the country level, which results from the accumulation and use of more and/or better inputs in the production process. Across time and in spite of its rich explanatory power of the stylized facts detailed in Footnote 25, the neoclassical growth model has been criticized with respect to several dimensions (e.g. Jones and Romer, 2010). It only covers one state variable, namely physical capital. The lacking microfoundations imply an exogenously assumed savings rate, since the resulting growth rate is not derived from individual optimization behavior. The model is also not able to explain the variation of growth rates between countries. In the model's equilibrium, growth per capita comes to a standstill with the consequence that the empirically observed positive growth rates can only be explained by exogenous technological progress. Due to its specification, technological change naturally remains a black box so that ultimately no clear-cut policy recommendation can be derived from the model in this respect.

Solow's basic growth model set the ground for several extensions, among them the inclusion of human capital or productive governmental activity. Further efforts led to the emergence of *endogenous growth theory* in the mid 1980s, which also provided insights into the utility maximization–growth nexus (i.e. the derived aggregate growth rate is based on individual optimization decisions (microfoundations)) thereby especially addressing the role of technological progress, human capital accumulation, or the role of institutions. Recent discussions of growth theorists distinguish 'proximate' from 'fundamental' causes of growth. The latter cover conditions such as luck, geography, culture, and institutions

[25] In detail, the Kaldor facts are the following: (i) Labor productivity has grown at a steady rate, (ii) capital per worker has continually increased, (iii) the real interest rate (return on capital) has been stable, (iv) the capital-output ratio has been constant, (v) the shares of capital and labor in national income have been stable, and (vi) among the fastest growing countries the growth rate has varied in the range of 2–5%.

while the former refer to production inputs such as physical and human capital as well as to their overall productivity, which is enhanced by technological progress.[26]

These new theoretical frameworks have now themselves undergone an empirical assessment. Again, the goal has been to review whether or not they are suited to contribute to a better understanding of the empirical facts and thus to link theory and empirical research within a consistent framework. Jones and Romer (2010) recently revisited the aforementioned Kaldor facts and updated them thereby also accounting for structural change over the last five decades.[27] Summarizing these findings, leads to the recognition that in order to match the requirements of today's empirical results and to develop a consistent theory, aside from physical capital more state variables – namely human capital, ideas, population – and also institutions as reflecting the "rules of the game" (North, 1990, 3) need to be considered. In the context of growing economies, the important role of institutions is emphasized in the seminal work of Acemoglu et al. (2001) and highlighted as well by, for instance, Rodrik et al. (2004).[28]

Altogether, within growth theory usually the platform of analysis is aggregate economies (countries or continents) where spatial components are not explicitly considered.[29] In case they are, the notion of 'geography' refers to natural conditions, which include a country's endowment with natural resources or its climatic conditions (e.g. Gallup et al., 1999; Dell et al., 2012). Put differently, as regards the spatial dimension, both the models and the stressed empirical studies refer to what sometimes is called 'first-nature geography', but do not make explicit man-made 'second-nature geography' conditions in a territorial sense.[30] The latter might, however, to some extent be understood as or be linked to institutional settings. Especially in Section 1.4 it is argued that the degree of integration might be interpreted as reflecting an explicit spatial dimension of institutions. In addition, Jones and Romer's (2010) stylized fact of 'increases in the extent of the market' implicitly

[26]Excellent overviews on issues related to economic growth are provided by Acemoglu (2009) – with a special emphasis on the distinction between fundamental and proximate causes of growth – or more broadly by Barro and Sala-i-Martin (2004).

[27]In detail, the Jones-Romer facts identify the following empirical regularities: (i) Increases in the extent of the market, (ii) accelerating growth, (iii) variation in modern growth rates, (iv) large income and total factor productivity differences, (v) increases in human capital per worker, (vi) long-run stability of relative wages.

[28]Note that in many discussions and theoretical models, the concept of institutions is often connected to the presence of property rights. Throughout this chapter, however, a broader interpretation is assumed and institutions are embedded as regards their impact on the degree of integration – as later e.g. in Section 1.3 where institutions are seen as being related to the variable 'freeness of trade' and the environments that allow for knowledge spillovers.

[29]A winged word in this context is that 'the world is flat', which refers to the title of a book by Friedman (2005).

[30]An exception is Bosker and Garretsen (2009), who demonstrate that second-nature (or relative) geography interpreted as the institutional quality of neighboring countries has an impact on a country's economic development.

incorporates a spatial dimension, as the observed increase is the outcome of both growth (at any place) and access to foreign markets due to globalization which extends the relevant market for any economy involved in (international) trade. To conclude: Until today, spatial aspects are, if at all, only implicitly or indirectly addressed in most models of endogenous growth.

1.2.3 Globalization

A glimpse at the discussion so far suggests that "agglomeration can be considered the territorial counterpart of economic growth" (Fujita and Thisse, 2002, 389). However, this viewpoint is too simplistic and without any explanatory power as regards the underlying interdependencies that drive this result given that countries or regions are not isolated actors, but are increasingly embedded in international value-creation processes in which due to globalization the places of production and consumption frequently are not located within the same national borders. Globalization in this context covers several dimensions, namely mobility of goods, people, ideas, and capital.[31] The exchange of these items is the immediate outcome of increased integration.[32] Important grounding in this context is given by rules that facilitate international/cross-border economic activity like the reduction of trade barriers, reduced transportation costs or the recognition of foreign degrees. The key driver that fosters the joint emergence of growth and agglomeration, however, is the increase in the international trade of goods (compare Baldwin et al. 2001). Concerning mobility of people, migration mainly occurs within, but also between countries (World Bank, 2009, 147), and global migration is increasing at least in absolute numbers. The world-wide stock of migrants has risen to 165 million in 2000 from a starting value of 92 million in 1960, although the migrant's share of the world population has fallen from 3.05% to 2.71% over this period (Özden et al., 2011, 15). For migration flows, analysis by Abel and Sander (2014) suggests that for the three 5-year periods from 1995 to 2010 a relatively stable share of approximately 0.6% of the world population has migrated internationally. With respect to the mobility of ideas, Jones and Romer (2010, 229) note that, for instance, the change in the share of patents granted to non-US entities by the U.S. Patent and Trademark Office can be interpreted as an indicator for the international flow of ideas. This share has increased from 18% in 1963 to 52% in 2012.[33]

[31] The latter, for instance, via foreign direct investment (FDI). The concept may be understood even more broadly as in the KOF Index of Globalization, which includes, amongst other things, information on cultural proximity (see `http://globalization.kof.ethz.ch/` (accessed: 11 August, 2015)).

[32] See Meissner (2014) for additional details on the concepts of globalization and integration.

[33] Compare the column 'Total Patent Grants, Foreign Origin Percent Share' under `http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm` (accessed: 11 August, 2015). The exchange of ideas is also immensely fostered by the fall in communication costs and the accompanying spread of the internet to which approximately 35.6% of the global population had access in 2012, whereas the corresponding shares were 0.05% in 1990 and 6.7% in 2000 (World Bank 2013; series IT.NET.USER.P2).

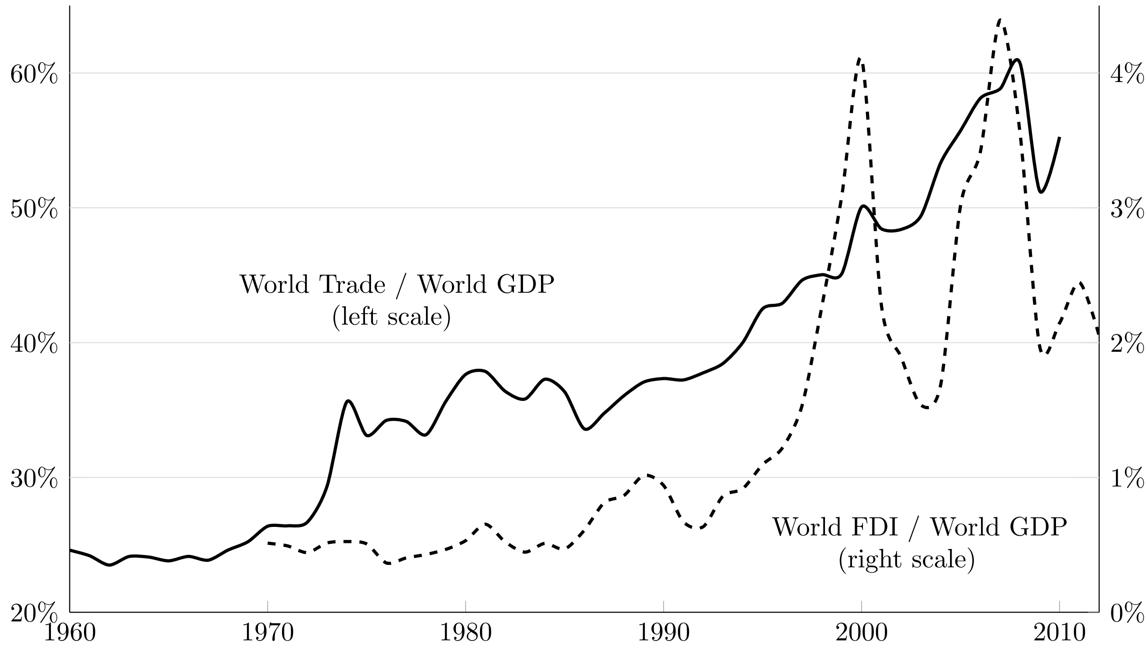


Figure 1.3: Globalization.

Note: World FDI over World GDP is series BX.KLT.DINV.WD.GD.ZS from the World Bank's *World Development Indicators* (World Bank, 2013), and the variable World trade / World GDP is calculated from Heston et al. (2012).

Figure 1.3 illustrates the phenomenon of globalization along the two dimensions 'trade' (mobility of goods) and 'FDI' (mobility of capital), which both have significantly increased during the last several decades although not steadily so. The sharp drop around the year 2000 coincides with the bursting of the dot-com bubble, and the onset of the recent financial crisis has led to a setback in the trend towards increasing globalization. These effects are particularly stark for the series of the ratio of World FDI to World GDP.

The increase in international trade over the past 50 years shown in Figure 1.3 overlaps with a period of ongoing trade liberalization. One measure for liberalizing trade are preferential trading agreements, and the cumulative number of these agreements in force has increased from a value in the single digits in the early 1950s to nearly 70 in 1990 before reaching almost 300 in 2010 (World Trade Organization, 2011, 54-55).[34] On a closer look, the increase in world trade over time hides an important aspect concerning the evolution of the composition of world trade. The share of *intra-industry* trade as opposed to *inter-industry* trade has increased as well. Brülhart (2009, 426) notes that the share of global *intra-industry* trade rose from roughly a quarter in 1962 to over 50% in 2006.[35]

[34]This is not to say that trade liberalization is the only or most important factor in the increase in world trade. See Baier and Bergstrand (2001) for an analysis of the relative importance of trade liberalization, transport costs, and other factors in this development.

[35]The underlying measure for *intra-industry* trade is the Grubel-Lloyd index at the 3-digit Standard International Trade Classification (SITC) level. At the more detailed 5-digit SITC level *intra-industry* trade has increased from ca. 10% to about 30% over the period.

A noteworthy aspect that accompanies the increasing share of *intra-industry* trade underlying this development is the composition of this form of trade in horizontal and vertical versions (see e.g. Thom and McDowell 1999 for this distinction). *Horizontal* intra-industry trade refers to trade in differentiated products at the same processing stage in a sector and is closely related to the love-of-variety effect that is, for example, present in the model discussed in the subsequent section.[36] *Vertical* intra-industry trade on the other hand takes place within a given sector, but at different stages of processing and is linked to comparative advantage and its resulting specialization. Distinguishing between final and intermediate goods instead, Brülhart (2009) notes that trade in final products has increased globally from a starting value slightly above 10% in 1962 to roughly a third in 2006, and the corresponding series for intermediate goods has followed the one for final goods closely until the mid 1970s, after which the shares were consistently higher and reached about 40% in 2006.[37]

On a theoretical level, neoclassical international trade theory in the form of the Ricardian and Heckscher-Ohlin models relies on the presence of comparative advantage due to different technologies or relative factor abundances to explain *inter-industry* trade between countries. However, neither is all trade between countries of the *inter-industry* variety nor is all trade between countries that differ in income, size, and relative factor endowments. Germany, for instance, exports cars to France and vice versa. In order to explain this *intra-industry* trade that takes place even in the absence of comparative advantage various assumptions of the neoclassical trade theory need to be relaxed. This is done in the field of new trade theory, which, for instance, no longer assumes that firms produce with a constant returns to scale technology and operate in an environment of perfect competition. Instead, increasing returns to scale at the firm level and imperfect competition are introduced.[38] The second change concerns the assumption of a homogeneous product in the neoclassical trade models, which is replaced in new trade theory by heterogeneous goods which are assumed to be imperfect substitutes for the consumers who exhibit love-of-variety preferences that are captured via a constant elasticity of substitution function.[39] The most widely adopted model of monopolistic competition with these characteristics is the one introduced by Dixit and Stiglitz (1977). Notice that these assumptions also mirror the transition from neoclassical to endogenous growth theory discussed in Section 1.2.2.

[36] This effect is however not constrained to final goods, but is also relevant for intermediate goods or services in firms' production processes (Hewings and Oosterhaven, 2014, 912-913).

[37] Data in Brülhart (2009) distinguishes only between trade in intermediate and final goods, but not between *horizontal* and *vertical* intra-industry trade. The shares are for the 5-digit SITC level.

[38] As Davis (1995) notes though, it is not necessary to assume increasing returns to scale at the firm level, since *intra-industry* trade can be accounted for without this assumption based on comparative advantage.

[39] These preferences imply that consumers receive more utility from consuming e.g. one unit each of seven different varieties instead of seven units of one particular variety.

Increased freeness of trade also enhances international competition between firms and industries and induces possible relocations of economic activity allowing for production at the most productive places. The consequences on the spatial distribution thereby involve two dimensions. Concerning *inter-industry* trade, industries that exhibit a comparative advantage will grow, whereas the disadvantaged industries will shrink. As regards *intra-industry* trade, liberalization increases international competition, and some firms are not able to cope with these conditions. Empirical studies have shown that the corresponding reallocation of firms is more pronounced as regards *intra-industry* than *inter-industry* trade (Brakman et al., 2009).

Analogous to the reasoning in Section 1.2.1, the resulting spatial equilibrium is the outcome of the interaction between local increasing returns to scale and trade costs in which agglomeration and dispersion forces offset each other. Reduced trade costs (which enhance the freeness of trade) are frequently interpreted as being the outcome of increased economic integration between formerly more or less autarkic economies. The driving force in all these models is a reduction of trade costs or put differently, increased integration. Trade costs are mostly argued to capture a reduction of trade barriers such as duties or other non-tariff barriers. Major components are also transportation costs that have undergone a significant decline throughout the last several decades.[40] This implicitly incorporates aspects of technological change, which allows e.g. for an increase in shipping capacities, but also reduced communication costs that are due to improvements of information and communications technology. Nevertheless, within the considered models and empirical studies the degree of integration is assumed to be exogenous. In Section 1.4 it is argued that integration itself undergoes an evolution together with the processes of growth and agglomeration. In order to capture the implications, integration should be related to the institutional view detailed throughout Section 1.2.2. A more pointed view would be that integration is a fundamental cause for the spatial shape of the economic landscape and has explicitly to be seen as a dynamic institution with a strong spatial dimension.

1.3 Model

Section 1.2 described a variety of stylized facts for three defining characteristics of 'modern' economic history: Growth, agglomeration, and integration. The former part of this chapter detailed these several lines of argumentation. Due to the dynamic perspective and the various interacting effects, their final impact on the spatial distribution of economic activity in growing economies is far from being trivial. To provide some guiding to

[40]The decline in costs is clearly present for air shipping, but less pronounced for ocean shipping as Hummels (2007) notes.

disentangle the main lines of reasoning within a consistent model, this section presents, without going deeply into its formal structure, a model by Baldwin and Forslid (2000) that highlights the connections between these three characteristics and thus attempts to account for the joint endogeneity of the location of industrial activity and long-run economic growth.

Baldwin and Forslid's starting point is the two region core-periphery model developed by Krugman (1991), which takes care of agglomeration and integration. In this model, there exists a traditional sector with constant returns to scale and perfect competition and an industrial sector with increasing returns to scale at the firm level, which is characterized by monopolistic competition. Goods from both sectors are traded, although trade costs occur only in the manufacturing sector. Whereas the global labor supply and the labor supply in the traditional sector are fixed, labor in the industrial sector is mobile, and the interregional labor distribution in this sector is determined endogenously, since migration depends on differences in the real wages in the two regions.[41] This model setup produces circular (cumulative) causation, which can be broken down to three forces that ultimately depend on the degree of integration as measured by transportation/trade costs. Two of these, the backward and forward linkages, are agglomeration forces and have already been described in Section 1.2.1 under the label market-size effects. The third force is the competition or market-crowding effect, which works against agglomeration, as firms prefer to locate away from their competitors. Whether or not a process of cumulative causation will be set in motion, depends on the relative strength of these forces. If, for instance, agglomeration forces are stronger than the dispersion force, then a shock to the system has the result that all industrial activity will locate in a single region, whereas the immobile factors of the traditional sectors determine the economic power of the periphery.

Baldwin and Forslid (2000) combine this framework with the endogenous growth model by Romer (1990) thereby incorporating the third characteristic. Both models are built upon the Dixit-Stiglitz approach to modeling monopolistic competition (Dixit and Stiglitz, 1977) mentioned in Section 1.2.3. This identical understructure greatly facilitates the integration of the two separate models. The new aspect in Baldwin and Forslid's model is to not only consider the implications of changes in the trading costs for goods as in standard new economic geography models, but also the implications of changes in the trading costs for ideas. New ideas or knowledge are the driving force of economic growth in the model, and they ultimately show up in the form of an increasing number of varieties produced in the monopolistically competitive manufacturing sector. Producing a new variety requires a fixed cost of one unit of capital K in the model in addition to a variable cost for labor.

[41] In a slight modification to the Krugman (1991) model, Baldwin and Forslid (2000) allow for forward-looking behavior in the migration decision instead of static expectations.

Capital in this model is viewed as "new knowledge embedded in a manufacturing facility that is immobile across regions" (Baldwin and Forslid, 2000, 310). Production of capital requires only labor as an input and takes place in the model's third sector, the investment good or innovation sector, which is characterized by perfect competition. The crucial feature of this sector is the presence of knowledge spillovers or technological externalities in the sense that the unit labor requirement falls with an increasing level of production in the investment good sector. The specific distribution of manufacturing activity over the two regions moreover has a bearing on the extent of knowledge spillovers. More precisely, knowledge accumulated in a given region, e.g. the north, is more beneficial to firms in the north than knowledge accumulated in the south. This specification receives empirical support by, for instance, the work by Eaton and Kortum (1996, 276), who demonstrate for the OECD that even though there is substantial diffusion of technology between countries, large impediments to its diffusions exist that "are sufficient to generate large differences in productivity across countries." Put differently, location matters for firm productivity.

Baldwin and Forslid formally model production in the innovation sector via the following production function:

$$Q_K(t) = \frac{L_I}{a_I(t)}; \qquad a_I(t) = \frac{1}{K(t-1) + \lambda K^*(t-1)}; \qquad 0 \leq \lambda \leq 1$$

in which the variable $Q_K(t)$ denotes the flow of new capital at time t, and employment in the investment sector is denoted by L_I. K and K^* are capital in the north and south, respectively, and the parameter λ signifies the degree of knowledge spillovers. The one-period lag, $t-1$, for the capital in the two regions in the expression for the unit labor requirement in the investment goods sector, $a_I(t)$, indicates that it takes a certain amount of time until knowledge produced in one region becomes available in the other region.[42]

A complete analytical derivation of the results in Baldwin and Forslid (2000) is outside the scope of this chapter. Suffice it to say at this point that the model's dynamics can be completely described by a system of three difference equations in the variables for labor in the north, $L(t)$, the north's share of global capital, $K(t)/\big(K(t)+K^*(t)\big) \equiv \theta_K(t)$, and the shadow value of migration, $W(t)$. As in the standard core-periphery model, three stable long-run equilibria exist:[43] One in which manufacturing activity is spread symmetrically between the north and the south, and in the other two equilibria manufacturing activity is either completely agglomerated in the north or the south. The stability of these equilibria is verified via an analysis of the eigenvalues of the Jacobian matrix of the system of difference equations linearized around the steady state under consideration.

[42]Baldwin and Forslid (2000, 313) take one period to last approximately 10 years.
[43]Additional unstable interior equilibria exist as well.

Despite the fact that the model is analytically rather complex, it has the neat feature that the stability properties of the respective equilibria depend only on two parameters. One is the already mentioned λ, denoting the degree of knowledge spillovers. The higher its value, the less localized are the technological externalities. Hence, this parameter can be interpreted as representing the cost of trading information. The second parameter, ϕ, is an index capturing the notion of freeness of trade and can be shown to vary between 0 and 1.[44] This specification makes it possible to summarize information on the stability of the equilibria in the comparatively simple diagram shown in Figure 1.4, which is divided into three sections.

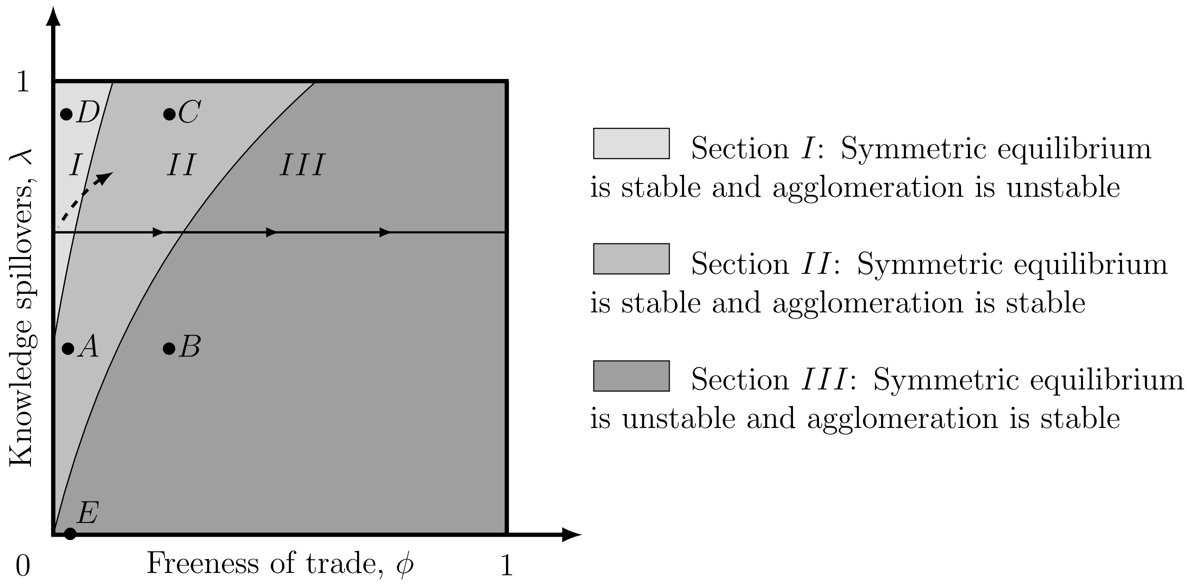


Figure 1.4: Stability of the Equilibria in Dependence on λ and ϕ (Adapted from Baldwin and Forslid (2000) and Brakman et al. (2009)).

Analysis of Figure 1.4 shows that the results from the standard core-periphery model of the Krugman variety carry over to the growth-augmented model. A fall in transport costs (equivalently a higher value for ϕ) has the implication that agglomeration becomes the only stable equilibrium.[45] This process is illustrated by the horizontal line with the three arrows in Figure 1.4. For a high enough level of knowledge spillovers, the economy moves from a situation in which only the symmetric equilibrium is stable (Section I) through a situation in which both agglomeration and the symmetric equilibrium are stable (Section II) to one in which only agglomeration is a stable equilibrium (Section III). This destabilizing aspect (in the sense of a tendency towards agglomeration) of closer integration is linked to the presence of an additional agglomeration force due to the introduction of endogenous growth into the model. However, the presence of knowledge spillovers influences the

[44] For infinitely high trade costs the index is zero, whereas in the absence of trade costs it equals one.

[45] Note that Figure 1.4 only establishes that agglomeration is a stable equilibrium, but not which region is or becomes the core. To determine this, the initial conditions and the specific shock that disturbs an unstable equilibrium need to be analyzed.

strength of this third circular causation chain and Baldwin and Forslid (2000) indicate that the strength of this effect can be counteracted by the model's second policy parameter, λ. Hence, combining a lowering of the costs of trading goods with a lowering of the costs of trading ideas opens up the possibility that the symmetric equilibrium remains stable for a wider range of values for ϕ. This possibility is shown by the dashed arrow moving from Section I to Section II.

It is also possible that a policy that lowers the costs of trading information sufficiently, leads to a spreading out of industrial activity. In Figure 1.4 this is captured by a 'world' economy starting in point A in a core-periphery equilibrium, which then becomes more integrated through higher knowledge spillovers and moves to point D in which only spreading is a stable equilibrium.

Therefore, integration needs to be viewed as a more complex process than in standard new economic geography models and embrace aside from mobility of goods and people also mobility of ideas and capital. Ever closer integration through a given reduction in the costs of trading goods does not necessarily lead to complete agglomeration. This process can be counteracted (compare a movement along the dashed arrow or from point A to C with a policy that moved the world economy from A to B) and possibly even reversed[46] through adequate policies that lead to a fall in the costs of trading information. Note that although integration is now more broadly specified than in the standard new economic geography model, it is still specified as an exogenous and static concept.

1.4 Size vs Scale: When, Where, and Why Does It Matter?

The considerations so far have mainly been focussed at a highly aggregate level, be it the arbitrarily chosen global regions highlighted within the empirical presentations in Figures 1.1 and 1.2, the world-wide view in Figure 1.3 or the economy-wide perspective assumed within the discussed theories in Sections 1.2.2 and 1.2.3 as well as within Section 1.3 and Figure 1.4. Besides, it is nearby to assume that the aggregation level also determines the size of the considered economy. Benefits of large economies arise e.g. as size co-determines the financial, political, business and cultural environments of people and firms. One example is the provision of public goods like national defense, public security, or the judicial system, for which the costs in larger nations can be spread over more taxpayers. Further benefits of large countries are insurance against asymmetric

[46]Consider a situation in which both regions are living in autarky, introduce some free trade which moves the world economy to point E where the core-periphery equilibrium is stable and then drastically reduce the cost of trading information so that the world economy ends up in point D where only spreading is a stable equilibrium.

regional shocks[47] or the possibility of nation-wide redistribution schemes that affect the after-tax income distribution in a way not feasible if the territories were independent entities (Alesina and Spolaore, 2003, 3-4). Counteracting these positive effects of size are more heterogeneous preferences in larger countries, leading to a higher probability that some subset of regions (or individuals) in a nation is not in agreement with the policies of the central government and thus poses a danger to its stability (Alesina and Spolaore, 2003, 4-5). Ethnic, linguistic, and religious heterogeneity play a role in this context. More prosaically, administrative costs may rise with ever larger size and issues of congestion, crime, and pollution may become a problem. It is thus nearby to assume that there is some endogenously resulting 'optimal' size of economic spaces, which differs according to their respective characteristics and evolves as production conditions, dynamic environments, or institutional settings (especially of interest are the rules underlying international trade or knowledge diffusion) change.

That size per se is not necessarily beneficial as regards prosperity is confirmed if one takes a more precise look at empirical regularities. Less aggregated data, for example, suggests that out of the twenty countries with the largest population only four (USA, Russia, Japan, and Germany) belong to the group of high-income economies according to the World Bank's classification.[48] Contrariwise, the mentioned high-income group includes countries with a relatively small population like Singapore, Luxembourg or Iceland. A similar result holds for selected US states. Out of the top ten according to population, only New York is also in the top ten with respect to per capita income. On the other hand, small states like New Hampshire are relatively rich.[49] Prosperous regions at different sizes have at various times included, for example, agglomerations in northwestern Europe for the continental scale, the Ruhr district in Germany for the country scale, or the city state of Hamburg in northern Germany on the local scale. Also at the city level, more and less prosperous districts might be observed[50] and even at the extreme micro level of one city block in New York City this phenomenon is observable (see Easterly et al. (2015)). From a dynamic perspective (i.e. focussing on the growth rate and not on the level of GDP per capita) OECD data for the second half of the 20th century at the country level also suggests that there is no conclusive evidence of a unique and positive relationship

[47] Just to mention one example: The lack of such an insurance in the Eurozone during the recent economic crisis has inhibited the recovery dramatically. Adversely affected US states, for instance, received federal transfers from their 'stronger' counterparts, whereas such fiscal transfers (e.g. from Germany to Spain) were not possible within the monetary union of the Eurozone as it is not a fiscal union.

[48] An overview of the classification scheme can be found here: `http://data.worldbank.org/about/country-classifications/country-and-lending-groups` (accessed: 11 August, 2015).

[49] See the 'State and County QuickFacts' data set available under: `http://quickfacts.census.gov/qfd/download_data.html` (accessed: 11 August, 2015).

[50] Compare e.g. for Karlsruhe the two districts of almost identical size 'Südweststadt' and 'Oststadt' with the former being relatively prosperous and the latter being inhabited by poorer individuals.

between GDP growth and measures of scale (compare e.g. Jones (1995a,b) or Backus et al. (1992)).[51]

The discussed ambiguity of the empirical studies is also mirrored within growth theory, which considers both the level but especially the growth rate of GDP per capita. An important class of growth models exhibits scale effects in the sense that variations in the size or scale of the economy permanently alter the long-run equilibrium growth rate per capita.[52] Within these models the relationship between scale and growth is unique – thereby contradicting the previously discussed empirical ambiguity – and depends upon both the existence and the nature of production externalities (e.g. via the provision of productive public inputs or various sorts of spillovers). This class of models, however, has strong formal requirements leading to restrictive knife-edge assumptions that have to be fulfilled in order to allow for long-run equilibrium growth.[53] Both the theoretical and empirical limitations of the mentioned models have led to the formulation of so-called non-scale growth models, i.e. models that exhibit equilibrium growth rates that are not subject to scale effects in spite of endogenously accumulated production factors and the existence of externalities. However, those models suffer from other limitations, e.g. they exhibit special stability characteristics together with transitional dynamics and hence are again only suited to explain details of the complex growth story, but not to resolve the aforediscussed tensions.[54] One might summarize that there is no clear-cut evidence on the relationship between size and prosperity neither as regards theory nor the empirical analyses. Thus it is also not clear how integration that alters the size of a considered economy affects wealth. This is a strong indicator for the fact that not all relevant aspects have been addressed yet to understand the relationship between agglomeration, growth, and integration and how their respective interdependencies and feedbacks shape the economic landscape.

One starting point for further thoughts consists in following the line of reasoning of Alesina and Spolaore (2003, 82), who argue that "whether country size matters for economic prosperity depends on a country's degree of integration with the rest of the world." This is a quite plausible argument, since, for instance, small countries like Luxembourg or Switzerland (0.5m and 7.5m inhabitants, respectively) dispose of intensive trade relations

[51] Over an extraordinary long time frame though, beginning in 1,000,000 B.C., Kremer (1993) finds support for the hypothesis of a positive link between population growth and economic growth. Notice that this long-run time scale includes the transition of various organizational forms of economic activity from hunter-gatherer over subsistence to nowadays industrialized economies.

[52] Examples are the models of Romer (1990) or Barro (1990); compare e.g. Turnovsky (2000, Chapter 14) for a comprehensive overview.

[53] These assumptions include constant returns to scale to rivalrous factors to allow for competitive factor markets. See Solow (1994) and Dalgaard and Kreiner (2003) for more on this.

[54] See Turnovsky (2000), Eicher and Turnovsky (1999) and Eicher and Turnovsky (2000) for more details on scale and non-scale models of growth.

and are well embedded in international value-creation networks. At the same time, landlocked countries of larger size as e.g. Uzbekistan (27.5m inhabitants), but with little access to world markets, are less prosperous.[55]

Though the argument of Alesina and Spolaore again starts at a national level, it is nearby to relax this perspective and to analyze the role of integration for prosperity and agglomeration also at a less aggregated level. It is conceivable that size may not matter for productivity at an aggregate level, but be quite relevant at a regional scale thereby relying on externalities associated with proximity-productivity linkages. It furthermore clarifies that the notion of scale (in the sense of size) has to be distinguished from scale effects, and one has to be precise by using the appropriate wording.

Scale may thus not simply be used as a synonym for population size. From the perspective of the production conditions, scale effects are especially linked to market forms[56] or various sorts of externalities as determinants of ongoing growth.[57] From a firm perspective, usually internal and external economies of scale are distinguished. The optimal firm size is reached whenever firms fully exploit existing internal economies of scale. In addition, there exist external economies of scale, and both scale effects in conjunction determine whether or not a firm decides to relocate. In this context, in particular firms that dispose of production conditions characterized by increasing returns to scale and the associated spillovers are important drivers for regional development. Nailing these thoughts down to a spatial component, it seems quite plausible to assume that e.g. knowledge flows are easier realized in places where more people are concentrated. Economies of scale are then rather linked to the idea of density than to mere size. It is thus important through which channel integration acts: Does it enhance the mere number of actors within the economy (scale) or does it affect their way of interaction (scale effects)?

In what follows it is argued that integration can and should be understood in a much broader sense than just as a reduction of trade costs and also cover, as argued within the model presented in Section 1.3, the impact of knowledge spillovers. Integration affects the environment of the firms in the sense that it has the power to transform size into density the latter being quite well an agglomeration force. In doing so, integration is especially apt to activate latent agglomeration and dispersion forces thereby shaping the economic landscape at various levels of aggregation. Besides, integration may be deployed more effectively as an agglomeration force if it can build upon a solid institutional base.

[55] In 2010 GDP per capita in Switzerland was more than four times the respective value for Uzbekistan (Bolt and van Zanden, 2013). The population numbers are also for 2010 and are taken from United Nations (2012, Table A.5).

[56] An example is the necessity of constant returns to scale of the private factor inputs so that competitive factor remunerations are guaranteed.

[57] A detailed overview on different scale effects is provided by the World Bank (2009, 128) in Table 4.1 entitled ‘A dozen economies of scale’.

So far, the role of institutions has been discussed as a fundamental cause of economic growth both within and across economies. Our approach exactly starts at this point and argues that the design of integration pins down the impact of institutions to a spatial dimension and may thereby alter the effective economic scale. This might be justified as follows: Integration affects both the size of the relevant market and the effectiveness of local increasing returns to scale. It thereby has the power to act as a major agglomeration force and thus has an impact on the spatial structure of economic activity at different levels of aggregation. Being more precise, integration may alter various returns to scale relevant for an individual firm. It acts via two separate channels. On the one hand, as it enhances the mobility of goods, people, capital, and ideas, integration, in the sense of higher freeness of trade also increases the size of the relevant market. On the other hand, it also affects the firms' environmental conditions, especially with respect to density as integration also impacts on knowledge spillovers. Formerly latent economies of scale might become active at a firm level, if, as a consequence of integration, a region becomes more dense. In sum, integration is not just an additional argument accompanying the agglomeration-growth nexus, but across time it is *the* key driver of spatial concentration in growing economies. It especially becomes powerful not only as an enabler but also as a magnifier of scale effects.

Hence, even though the empirical regularities discussed in Section 1.2 and the presented theory in Section 1.3 are strong indicators for a co-evolution of urbanization and growth, spatial scaling (in the sense of zooming in from continental over national and regional to even city level) together with a differentiated analysis of the impact of integration both requires and allows for a more sophisticated view.

One caveat concerns the missing linkage between scale and institutions and how these are related to space; another concerns the dynamics and endogeneity of institutions. There is a large discussion on the role of national and regional innovation systems that deals with institutional and organizational dimensions that already address the evolution of institutions within dynamic economies. The analysis, however, is carried out mostly within isolated economies (compare e.g. Cooke et al. (1997)). Analogous reasonings as regards the institutional embedding of integration and its feedback on diverse levels of aggregation are still far from being understood in depth. What most theories do so far is an analysis of how static and exogenously given levels of integration affect the dynamic concepts of agglomeration and growth. However, the corresponding feedback link from agglomeration and growth on the evolution of integration and its interpretation as an institution that is shaped by those economies that are linked via integration is missing. But these considerations are mandatory to understand integration as a dynamic and endogenous concept that is linked to space. The following section even goes one step further and argues that if integration is interpreted as a dynamic spatial institution, the concept and its design become a fundamental cause not only for growth but also for agglomeration on which economic policy has a large impact. The corresponding policy implications therefore have to accommodate these various interdependencies.

1.5 Policy Implications

The previous sections have presented a multitude of empirical facts on the issues of growth, urbanization, and globalization and explored the connections between these also on a theoretical level before the matters of size and scale and the role integration plays herein have been discussed more deeply. Implications for policy have so far been delegated to the background.

Which policy implications can be derived from the Baldwin and Forslid (2000) model? Naturally, these depend on the particular objectives set by policy makers and how they relate these objectives to overall welfare. For instance, a preference for a symmetric outcome[58] can be achieved by adjusting the policy variables for the costs of trading goods and information adequately. However, as long as one neglects arguments of redistribution, such an objective contrasts with the goal of maximizing overall real income. The technological externalities in the innovation sector imply that agglomeration of economic activity is growth enhancing in the sense that more varieties are produced if all manufacturing activity is located in a single region (unless $\lambda = 1$). Workers in every region and sector gain from this increased number of varieties via the love-of-variety effect. This dynamic gain needs to be seen alongside the static welfare loss for consumers in the periphery's traditional sector, which arises since these workers have to import all manufactured varieties, leading to a higher price index in the peripheral region (for $\phi < 1$). Baldwin and Forslid (2000) demonstrate that the dynamic gains can only mitigate, but not compensate the welfare losses for the workers in the periphery's traditional sector. Hence, starting from a symmetric equilibrium no clear-cut policy implications exist, which lead to a Pareto improvement. The presence of the externality moreover implies that, as in standard expanding-variety growth models, the decentralized growth rate in the Baldwin and Forslid (2000) model is not Pareto optimal. Combining completely free trade (i.e. $\phi = 1$) with, for instance, a subsidy to production in the innovation sector would be an example for a policy that leads to the socially optimal growth rate and, assuming the starting point is the symmetric equilibrium, brings about a Pareto improvement for all workers (Baldwin and Forslid, 2000, 323). As described above, building upon an institutional foundation, policies for closer integration become more important in the process of

[58] Perhaps inspired by Article 72 (2) of the Basic Law for the Federal Republic of Germany, which tasks the Federation with the "establishment of equivalent living conditions throughout the federal territory" (see `http://www.gesetze-im-internet.de/englisch_gg/` (accessed: 11 August, 2015).

development. Along the lines of the model in Section 1.3, these fall into two categories: Integrating policies that lead to lower costs of trading goods and integrating policies that lower the costs of trading information. That tighter integration brings immense benefits is highlighted by e.g. Eaton and Kortum (1996, Table 5), who note that for all OECD countries except the US the majority of the contribution to productivity growth comes from abroad.

Taking a step back from the model, it is evident that economic activity is spatially concentrated.[59] This need not imply differences in living standards between regions in which economic activity is concentrated and regions in which it is not, even though it can – the relationship resembles an inverted U over the course of economic development (World Bank, 2009, 74) and can be interpreted as a variant of a 'Kuznets curve'.[60] In reaction to this situation many policy makers pursue the goal of convergence of living standards as measured by GDP per capita. The European Union, for instance, tries to promote this goal by allocating a significant share of its funds for cohesion policy for the period 2014-2020 to economically lagging regions.[61] It is important to keep in mind here that pursuing economic convergence of regions does not imply striving for an equal spatial distribution of economic activity, as this would prevent regions from benefiting from agglomeration economies with possible negative repercussions on aggregate economic performance.[62]

According to the World Bank (2009, e.g. 41), economic integration is expected to resolve the tension between the fact of concentration and the objective of convergence. However, the utilized concept of integration is fuzzy, and the corresponding policy implications need to be differentiated according to the scale at which the policies are applied as well as to a country's level of development thereby taking an implicit dynamic perspective, which also addresses the evolution of integration and thus represents a dynamic concept. Such an approach allows taking into consideration the relative importance of the various agglomeration economies over the course of development and furthermore means that policies that aim at integration within or between cities differ from policies intended to integrate regions or nations.

[59] For an additional illustration of the global situation see for example: `http://gecon.yale.edu/large-pixeled-contour-globe` (accessed: 11 August, 2015).

[60] The relationship considered here is not quite the same as the one for the original Kuznets curve, which considered interpersonal income inequality instead of spatial inequality, though the underlying reasoning is the same for both curves (World Bank, 2009, 293).

[61] Defined for this purpose as regions with a GDP per capita less than 75% of the EU-27 average. See the European Union's Regional Policy website for more detailed information: `http://ec.europa.eu/regional_policy/index_en.cfm` (accessed: 12 March, 2014). At the national level in Germany, the federal financial equalization system (Länderfinanzausgleich) indirectly pursues a similar objective by utilizing a large budget to redistribute money from prosperous to weaker federal states.

[62] The example of the industrial policy in the former Soviet Union has demonstrated the problems with such an approach (World Bank, 2009, 256).

Consider, for example, the policies the World Bank (2009, 229) advocates for an urbanization strategy that explicitly aims at increasing density in order to foster benefits from the proximity-productivity linkage. The report categorizes territories according to their level of urbanization. Given *incipient urbanization*, characterized by a share of the urban population of less than 25%, density is comparatively low so that policy makers' objective should be to enable the aforementioned agglomeration economies. In addition, at this stage of development there is room for individual plants and firms to more fully exploit internal returns to scale. However, the choice of policy instrument is important. Policy makers, for instance, face a risk of favoring one place or industry over another by spatially targeted interventions, whereas markets may assess the situation in a different light. This may lead to inefficient economic structures, which may persist over time. Hence, a strategy avoiding such an outcome would place emphasis on establishing adequate institutions like secure property rights in land markets and provision of basic social services for health and education without applying a spatial focus. This aspect remains important in areas with *intermediate urbanization* shares of about 50%. Firms located in these areas are embedded in various emerging networks where cooperation arises both within and between different places. Through the co-location of firms in the same or closely related sectors, these firms benefit from localization (MAR) externalities, which lead to increasing returns to scale within the considered region. The promotion of these agglomeration economies should be high on the policy makers' agenda in these regions, but not the only item on it. Investments in infrastructure are important as well in order to ease congestion and to better integrate people and places. These policies continue to be relevant for areas of *advanced urbanization* where around 75% of the population live in urban areas. Benefits in these areas arise mostly from urbanization economies dating back to Jacobs (1969) and which point to diversity and intense economic activity. 'Livability' is the watchword from the World Bank (2009, 201) for these areas, and this could be reflected in polices that try to reduce crime and pollution or provide amenities.

Policies should thus reflect that as an economy evolves, it passes from states with constant returns to scale through states characterized by specialization/localization economies to states of urbanization economies. Throughout this process of transformation, frequently internal economies of scale turn into external economies, which then simultaneously act as agglomeration forces for those being located in close proximity. This argument is especially intuitive for growing agglomerations where development prepares the ground for the gradual evolution of clusters, which on their own are specialized, but which – given sufficient size – due to their interaction finally also allow for diversification. Integration plays a major role herein.

1.6 Conclusion for Chapter 1

Agglomeration, economic growth, and integration are three main aspects that characterize recent world economic history. This chapter has highlighted the relevant empirical facts on these and summarized how they relate to each other by presenting a model by Baldwin and Forslid (2000) that combines two seminal models from the endogenous growth (Romer, 1990) and new economic geography literature (Krugman, 1991). A non-formal analysis of the combined model has shown that lowering the costs of information tends to weaken the agglomeration force that is set in motion by a lowering of trade costs. Despite these important insights, the treatment of integration in this model is still on a very rough level.

The issue of scale in its various guises and how it is affected via integration has received particular attention. It has been demonstrated that the presence of agglomeration economies possibly depends on the level of aggregation (continent, nation, region, city, district). Analyzing this issue in more depth and investigating the presence of possible threshold effects seems to be worthwhile. Major weaknesses are that the degree of integration is exogenously given in most analyses and thus does not allow for either economic development or structural change or spatial scaling. Especially in dynamic economies where the economic landscape is shaped by the interactions of growth and agglomeration such a perspective of integration is much too simplistic. Integration in a comprehensive sense goes far beyond being the mere enabler and magnifier of mobility of various items. It also implies changes in the organization of economic processes at a spatial scale as it changes the way in which different agents in a system relate to each other. The design of integration thus complements the respective institutional settings of interacting territories. Due to the dynamic environments, also integration itself evolves and becomes a dynamic and endogenous concept. Future research thus requires a far more precise view on the endogenous determination and evolution of integration and how this interacts within supra-national, national and regional institutional settings thereby shaping the economic landscape. Integration should be clearer understood as a process that itself undergoes dynamic changes, too. As the economy evolves, also the design of the prevailing integration mechanism has to be continuously adjusted. In doing so, the respective level of agglomeration at which integration becomes active has to be considered as well as the dynamic characteristics of the investigated economic spaces.

Chapter 2: Knowledge Spillovers: On the Impact of Genetic Distance and Data Revisions[63]

2.1 Introduction to Chapter 2

Countries do not develop in isolation from each other, but are connected and interact in many different ways. A key aspect of this interdependence concerns technology, in particular technological knowledge spillovers. Accounting for this technological interdependence both on an empirical and theoretical level requires a notion of how to model the interaction between countries. Empirical evidence suggests that knowledge spillovers decline with the distance in geographic terms between countries (Keller, 2002, 136). This insight has, for instance, been picked up by Ertur and Koch (2007), who develop a theoretical model of economic growth that incorporates technological knowledge spillovers between countries. In the empirical part of their paper, they employ a specification, which qualitatively replicates the effect identified by Keller (2002).

However, geographic distance is only one possible measure to model interaction between countries. The concept is more general and encompasses "any kind of network structure" (Ertur and Koch, 2011, 236). For example, data on genetic distance, which is defined as the time, since two populations have shared a common ancestor (Spolaore and Wacziarg, 2009, 470), can be used to build this structure.

The general possibility of implementing this concept in this way is noted briefly by Ertur and Koch (2011, 236-237, 249), and it follows as Spolaore and Wacziarg (2009) demonstrate that genetic distance has an effect on cross-country income differences. They propose the following mechanism for this result and also provide empirical evidence consistent with it:[64] Within populations, characteristics like habits, implicit beliefs or conventions

[63]Chapter 2 has been published in similar form as Working Paper No. 74 in the Working Paper Series in Economics at KIT (Deeken, 2015a). An abridged version of this chapter is also available on the website of the XVIII Applied Economics Meeting in Alicante (`http://www.alde.es/encuentros/trabajos/d/pdf/156.pdf` (accessed: 19 August, 2015)).

[64]It needs to be pointed out though that their empirical results are not uncontroversial, and have, for instance, been challenged by Campbell and Pyun (2015).

are transmitted across generations biologically and culturally, and genetic distance can be viewed as a summary statistic that picks up a divergence across populations in characteristics that are slowly changing over time (see, also Spolaore and Wacziarg (2015)). The next step in their argument is the assumption that these differences in characteristics between populations introduce barriers to communication and understanding which then hinder the diffusion of technology. Hence, by using genetic distance, this chapter contributes to the literature by providing an important robustness check for the empirical results in the influential model by Ertur and Koch (2007) which relies on geographic distance to model interaction.

A further motivation for employing data on genetic distance is that this approach captures interactions between economies that geographic distance is missing. For instance, Lindner and Strulik (2014, 18) note (without any reference to genetic distance) that it might be the case that knowledge exchange between the United States and the United Kingdom is higher than between the United States and Guatemala even though geographic distance would suggest otherwise. By modeling interaction through genetic distance instead of geographic distance however, stronger knowledge spillovers between the United States and the United Kingdom compared to between the United States and Guatemala would be in line with the data on genetic distance, as the United States and United Kingdom populations are genetically closer to each other than the ones in the United States and Guatemala.

The second contribution of this chapter is the assessment of the robustness of the results by Ertur and Koch (2007) to data revisions. In their econometric analysis, they rely on data from Penn World Table (PWT) Version 6.1 (Heston et al., 2002). Since the publication of their article, newer versions of the PWT have become available, and in each update the data has been revised. Ideally, empirical results should be robust to different versions of the PWT. However, this is not a foregone conclusion, and Ponomareva and Katayama (2010) find that conclusions from cross-country growth studies might change even for the same period and units of observation, depending on the version of the PWT. More recently, Johnson et al. (2013) have also investigated this issue. They find that some data revisions have been relatively minor. For instance, the average growth rate of GDP over the period 1975-1999 for Morocco was 1.6% when calculating it using PWT 6.1 and 1.7% when basing the calculations on PWT 6.2 (Johnson et al., 2013, Table 1). Other revisions were drastic, showing high variability in the estimates, as exemplified by the case of Equatorial Guinea. Taking the data from PWT 6.1, its average GDP growth rate in the period 1975-1999 was -2.7%, making it the worst performing of 40 African countries that are covered in both PWT 6.1 and 6.2. On the other hand, for the data from PWT 6.2 its average GDP growth rate over the same period was 4%, thereby becoming the second-best performer in the list of 40 African countries after Botswana (Johnson et al., 2013, 255-256). Hence, the fact that robustness to different versions is an issue for some studies is not too surprising. However, they also argue, based on the results of a series of replication exercises for prominent articles investigating economic growth that results from cross-sectional estimations tend to be robust to changing the version of the PWT (Johnson et al., 2013, 273). This chapter investigates whether this is also the case for the results in Ertur and Koch (2007) by estimating the model for the same set of countries and the same time period (1960-1995), but with data taken from PWT Versions 6.2 and 7.1. The importance of checking the robustness of a study's results to data revisions has also been highlighted, for example, in the debate on the relationship between public debt levels and economic growth (see Reinhart and Rogoff (2010) and Herndon et al. (2014)). In this regard, the implications for providing policy advice based on results that are, for instance, sensitive to the specific version of the data set that is used, cannot be neglected. Policy makers' awareness of this issue needs to be raised.

The third contribution of this chapter lies in the quantification of the strength of the indirect (spillover) effects from, for instance, physical capital investment on steady-state per capita income in the model by Ertur and Koch (2007). In the original study, only the magnitude of the direct effects is presented. New methods have been developed by LeSage and Pace (2009) that are applied here which allows for providing important results concerning knowledge spillovers which are not highlighted in Ertur and Koch (2007).

This chapter is organized as follows: Section 2.2 introduces the concept of genetic distance. The following section briefly motivates the need to incorporate knowledge spillovers in theoretical models, introduces the concept of spatial dependence and provides indicative evidence for its existence before presenting the model by Ertur and Koch (2007) in detail. In Section 2.4, the empirical specification and estimation strategy are discussed, and Section 2.5 presents and discusses the estimation results. Section 2.6 concludes this chapter.

2.2 Genetic Distance

Genetic data is increasingly used in economic studies.[65] Nonetheless, a brief summary of relevant concepts might be helpful in order to better understand the measure of genetic distance employed in the empirical part of this chapter. A gene, i.e. a string of DNA encoding a protein, can exist in numerous forms, and a particular form of this gene is called an allele (Giuliano et al., 2014, 182). Individuals with different alleles may have different observable (phenotypic) traits, for instance, eye color; although different

[65]See, for instance, Spolaore and Wacziarg (2009), Giuliano et al. (2014), Desmet et al. (2011) or Ashraf and Galor (2013).

alleles between individuals need not result in different observable characteristics (ibid.). It is important to note that the frequency of alleles is not constant across populations, as this is the information used to calculate measures of the genetic distance between populations (Spolaore and Wacziarg, 2009, 480). In principle, on which particular genes' allelic frequency[66] this computation is based would not matter. In practice, however, the measure is based on neutral genes. These are genes that do not endow an individual with a selective advantage (Giuliano et al., 2014, 182). This implies that the measure of genetic distance provides no information about specific genes that have a direct impact on fitness and survival or income and productivity (Spolaore and Wacziarg, 2009, 470).

The particular index of genetic distance mainly considered in this chapter, F_{ST} distance, measures the probability that the alleles for a gene selected at random from two populations will be different (Spolaore and Wacziarg, 2009, 481).[67] For identical allele distributions this index equals zero, and it increases with differences in the distributions.[68] As Spolaore and Wacziarg (2009) argue, these allele differences increase due to the presence of random (or genetic) drift. This concept may be illustrated through an example by Masel (2011, R837): Imagine a population of 5,000 people in which, due to the general diploid nature of human somatic cells (the gametes, ovum and sperm, in contrast, are haploid), 10,000 copies of each gene exist.[69] If now, for instance, 3,000 of those copies are of a particular form or allele, then in the next generation there might be more or fewer than 3,000 copies, as out of all possible gametes, only some are randomly picked out. When populations become separated, and for constant drift rates (see Kimura (1968) for evidence on this), genetic distance can then be used to measure the time that has passed, since populations have become separated (or, in other words, their degree of genealogical relatedness). It is in this sense that genetic distance can be understood as the time that has elapsed, since populations have shared a common ancestor. Spolaore and Wacziarg (2009, 470-471) furthermore hypothesize that populations that are genetically more distant, have diverged more strongly in characteristics that are variably transmitted across generations, like habits, norms, or implicit beliefs, and that this divergence hinders, for instance, communication and understanding and thereby creates barriers to the diffusion of development or technology. Applying this line of thought to the example mentioned in

[66] A database on allele frequencies is available under: `http://alfred.med.yale.edu` (accessed: 11 August, 2015).

[67] Data from Spolaore and Wacziarg (2009) on an index with different theoretical properties, Nei's distance (see Nei (1972) and Cavalli-Sforza et al. (1994)), which however is highly correlated with F_{ST} distance, will be used to assess the robustness of the empirical results as well.

[68] This index from Cavalli-Sforza et al. (1994) uses the frequency of 128 alleles that are related to 45 genes, which fulfill the conditions that they are both selectively neutral and easy to collect (Giuliano et al., 2014, 183).

[69] A human cell is called haploid if its nucleus has a single set of 23 chromosomes and it is diploid if its nucleus has a double set of 23 chromosomes.

the introduction: The United States are genetically closer to the United Kingdom than to Guatemala (the pairwise genetic distances are 0.033 and 0.091, respectively) so that with regard to this concept fewer barriers to knowledge diffusion should exist between the United States and the United Kingdom than between the United States and Guatemala.[70] Note that the stated genetic distances in this example are weighted F_{ST} genetic distances, which take into account that some countries, like the United States or Australia, consist of genetically distant subpopulations (see Spolaore and Wacziarg, 2009, 484-485).[71]

2.3 Spatial Dependence and Model Setup

This section motivates the need for including technological interdependence across geographical regions in theoretical models and introduces the spatial Solow model as it was developed by Ertur and Koch (2007). Section 2.3.1 develops the concept of spatial dependence and illustrates the concept with a brief example. Section 2.3.2 describes in detail the specification of technological progress and technological interdependence in this model before Section 2.3.3 investigates the transition dynamics and derives an equation for the steady-state income per worker.

2.3.1 Spatial Dependence

Knowledge spillovers have been discussed by economists for quite a long time, going back to Marshall (1890). His description of these effects was completely verbal however, and the first attempts to incorporate these effects within a theoretical model are due to Arrow (1962) and Romer (1986). These authors made the assumption that knowledge generated in a single firm is not confined to this particular firm, but might spill over to other firms in a given geographical region as knowledge is considered a non-rival input. While this is an improvement on earlier models like the one by Solow (1956), it remains unclear why knowledge diffusion should stop at a given border. Learning-by-doing, for instance, can result as a by-product of mergers and acquisitions, be a result of interfirm cooperation or the meeting of different people at conferences and seminars (Fischer, 2011, 420). None of these activities is necessarily confined within an arbitrary geographical unit. With respect to physical capital externalities, for example, López-Bazo et al. (2004, 44), note that "there is no a priori reason to constrain spillovers within the barriers of the economy where the agent making the investment is located". Diffusion of these knowledge spillovers across boundaries can then be viewed as a spatial externality, implying that, for instance,

[70]Considering geographic distances between the country capitals suggests that Washington, D.C. is closer to Guatemala City (distance = 3,007km) than to London (distance = 5,909km). See Equation (B.5) in Appendix B.2 for the general formula to calculate these distances.

[71]See Appendix B.1 for a formal definition of F_{ST} genetic distance. The formula for the weighted version is provided in Equation (B.3).

the economic development of neighboring countries is related. Before presenting how this effect is picked up in a theoretical model, the concept of spatial dependence will be introduced to provide indicative evidence for the relevance of these spatial externalities.

As countries interact with each other in numerous ways, it is straightforward to assume that the development of one country may be influenced by the development of nearby countries. This latter idea is captured in the (spatial) econometric literature by the concept of spatial dependence. More precisely, spatial dependence captures situations in which the values observed in e.g. country i depend on the values observed in neighboring countries (see LeSage and Pace, 2009, 2).

As an example consider Figure 2.1, which depicts total factor productivity (TFP) levels relative to those of the United States for 110 countries for the year 2007 in shades of green. Visual inspection of this figure suggests that the TFP levels are not distributed

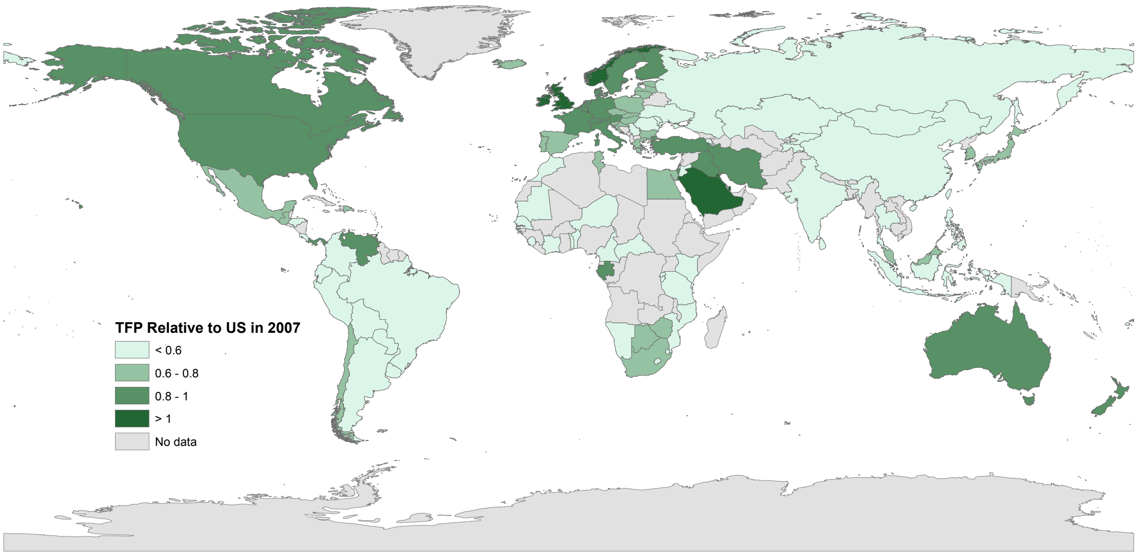


Figure 2.1: Total Factor Productivity Relative to the United States for 110 Countries in 2007 (Data from Feenstra et al., 2013b).

randomly. Countries with comparatively low TFP levels (less than 80% of the value for the United States) are, for instance, concentrated in South America, southeastern Europe or east and southeast Asia, whereas regions with higher TFP levels (above 80% of the level in the United States) can be found in northwestern Europe.

An alternative visualization of these data is provided by a Moran scatterplot in Figure 2.2. When interpreting this figure, it is important to note that the variables are in deviations-from-the-mean form. The meaning of the variable "Spatial Lag of TFP" on the ordinate might not be immediately clear. In general, a spatial lag is a weighted average of the

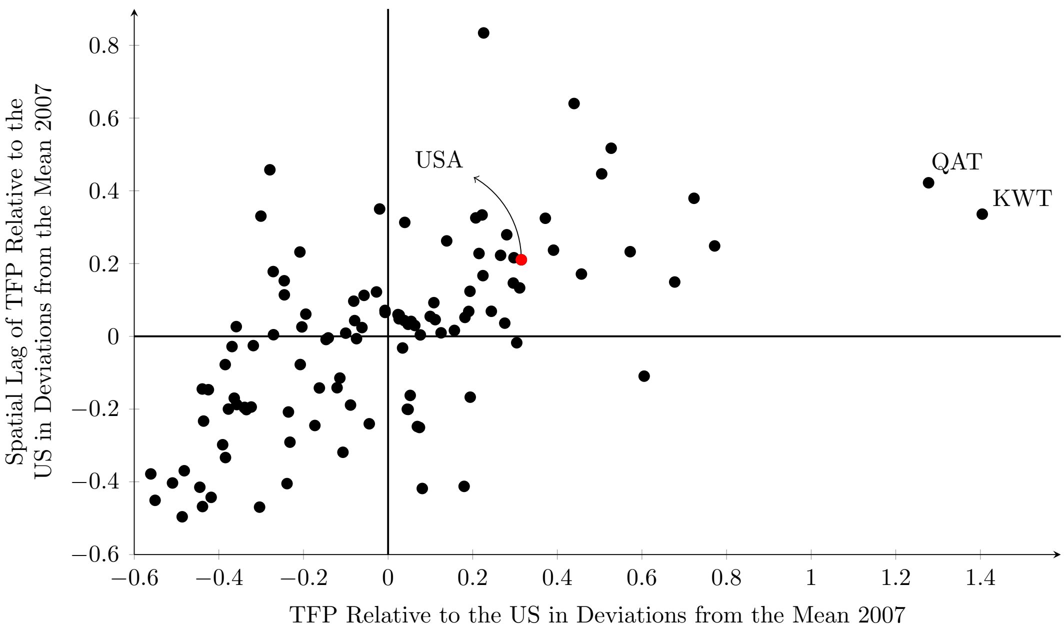


Figure 2.2: Moran Scatterplot of TFP Relative to the United States for 110 Countries in 2007 (Data from Feenstra et al., 2013b).

Note: The variables are in the form of deviations from the mean so that the value 0 on the abscissa is equivalent to the mean value of 0.685.

values for a variable from countries that are neighbors to country i (see LeSage and Pace, 2009, 8).[72]

The Spatial Lag of TFP on the ordinate in Figure 2.2 thus has the interpretation that for a given observation i, this variable shows the deviation of the TFP for country i from the mean of TFP of its neighbors. Hence, in the lower left quadrant of the figure, one finds countries for which not only their own TFP is below the mean, but also the TFP of its neighbors is below the mean. Whereas, in the upper right quadrant countries cluster whose own TFP as well as the one of its neighbors is above mean.[73]

Figures 2.1 and 2.2 have provided indicative evidence of spatial dependence (or spatial autocorrelation) in country-level data. Theoretical models should therefore not disregard

[72]For expository reasons the term "neighbor" will be slightly abused in this section. In fact, in the calculation of the spatial lag of TFP for country i, all countries for which data is available are included and not only neighboring countries. However, countries that are geographically closer to country i receive a higher weight in the calculation of the spatial lag. The precise formal specification of this idea is provided in Section 2.5.1. Note that countries whose TFP levels exceed those of the US are mainly oil-rich countries, like Saudi Arabia, Qatar (QAT) or Kuwait (KWT) for which TFP will be overstated as data is lacking to include also "subsoil assets" in the underlying methodology (see Feenstra et al. (2013a, 35-36) and Inklaar and Timmer (2013)) as well as Singapore or the Special Administrative Regions of Hong Kong and Macao.

[73]Note that this result of spatial dependence is not particular to country-level data. Looking, for example, at the distribution of the TFP levels of European NUTS 2 regions gives a similar result (Derbyshire et al., 2011).

this characteristic of the data, but instead try to represent it. The following section shows one possible way to achieve this.

2.3.2 Specification of Technological Progress

The aggregate production for each country $i = 1, \dots, N$ at time t in the model developed by Ertur and Koch (2007) is described by the Cobb-Douglas production function

$$Y_i(t) = A_i(t) K_i(t)^{\alpha} L_i(t)^{1-\alpha} \qquad \text{with} \qquad 0 < \alpha < 1 \tag{2.1}$$

where output, $Y_i(t)$, is produced with the three input factors labor, $L_i(t)$, physical capital, $K_i(t)$ and technology, $A_i(t)$. This function is linearly homogenous in the two input factors capital and labor and thus has constant returns to scale with respect to these two factors. The aggregate level of technology in country i is described by

$$A_i(t) = \Omega(t) k_i(t)^{\phi} \prod_{j \neq i}^{N} A_j(t)^{\gamma w_{ij}}. \tag{2.2}$$

Basically, overall technological progress is assumed to be due to three different factors in Equation (2.2) which are (imperfect) substitutes. The first factor, $\Omega(t)$, reflects exogenous (Harrod-neutral) technological progress as modeled in the original contributions by Solow (1956, 85) and Swan (1956). In formal terms, this is captured by the equation

$$\Omega(t) = \Omega(0) e^{\mu t},$$

with μ as the constant rate of technological progress and $\Omega(0)$ the initial level.

The second term models the influence of the physical capital per worker, $k_i(t) = \frac{K_i(t)}{L_i(t)}$, on aggregate technology in country i. The level of technology increases with the level of capital per worker $k_i(t)$, modeling the assumption that physical capital externalities exist in general. Their strength is governed by the parameter ϕ for which $0 \leq \phi < 1$ holds so that perfect knowledge spillovers from a capital investment in a given firm in country i to the remaining firms in this country are ruled out, as diffusion is not frictionless and some knowledge is "lost in transmission". The assumption that all firms in a country gain a higher level of technology, if one firm increases its physical capital per worker is due to Arrow (1962) and Romer (1986). As has been mentioned above, the assumption that these knowledge spillovers should be constrained within a single region or country is tenuous. Why should knowledge diffuse only within a country but not across countries?

The strength of the spillovers might be dampened (and there is indeed empirical evidence to that extent[74]), but they should be present nonetheless.

The third factor in Equation (2.2) picks this up. From a formal perspective, this factor is a weighted geometric mean of the level of technology in all countries $j = 1, \ldots, N$ connected to country i. The strength of these cross-border spillovers or spatial externalities is governed by two factors. The parameter γ, for which $0 \leq \gamma < 1$ holds, gauges which fraction of knowledge generated in, for example, country j' spills over into country i. This value is the same for all units of observation. The second factor concerns the weights w_{ij}. In general, these are allowed to differ across countries, and they specify the way in which countries are connected to each other. It is important to note that how strong country i benefits from knowledge spillovers depends on the way it is connected to all other countries under consideration. This implies that the net effect on a country's level of technology due to spatial spillovers will differ across countries. For a given degree of spillovers, relatively isolated countries will benefit less than more integrated countries. With respect to the spatial weights, it is assumed that these are non-negative, which leaves open the possibility that countries might not be connected to each other at all so that spatial externalities are absent between particular pairs of countries, non-stochastic, implying that the weights are fixed over time, and finite. In addition, the weights w_{ij} lie in the interval $[0, 1]$ and for $i = j$ $w_{ij} = 0$ holds, excluding the case of self-influence. Finally, the weights sum to one.[75] Summarized, the spatial weight matrix or more generally interaction matrix, $\mathbf{W}$, is thus row-stochastic (LeSage and Pace, 2009, 9-10).[76]

Applying the natural logarithm to Equation (2.2), it can be rewritten as

$$\ln A_i(t) = \ln \Omega(t) + \phi \ln k_i(t) + \gamma \sum_{j \neq i}^{N} w_{ij} \ln A_j(t). \tag{2.3}$$

[74] See, for example, Keller (2002), who estimates that at a distance of about 1,200 kilometers from the country in which the knowledge originates, 50% is still available.

[75] On the assumptions for the spatial weights see Ertur and Koch (2007, 1036, Footnote 2) and Fischer and Wang (2011, 20).

[76] An illustration of two spatial weight matrices is given in Appendix B.2, which also describes the calculation of the spatial weights based on great circle distances between country capitals in detail.

Stacking the equations for all countries $i = 1, \ldots, N$ at time t, the level of technology can be expressed as

$$\underbrace{\begin{pmatrix} \ln A_1(t) \\ \vdots \\ \ln A_N(t) \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}} = \underbrace{\begin{pmatrix} \ln \Omega(t) \\ \vdots \\ \ln \Omega(t) \end{pmatrix}}_{\substack{=\boldsymbol{\Omega} \\ (N\times 1)}} + \phi \underbrace{\begin{pmatrix} \ln k_1(t) \\ \vdots \\ \ln k_N(t) \end{pmatrix}}_{\substack{=\boldsymbol{k} \\ (N\times 1)}} + \gamma \underbrace{\begin{pmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{pmatrix}}_{\substack{=\boldsymbol{W} \\ (N\times N)}} \cdot \underbrace{\begin{pmatrix} \ln A_1(t) \\ \vdots \\ \ln A_N(t) \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}} \tag{2.4}$$

$$\iff \qquad \boldsymbol{A} = \boldsymbol{\Omega} + \phi\boldsymbol{k} + \gamma\boldsymbol{W}\boldsymbol{A}.$$

Given that spatial dependence is positive, $\gamma \neq 0$, and that the inverse $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$ exists[77] the previous equation is equivalent to

$$\boldsymbol{A} = (\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}\boldsymbol{\Omega} + \phi(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}\boldsymbol{k}. \tag{2.5}$$

From this expression, it follows that the level of technology for a given country i can be written as

$$A_i(t) = \Omega(t)^{\frac{1}{1-\gamma}} k_i(t)^{\phi} \prod_{j=1}^{N} k_j(t)^{\phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W}^r)_{ij}} \tag{2.6}$$

where $(\boldsymbol{W}^r)_{ij}$ are the individual entries in row i and column j of the matrix $\boldsymbol{W}$ taken to the power of r. Since the derivation of Equation (2.6) is not immediately obvious, important intermediate results are provided in Appendix B.4.1. In particular, it is proved that the inverse matrix $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$, which is also called the inverse spatial transformation (Le Gallo, 2014, 1515), can be written as an infinite series, i.e.

$$(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1} = \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r.$$

With respect to Equation (2.5), it follows then that the level of technology in every country is correlated with the level of technology in every other country and closer countries are more closely related.[78] The effect of the inverse spatial transformation is often referred to as the spatial multiplier effect (see, for example, Ertur and Koch (2007, 1044) or Le Gallo (2014, 1515)).

[77] This inverse exists if $\frac{1}{\gamma}$ is not an eigenvalue of the spatial weight matrix. However, it is not necessarily guaranteed that the inverse exists for the parameter space for γ assumed here. See Appendix B.3 for a proof that the inverse exists in this case as well.

[78] As Anselin (2003, 155) mentions in a slightly different context, which is nonetheless applicable here, this is in effect a reformulation of the first law of geography by Tobler, which states that "everything is related to everything else, but near things are more related than distant things" (1970, 236).

The results derived with respect to the level of technology are helpful in rewriting the production function. This function exhibits constant returns to scale in capital and labor, which implies that Equation (2.1) can be written in per capita terms

$$y_i(t) = A_i(t)k_i(t)^{\alpha} \tag{2.7}$$

where $y_i(t) = \frac{Y_i(t)}{L_i(t)}$. Inserting the expression for the level of technology in Equation (2.6) into the per worker production function leads to

$$y_i(t) = \Omega(t)^{\frac{1}{1-\gamma}} \cdot k_i(t)^{\alpha+\phi\left(1+\sum_{r=1}^{\infty}\gamma^r(\boldsymbol{W^r})_{ii}\right)} \cdot \prod_{j\neq i}^{N} k_j(t)^{\phi\sum_{r=1}^{\infty}\gamma^r(\boldsymbol{W^r})_{ij}}.$$

Now define

$$u_{ii} \equiv \alpha + \phi\left(1 + \sum_{r=1}^{\infty}\gamma^r(\boldsymbol{W^r})_{ii}\right) \qquad \text{and} \qquad u_{ij} \equiv \phi\sum_{r=1}^{\infty}\gamma^r(\boldsymbol{W^r})_{ij} \tag{2.8}$$

and substitute for the exponents of physical capital per worker so that the per worker production function can be written more compactly as

$$y_i(t) = \Omega(t)^{\frac{1}{1-\gamma}} \cdot k_i(t)^{u_{ii}} \cdot \prod_{j\neq i}^{N} k_j(t)^{u_{ij}}. \tag{2.9}$$

From this function it can be seen that in contrast to the standard Solow model, the model presented here implies heterogeneity in the social elasticities of income per worker with respect to capital per worker. If, for instance, country i increases its own stock of physical capital per worker, the social return (or elasticity) is[79]

$$\frac{\partial y_i(t)}{\partial k_i(t)}\frac{k_i(t)}{y_i(t)} = u_{ii}.$$

In case all countries except country i simultaneously increase their stocks of physical capital per worker, then the corresponding elasticity is

$$\sum_{j\neq i}^{N}\frac{\partial y_i(t)}{\partial k_j(t)}\frac{k_j(t)}{y_i(t)} = \sum_{j\neq i}^{N} u_{ij}.$$

[79]The term social in contrast to private is warranted in this case, as the elasticity calculated here includes the physical capital externalities, ϕ, within a country (see the definitions in Equation (2.8)).

Hence, if all countries $i = 1, \dots, N$ together increase their stocks of physical capital per worker, then

$$\frac{\partial y_i(t)}{\partial k_i(t)} \frac{k_i(t)}{y_i(t)} + \sum_{j \neq i}^{N} \frac{\partial y_i(t)}{\partial k_j(t)} \frac{k_j(t)}{y_i(t)} = u_{ii} + \sum_{j \neq i}^{N} u_{ij} = \alpha + \frac{\phi}{1-\gamma} < 1 \tag{2.10}$$

is the social output elasticity per worker in the situation in which all countries simultaneously increase their capital stock per worker.[80] The inequality $\alpha + \frac{\phi}{1-\gamma} < 1$ is an assumption made by Ertur and Koch (2007, 1037), since otherwise the per worker production function in Equation (2.9) would not have decreasing returns to (all) physical capital, and the model would exhibit endogenous growth.

2.3.3 Transition Dynamics and Steady State

Capital accumulation is described by the fundamental dynamic equation of the Solow model, i.e.

$$\dot{k}_i(t) = s_i y_i(t) - (n_i + \delta) k_i(t) \tag{2.11}$$

where $\dot{k}_i(t) = dk_i(t)/dt$ denotes a time derivative, s_i is the country specific constant saving rate (the fraction of output invested in physical capital), n_i is the constant growth rate of labor for country i, and δ is the depreciation rate, which is assumed to be identical for all countries.

Due to the decreasing returns to capital per worker (it holds that $0 < \alpha < 1$, see Equation (2.1)), $k_i(t)$ converges monotonically to its steady-state value or value on the balanced growth path, $k_i^*(t)$.[81] When this value is reached, capital (and by implication output)

[80] The result before the inequality follows since

$$\begin{aligned} u_{ii} + \sum_{j \neq i}^{N} u_{ij} &= \alpha + \phi + \phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W^r})_{ii} + \sum_{j \neq i}^{N} \phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W^r})_{ij} \\ &= \alpha + \phi \left(1 + \sum_{j=i}^{N} \cdot \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W^r})_{ij} \right). \end{aligned}$$

The matrices $\boldsymbol{W^r}$ are Markov matrices, and in this case it is the rows that sum to one, meaning that $\sum_{j=i}^{N} \boldsymbol{W^r}_{ij} = 1 \, \forall \, r$ so that $u_{ii} + \sum_{j \neq i}^{N} u_{ij} = \alpha + \phi \left(1 + \sum_{r=1}^{\infty} \gamma^r\right)$, and the term in parentheses can be rewritten as $1 + \sum_{r=1}^{\infty} \gamma^r = 1 + \sum_{r=1}^{\infty} \gamma^r + \gamma^0 - \gamma^0 = 1 + \sum_{r=0}^{\infty} - \gamma^0 = \frac{1}{1-\gamma}$. With this result, the social returns are

$$u_{ii} + \sum_{j \neq i}^{N} u_{ij} = \alpha + \frac{\phi}{1-\gamma}.$$

[81] Similar to Fischer (2011, 425), the balanced growth path is defined as a situation in which the physical capital per worker grows at rate g, the investment rate for physical capital, the employment growth rate and the growth rate of the exogenous part of technology are constant.

per worker grow at the balanced growth rate $g = \mu\left[(1-\alpha)(1-\gamma)-\phi\right]^{-1}$.[82] This rate increases if, for instance, ϕ, the parameter indicating the strength of knowledge spillovers within a country, increases or if γ increases so that knowledge spillovers between countries are stronger.[83]

The steady-state value $k_i^*(t)$ can be calculated by noting that from Equation (2.11) on the balanced growth path

$$g = s_i \frac{y_i^*(t)}{k_i^*(t)} - (\delta + n_i) \quad \Longleftrightarrow \quad k_i^*(t) = \frac{s_i}{n_i + \delta + g} y_i^*(t) \tag{2.12}$$

holds. Inserting Equation (2.9) into the right-most expression above and solving for $k_i^*(t)$ yields

$$k_i^*(t) = \Omega(t)^{\frac{1}{1-\gamma(1-u_{ii})}} \left(\frac{s_i}{n_i + \delta + g}\right)^{\frac{1}{1-u_{ii}}} \prod_{j \neq i}^{N} \left(k_i^*(t)\right)^{\frac{u_{ij}}{1-u_{ii}}}.$$

The steady-state value of real income per capita in country i, $y_i^*(t)$, can be derived by first taking the logarithm of the production function in Equation (2.7), writing it in matrix form (compare Equation (2.4)) to obtain

$$\mathbf{y}^* = \mathbf{A}^* + \alpha \mathbf{k}^*,$$

where the asterisks denote steady-state values, and then inserting the expression for $\mathbf{A}$ from Equation (2.5) evaluated at steady state therein and finally solving for $\mathbf{y}^*$, which yields

$$\mathbf{y}^* = \boldsymbol{\Omega} + (\alpha + \phi)\mathbf{k}^* - \alpha\gamma \mathbf{W}\mathbf{k}^* + \gamma \mathbf{W}\mathbf{y}^*.$$

Writing this equation for a single country i at time t results in

$$\ln y_i^*(t) = \ln \Omega(t) + (\alpha + \phi) \ln k_i^*(t) - \alpha\gamma \sum_{j \neq i}^{N} w_{ij} \ln k_j^*(t) + \gamma \sum_{j \neq i}^{N} w_{ij} \ln y_j^*(t). \tag{2.13}$$

[82] The rate can be calculated by taking the derivative of Equation (2.13) with respect to time, then using $\ln k_i^*(t)/dt = g = \ln y_i^*(t)/dt$, and solving the derivative for g.

[83] These results hold due to the inequality $\alpha + \frac{\phi}{1-\gamma} < 1$.

Inserting now the expression for the capital-output ratio on the balanced growth path from Equation (2.12) into this expression and solving for $y_i^*(t)$ leads to the final result

$$\begin{aligned}\ln y_i^*(t) = {} & \frac{1}{1-\alpha-\phi}\ln\Omega(t) + \frac{\alpha+\phi}{1-\alpha-\phi}\ln s_i \\ & - \frac{\alpha+\phi}{1-\alpha-\phi}\ln(n_i+g+\delta) - \frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i}^{N} w_{ij}\ln s_j \\ & + \frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i}^{N} w_{ij}\ln(n_j+g+\delta) + \frac{\gamma(1-\alpha)}{1-\alpha-\phi}\sum_{j\neq i}^{N} w_{ij}\ln y_j^*(t).\end{aligned} \tag{2.14}$$

In line with the standard Solow model, this equation states that the per worker income in steady state in country i is positively influenced by an increase in its own saving rate, s_i, since an increase in savings leads to more investment and a higher capital stock per worker, which in turn leads to a higher per worker income in steady state. Increases in the labor force (note that g and δ are constant) reduce steady-state income, since for a given saving rate the capital stock must now be spread over more workers so that $k_i^*(t)$ falls, implying a decrease of $y_i^*(t)$. In addition to these standard effects, Equation (2.14) suggests that the steady-state value also depends negatively on increases in the saving rates of the other countries and positively on the increases in the population growth rate and steady-state levels of the remaining countries. Why this should be the case is not immediately obvious. However, at this point it needs to be taken into account that the steady-state values in the neighboring countries of i depend, for instance, positively on their own saving rates. Higher capital stocks in neighboring countries lead to a higher level of technology in these countries (see Equation (2.2)). A fraction of this knowledge spills over into country i, which therefore benefits via these spatial externalities. The elasticity $\eta_{s_j}^i$ of income per worker in steady state in country i with respect to the saving rate in the neighboring countries is given by[84]

$$\eta_{s_j}^i = \frac{\phi}{(1-\alpha)(1-\alpha-\phi)}\sum_{r=1}^{\infty}\left(\boldsymbol{W}^r\right)_{ij}\left[\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\right]^r. \tag{2.15}$$

This expression is clearly positive (compare Equation (2.10)). The corresponding elasticity with respect to population growth, $\eta_{n_j}^i$, equals the expression above with a negative sign. A further point to note is that the effect on a country's per capita income from increasing

[84] See Appendix B.4.2 for the derivation.

its own saving rate (or decreasing its own population growth rate) is higher in this model than in the standard Solow model. This elasticity is given by

$$\eta^i_{s_i} = \frac{\alpha + \phi}{1-\alpha-\phi} + \frac{\phi}{(1-\alpha)(1-\alpha-\phi)} \sum_{r=1}^{\infty} (\boldsymbol{W}^r)_{ii} \left[\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\right]^r . \tag{2.16}$$

The reason is that the knowledge generated by the increase in the capital stock per worker diffuses to the neighboring countries, leading to a higher income per worker in these countries, which again has a positive impact on the income per worker in country i. This feedback effect follows from the model's setup, since even though the diagonal entries of the spatial weight matrix $\boldsymbol{W}$ are zero, this is not the case for higher orders of the matrix, as, for instance, each country is a second-order neighbor to itself or in other words a neighbor to its first-order neighbor (see LeSage and Pace, 2009, 9).

2.4 Empirical Specification, Estimation Strategy, and Model Interpretation

This section presents details on the empirical specification of the model from Section 2.3, develops the spatial econometric estimation strategy, and addresses the interpretation of parameters from the estimation. It will first be shown that ordinary least squares (OLS) estimators of the model's parameters are biased and inconsistent. Thereafter, maximum likelihood estimators (ML) will be presented as an alternative to OLS.

2.4.1 Econometric Specification of the Model

Equation (2.14) from Section 2.3 has the empirical counterpart at $t = 0$ (both the time index and the star to indicate the steady-state value of per worker income are now dropped to enhance readability)

$$\begin{aligned} \ln y_i = \beta_0 + \beta_1 \ln s_i + \beta_2 \ln(n_i + g + \delta) + \theta_1 \sum_{j \neq i}^{N} w_{ij} \ln s_j \\ + \theta_2 \sum_{j \neq i}^{N} w_{ij} \ln(n_j + g + \delta) + \rho \sum_{j \neq i}^{N} w_{ij} \ln y_j + \varepsilon_i \end{aligned} \tag{2.17}$$

where $\frac{1}{1-\alpha-\phi} \ln \Omega(0) = \beta_0 + \varepsilon_i$ for $i = 1, \ldots, N$ and β_0 is a constant and ε_i is a country-specific shock. From the development of the theoretical model, the empirical specification above implies the following constraints on the coefficients $\beta_1 + \beta_2 = 0$ and $\theta_1 + \theta_2 = 0$ (see Equation (2.14)).

In matrix form, Equation (2.17) is equivalent to[85]

$$\boldsymbol{y} = \boldsymbol{\iota_N}\beta_0 + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \rho\boldsymbol{Wy} + \boldsymbol{\varepsilon}. \tag{2.18}$$

and the definitions of the respective variables are in order of appearance in the equation above provided in the list below

- $\boldsymbol{y}$ is an $N \times 1$ vector of real income per worker in logarithms,
- $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones,
- β_0 is a scalar (constant parameter),
- $\boldsymbol{X}$ is an $N \times 2$ matrix of the exogenous explanatory variables (investment rate and population growth rate) in logarithms for the N observations,
- $\boldsymbol{\beta}$ is a 2×1 vector $[\boldsymbol{\beta} = (\beta_1, \beta_2)']$ of the regression parameters for the investment rate and population growth rate,
- $\boldsymbol{W}$ is the $N \times N$ spatial weight matrix in row-standardized form,
- $\boldsymbol{WX}$ is the $N \times 2$ matrix of the spatially lagged explanatory variables,
- $\boldsymbol{\theta}$ is a 2×1 vector $[\boldsymbol{\theta} = (\theta_1, \theta_2)']$ of the regression parameters for the spatially lagged explanatory variables,
- ρ is the spatial autoregressive coefficient, $\rho = \frac{\gamma(1-\alpha)}{1-\alpha-\phi}$,
- $\boldsymbol{Wy}$ is an $N \times 1$ vector representing the spatially lagged endogenous variable,
- $\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of errors, for which the assumption of normal and identical distribution with mean zero and variance $\sigma^2\boldsymbol{I}$ holds, i.e. $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

Equation (2.18) includes spatial lags of both the endogenous variable and the explanatory variables on the right-hand side. This specification is called a Spatial Durbin Model (SDM) (see e.g. Anselin, 1988b, 111). By redefining $\boldsymbol{Z} = [\boldsymbol{\iota}_N \boldsymbol{X} \boldsymbol{WX}]$ and $\boldsymbol{\delta} = [\beta_0, \boldsymbol{\beta}, \boldsymbol{\theta}]'$, this model can be rewritten as (see, for instance, LeSage and Pace, 2009, 46)

$$\boldsymbol{y} = \rho\boldsymbol{Wy} + \boldsymbol{Z\delta} + \boldsymbol{\varepsilon} \tag{2.19}$$

[85] The notation here and in the list below follows Fischer (2011) and thus differs slightly from the one in Ertur and Koch (2007). The reason for this is to be precise and clear in the notation. In particular, by using the notation in Fischer (2011), having $\boldsymbol{X}$ denote two different matrices depending on context, is avoided.

which is a spatial autoregressive (SAR) model. This specification will be used to demonstrate that the OLS estimates are biased and to derive the ML estimates for this model.[86]

In reduced form (i.e. solved for the endogenous variable), the specification in Equation (2.19) can be expressed as[87]

$$\boldsymbol{y} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{Z}\boldsymbol{\delta} + (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}.$$

This specification implies that the spatial lag of the endogenous variable and the error term are correlated with each other, as

$$Cov[(\boldsymbol{W}\boldsymbol{y}), \boldsymbol{\varepsilon}] = E[(\boldsymbol{W}\boldsymbol{y})\boldsymbol{\varepsilon}'] - E[\boldsymbol{W}\boldsymbol{y}] = \boldsymbol{W}(\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\sigma^2$$

so that the OLS parameter estimators are biased and inconsistent (Davidson and MacKinnon, 2004) and an alternative estimation strategy is thus necessary.

2.4.2 Estimation Strategy

Given these problems, LeSage and Pace (2009, 45) note with reference to Lee (2004) that maximum likelihood is a viable alternative to OLS.[88] Assuming that the errors are normally distributed, the specification in Equation (2.19) has the following log-likelihood function.

$$\begin{aligned}\ln L(\boldsymbol{y}; \boldsymbol{\delta}, \rho, \sigma^2) = &-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I} - \rho\boldsymbol{W}| \\ &- \frac{1}{2\sigma^2}\left[(\boldsymbol{I} - \rho\boldsymbol{W})\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I} - \rho\boldsymbol{W})\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\delta}\right].\end{aligned}$$

Finding the maximum for this function requires calculating the partial derivatives with respect to all parameters, setting these necessary conditions equal to zero, and solving the system for the parameters. Instead, yielding identical results, this multivariate optimization problem can be reduced to a univariate optimization problem by concentrating the log-likelihood function with respect to the parameters $\boldsymbol{\delta}$ and σ^2 (LeSage and Pace, 2009, 47). This concentrated log-likelihood function depends, in addition to the sample data, only on the single parameter ρ and is given by

$$\ln L(\boldsymbol{y}; \rho) = -\frac{N}{2}\left[\ln(2\pi) + 1\right] + \ln|\boldsymbol{I} - \rho\boldsymbol{W}| - \frac{N}{2}\ln\left[\frac{(\hat{\boldsymbol{e}}_O - \rho\hat{\boldsymbol{e}}_L)'(\hat{\boldsymbol{e}}_O - \rho\hat{\boldsymbol{e}}_L)}{N}\right] \tag{2.20}$$

[86]The SAR model is nested in the SDM model and so with the above rewriting their likelihood functions coincide (LeSage and Pace, 2009, 46). Using the SAR model here is simply done to save on notation.

[87]Note that this only holds if $(\boldsymbol{I} - \rho\boldsymbol{W})$ is non-singular. See Appendix B.3 for the proof.

[88]Other approaches like instrumental variables (IV), generalized methods of moments (GMM) or Bayesian Markov Chain Monte Carlo (MCMC) might be alternatives (see Elhorst, 2010, 15).

where $\hat{\boldsymbol{e}}_O$ are the estimated residuals from a regression of $\boldsymbol{y}$ on $\boldsymbol{Z}$ and $\hat{\boldsymbol{e}}_L$ those from a regression of $\boldsymbol{Wy}$ on $\boldsymbol{Z}$ (see Fischer, 2011, 427). Maximizing Equation (2.20) yields a ML estimate $\hat{\rho}$, which can then be used to compute the ML estimates $\hat{\boldsymbol{\delta}}$ and $\hat{\sigma}^2$.

2.4.3 Model Interpretation

Due to the presence of the spatial lags $\boldsymbol{WX}$ and $\boldsymbol{Wy}$ in Equation (2.18), the interpretation of the parameters is a bit more complicated than in standard linear regression models, since the feedback effects mentioned in Section 2.3.3 need to be taken into account. The partial derivatives of Equation (2.18) with respect to, for example, the investment rate, are given by

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{X}_1'} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} (\boldsymbol{I}\beta_1 + \boldsymbol{W}\theta_1) . \tag{2.21}$$

This expression is an $N \times N$ matrix, which represents the non-linear impacts on all countries that result from a change in the investment rate in any country (Fischer, 2011). As LeSage and Pace (2009, 36) point out, in general the impact of a change in an explanatory variable in this type of model will not be identical across all observations. Therefore, they suggest a summary measure of these differing impacts. The row sums in the matrix in Equation (2.21) represent the total impact to an observation, i.e. the impact of a change in the investment rate in all countries on steady-state income in country $i = 1, \dots, N$. The average of these row sums is then labeled the average total impact to an observation by LeSage and Pace (2009). On the main diagonal of the matrix are the own partial derivatives or direct impacts from a change in the explanatory variable. These derivatives capture the effect of a change in, for example, the investment rate in country i on steady-state income in country i, and these impacts are summarized via averaging the entries on the diagonal of the matrix. LeSage and Pace (2009, 37) note that this corresponds, at least to a certain extent, to the typical interpretation of regression coefficients. Finally, the off-diagonal elements in the matrix are the cross-partial derivatives and represent the indirect (or spillover) impacts, which are again summarized by averaging the row sums of the respective matrix elements. In other words, this measure records the effect on the steady-state level income in country i resulting from a change in the investment rate in all countries except country i. Hence, the average indirect impact is given by the difference between the average total impact and the average direct impact.

2.5 Data, Estimation Results, and Robustness

This section starts by providing information on the data sources used to assemble the data set for the empirical analyses and on how the variables were constructed from the source data. Thereafter, robustness checks on the results in Ertur and Koch (2007) are conducted and discussed. The first set of robustness checks in Section 2.5.2 considers the sensitivity of the results to changing the version of the PWT from 6.1 to 6.2 and 7.1, respectively. Next, in Section 2.5.3, estimation results are reported and discussed for the specification in which technological interdependence is modeled via genetic distance. Again, sensitivity of the results is assessed by estimating the model with data from the three different versions of the PWT.

2.5.1 Data

The main data source for the replication exercise is PWT 6.1 (Heston et al., 2002), while for the robustness checks PWT 6.2 (Heston et al., 2006) and 7.1 (Heston et al., 2012) are used. Additional versions of the PWT exist as well, for example, PWT 6.3 and 7.0. However, Breton (2012) has noted substantial issues with Version 7.0. Moreover, as Johnson et al. (2013, 257) point out, the exposition would soon become intractable if one aimed at a comparison between every single version.[89] As in Ertur and Koch (2007, 1042), the initial sample covers the 91 countries of the non-oil sample in Mankiw et al. (1992), for which data is available over the period 1960-1995.[90] In contrast to the theoretical model, GDP per capita and GDP per worker are not in fact identical, as not the whole population in a country is employed. Hence, for the empirical exercise, the dependent variable, y, is real GDP (evaluated via the chain method) per worker (variable *rgdpwok* in PWT). The investment rate, s, is the real share of investment in real GDP (variable *ki* in PWT) averaged over the respective years. For the average growth rate of workers, n, no directly corresponding variable is available in PWT. A number for the size of the working-age population can be recovered however by noting that the series for real GDP per capita and population are available so that the number of workers can be calculated by multiplying real GDP per capita (*rgdch* in PWT) by the size of the population (*pop* in PWT) and dividing the result by the value of real GDP per worker (Ertur and Koch (2007, 1042) refer to Caselli (2005, 685) for this method). The average growth rate of the working-age population is then calculated as an approximation (though this is not stated explicitly in Ertur and Koch (2007)) by taking the natural logarithm of the number of workers in 1995, subtracting the natural logarithm of the number of workers in 1960, and dividing the result by the number of years, i.e. 35.

[89] See Table 2 in Johnson et al. (2013) for an overview of the evolution of the PWT up to Version 7. More recent versions of the Penn World Table (8.0 and 8.1, respectively) are also available (Feenstra et al., 2015). These data sets will however not be used in this analysis, as these versions lack data on the real share of investment in real GDP. See Table A3 in the document "variable correspondence" available under `http://www.rug.nl/research/ggdc/data/pwt/pwt-8.0` (accessed: 11 August, 2015), which states that not only is this variable not reported in PWT 8.0, but neither is it possible to construct it from the source data. This continues to be valid for PWT 8.1 (see `http://www.rug.nl/research/ggdc/data/pwt/pwt-8.1.xml` (accessed: 11 August, 2015).

[90] Appendix B.6 lists these countries.

For the construction of the interaction matrices the general assumptions made in Section 2.3.2 are valid. An additional important point to note is that the weights in these matrices should be exogenous with respect to model (Ertur and Koch, 2007, 1042). This makes geographic and genetic distance ideal candidates.[91] The matrices that are based on spatial distances use as weights the great circle distances, d_{ij}, between country capitals i and j. There is however some scope in pinning down the latitude and longitude of a capital, and Ertur and Koch provide no information for their source of this data. In this chapter, in all calculations that rely on latitude and longitude, the coordinates are taken from the CIA's World Factbook (Central Intelligence Agency, 2013), and the distances are calculated as described in Appendix B.2.[92] As a final step, the weights for the interaction matrices are given by $w_{ij}(1) = w^*_{ij}(1)/\sum_j w^*_{ij}(1)$ and $w_{ij}(2) = w^*_{ij}(2)/\sum_j w^*_{ij}(2)$, and the weights are based on the following functional forms

$$w^*_{ij}(1) = \begin{cases} 0 & \text{if } i = j \\ d_{ij}^{-2} & \text{otherwise} \end{cases} \tag{2.22a}$$

$$w^*_{ij}(2) = \begin{cases} 0 & \text{if } i = j \\ e^{-2d_{ij}} & \text{otherwise.} \end{cases} \tag{2.22b}$$

Applying the transformations $w_{ij}(1) = w^*_{ij}(1)/\sum_j w^*_{ij}(1)$ and $w_{ij}(2) = w^*_{ij}(2)/\sum_j w^*_{ij}(2)$, ensures that the row entries in the interaction matrix indeed sum to one. Adopting the inverse of the squared distance as a functional form in Equation (2.22a) reflects a gravity function (Ertur and Koch, 2007, 1042) and captures that the effect of the spatial externalities weakens more than proportionally with distance; a result that has received support in the empirical literature (see e.g. Keller, 2002). The spatial weight matrix based on the weights in Equation (2.22a) is called $\boldsymbol{W_1}$ and the one in Equation (2.22b), which Ertur and Koch (2007) employ as a robustness check, is $\boldsymbol{W_2}$.[93]

[91] Another interesting variable on which to base the weights would be, for example, a measure of technological proximity between countries. However, this measure could not be considered exogenous to the model for the sample period considered in this chapter and it would be problematic in the case of technology to disentangle the effects on income per worker due to changes in $\boldsymbol{X}$ from those effects due to changes in $\boldsymbol{W}$.

[92] For some countries the capital has moved to a different city over the period from 1960 to 1995. The capital for Côte d'Ivoire, for example, has moved from Abidjan to Yamoussoukro in 1983. In these cases, the coordinates of the city, which was the capital over the longer period with respect to the sample horizon was used. Recent versions of the World Factbook, however, lack geographic coordinates for former capitals so that for these capitals the coordinates have been gathered via Google Maps. This approach has also been employed for cities like Hong Kong that are not listed in the World Factbook.

[93] This latter matrix is also the one that has been used to calculate the spatial lag of TFP in the Moran scatterplot in Figure 2.2.

The data on genetic distance is taken from the data set of Spolaore and Wacziarg (2009), who rely on data assembled by Cavalli-Sforza et al. (1994). Following the construction of the original weight matrices based on geographic distance, the functional form in Equation (2.22a) has been chosen for the interaction matrices based on genetic distances as well so that a straightforward robustness check is possible. For the interaction matrix $\boldsymbol{W_3}$, the distances d_{ij} are based on the concept of weighted F_{ST} genetic distance (see Section 2.2 and Appendix B.1 for this measure) and for matrix $\boldsymbol{W_4}$ on weighted Nei's genetic distance (Nei, 1972, 1973).

2.5.2 Results – Interaction Matrix Based on Geographic Distance

Estimation results are presented in Table 2.1.[94] The first two columns replicate the results from Table 1 in Ertur and Koch (2007) and serve as a benchmark compared to which all subsequent robustness analyses will be assessed.[95]

In this analysis, the interaction matrix based on the weights in Equation (2.22a) has been used. Column 1 shows in the upper half the results for the standard Solow model estimated by ordinary least squares (OLS). The estimated coefficients on the investment rate and on the population growth rate have the signs expected from the theoretic model and in addition are highly significant. In the lower half, this model is estimated with the restriction $\beta_1 = -\beta_2$ imposed. This restriction is tested with a Wald test and rejected (p-value = 0.038). Furthermore, the implied value for the capital share, $\alpha = 0.58$, is too high compared to empirical estimates. Gollin (2002, 458), for example, estimates that the capital shares for most countries lie in the range of 20% to 35%. Also, Moran's I test indicates spatial autocorrelation in the error term. Based on these results, Ertur and Koch (2007, 1046) thus conclude that the standard model is misspecified as it does not account for physical capital externalities and technological interdependence between countries.

Column 2 shows that the estimation results support the implications of the spatially augmented model. All coefficients have the signs predicted from theory (compare Equation (2.17)), even though, for instance, the estimated coefficient associated with the spatial

[94] All estimations have been carried out in Matlab using the Spatial Econometrics Toolbox by LeSage, which is publicly available under: `http://www.spatial-econometrics.com/` (accessed: 11 August, 2015).

[95] Note that since the analysis here is based on the geographic coordinates from the CIA's World Factbook, and these coordinates differ in some cases slightly from the ones in Ertur and Koch (2007), the values for Moran's I test in the unrestricted and restricted versions of the standard Solow model in Column 1 as well as the values for the spatially augmented Solow model in Column 2 are somewhat different. Qualitatively, the results are not affected though. Also, there is a small mistake in Ertur and Koch's Table I, as the values for Moran's I test in the unrestricted Solow model belongs to the restricted Solow model and vice versa.

lag of the population growth rate is insignificant (p-value $= 0.479$). The likelihood ratio test does not reject the joint theoretical restriction $\beta_1 + \beta_2 = 0$ and $\theta_1 + \theta_2 = 0$, as the p-value is 0.419, which supports the validity of the spatially augmented model. In addition, the (significant) implied value for the capital share of income is $\alpha = 0.284$ and thus falls approximately right in middle of the range of estimates by Gollin (2002). Furthermore, the parameter ϕ reflecting physical capital externalities is positive and significant at the 10%-level. Also, the implied value for γ, which gauges the degree of technological interdependence among the countries is positive and highly significant, implying that this characteristic indeed needs to be taken into account in growth models, as economies cannot be considered as independent observations (Fischer, 2011, 432). Finally, the value of $\alpha + \phi/(1 - \gamma)$ is below 1, implying that the externalities in the model are not strong enough to lead to endogenous growth (Ertur and Koch, 2007, 1048). In sum, the estimation results therefore provide rather strong support for the model developed by Ertur and Koch.

The next columns in Table 2.1 assess the sensitivity of these results when changing the underlying data source to more recent versions of the PWT. Due to missing data, for instance, values for the variable capital investment are not available for some countries in PWT 6.2, the sample size needs to be reduced to 83 countries in the estimations based on this data source.[96] In order to obtain estimation results for a balanced sample across all three versions of the PWT considered in this chapter, Columns 3 and 4 show estimation results for the 83-country sample with data from PWT 6.1. For the standard model, the results are virtually identical (Column 3) to those from the full sample. However, dropping these 8 observations from the sample affects the results in the spatial model. The implied values for α and ϕ are comparable in size to the full sample with 91 countries, but they are insignificant in the smaller sample. Hence, dropping these 8 countries from the sample already puts a small dent in the robustness of the results obtained by Ertur and Koch (2007).

Columns 5 and 6 change the data source to PWT 6.2. In Column 5 of Table 2.1 the estimation results are in line with those from Columns 1 and 3. The only exception is that for this data source the restriction $\beta_1 = -\beta_2$ is not rejected (p-value $= 0.476$), suggesting a good fit between the model and the data, except that the implied value for the capital share is still too high with $\alpha = 0.576$. For the unconstrained estimation of the spatial model, Column 6 shows that compared to Columns 2 and 4, the coefficient for the population growth rate still has the sign implied by the theoretical model, but is now insignificant (p-value $= 0.347$). The results from the estimation with the joint parameter restriction applied, show that, as for the results for the 83-country sample with data from

[96]See Appendix B.6 for the eight countries with missing data.

Table 2.1: Estimation Results for the Standard and Spatially Augmented Solow Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_1}$ (Geographic Distance).

Data set	PWT 6.1				PWT 6.2		PWT 7.1	
Model	Stand.	Spatial	Stand.	Spatial	Stand.	Spatial	Stand.	Spatial
Number of observations	91	91	83	83	83	83	83	83
Unconstrained estimation:								
Constant	4.651	0.886	4.609	0.518	7.130	2.780	2.976	1.828
	(0.010)	(0.635)	(0.017)	(0.796)	(0.000)	(0.181)	(0.189)	(0.399)
$\ln s_i$	1.276	0.836	1.234	0.789	1.319	0.876	1.697	0.944
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−2.709	−1.538	−2.701	−1.449	−1.835	−0.689	−3.428	−1.441
	(0.000)	(0.006)	(0.000)	(0.021)	(0.008)	(0.347)	(0.000)	(0.081)
$\boldsymbol{W} \ln s_j$	—	−0.347	—	−0.314	—	−0.160	—	0.710
		(0.057)		(0.137)		(0.514)		(0.110)
$\boldsymbol{W} \ln(n_j + 0.05)$	—	0.591	—	0.343	—	−0.191	—	−0.298
		(0.479)		(0.705)		(0.843)		(0.793)
$\boldsymbol{W} \ln y_j$	—	0.742	—	0.732	—	0.608	—	0.595
		(0.000)		(0.000)		(0.000)		(0.000)
Moran's I test	0.432	—	0.397	—	0.346	—	0.389	—
	(0.000)		(0.000)		(0.000)		(0.000)	
Constrained estimation:								
Constant	8.375	2.118	8.407	2.220	8.465	3.158	7.321	1.939
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)
$\ln s_i - \ln(n_i + 0.05)$	1.379	0.855	1.354	0.813	1.356	0.871	1.904	0.958
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\boldsymbol{W}[\ln s_j - \ln(n_j + 0.05)]$	—	−0.292	—	−0.230	—	−0.149	—	0.692
		(0.098)		(0.270)		(0.527)		(0.109)
$\boldsymbol{W} \ln y_j$	—	0.735	—	0.721	—	0.613	—	0.608
		(0.000)		(0.000)		(0.000)		(0.000)
Moran's I test	0.415	—	0.4397	—	0.342	—	0.377	—
	(0.000)		(0.000)		(0.000)		(0.000)	
Test of restriction	4.427	1.738	4.066	1.474	0.514	0.127	3.805	0.358
	(0.038)	(0.419)	(0.047)	(0.479)	(0.476)	(0.938)	(0.055)	(0.836)
Implied α	0.580	0.284	0.575	0.242	0.576	0.196	0.656	8.261
	(0.000)	(0.012)	(0.000)	(0.120)	(0.000)	(0.403)	(0.000)	(0.852)
Implied ϕ	—	0.177	—	0.206	—	0.270	—	−7.772
		(0.082)		(0.139)		(0.213)		(0.861)
Implied γ	—	0.554	—	0.525	—	0.408	—	−0.043
		(0.000)		(0.000)		(0.009)		(0.868)
$\alpha + \frac{\phi}{1-\gamma}$	—	0.680	—	0.676	—	0.651	—	0.808
		(0.000)		(0.000)		(0.000)		(0.000)

Note: p-values are given in parentheses. For the standard Solow model the restriction is tested with the Wald test and for the spatially augmented model the restriction is tested with the likelihood ratio (LR) test.

PWT 6.1, the implied share of capital income and the parameter for the physical capital externalities are insignificant (p-values of 0.403 and 0.213, respectively). Hence, changing the data source from PWT 6.1 to 6.2, suggests that while many results (e.g. concerning the implied value of γ or the test of the joint restriction) are not sensitive to this change, the original results by Ertur and Koch (2007) concerning the implied capital share of income and the parameter ϕ are not robust.

More drastic changes to the original results are visible when moving to PWT 7.1 in Columns 7 and 8. For the standard model in Column 7, the signs of the coefficient estimates have the expected signs, and Moran's I test indicates misspecification with respect to spatial correlation in the error term. In accordance with the results for the PWT 6.1 sample, the parameter restriction $\beta_1 = -\beta_2$ is rejected (p-value $= 0.055$), though this time at the 10% level instead of the 5% level. However, in the spatially augmented model in Column 8, the constrained estimation implies an implausibly large share of capital income. The estimated value for α is 8.261 (although this value is not significant with a p-value of 0.852). Moreover, the value for the physical capital externalities is now negative, but also not significant ($p = 0.861$). The same holds qualitatively for the parameter measuring technological interdependence. These estimates imply that using a more recent data source, leads to drastic changes in the empirical results compared to the benchmark results.[97]

It needs to be kept in mind though that in addition to the results in Table 2.1 the model's interpretation relies on the calculation of the direct and indirect effects from changes in the exogenous variables via the approach presented in Section 2.4.3. The results for these impacts are presented in Table 2.2 for all four samples considered in this chapter. In the paper by Ertur and Koch only the direct effects are reported (though without any reference to the significance of these estimates). Here, a richer analysis is presented by also reporting estimates for the indirect and total impacts on steady-state per worker income due to changes in the exogenous variables and by providing information about the significance of all three impacts as well.

Concerning the direct impacts, the results show that across all four samples an increase in the investment rate in physical capital is approximately comparable in size and significance. The estimated coefficients are highly significant and imply, due to the logarithmic specification of the model, that a 10% increase in the investment rate would result in an increase in per capita income between 8.6% and 11.6%. The results for the indirect impacts of changes in the investment rate, resulting from spatial spillovers, differ however across the samples. Whereas these impacts are comparable in size for the first three

[97]That changing the data source from e.g. PWT Version 6.1 or 6.2 to 7.1 can lead to different results in models similar to the one considered here has also been pointed out by Johnson et al. (2013, 270). They find that in the Solow model augmented with human capital, developed by Mankiw et al. (1992), the coefficient on the investment share is reduced in size close to zero, when the estimation is based on a more recent version of the PWT (7.0 in their case). This finding is attributed to the investment series being more variable in this version of the PWT. However, they also state that the reason for this higher variability is unclear (Johnson et al., 2013, 270).

Table 2.2: Estimation Results for the Direct, Indirect and Total Impacts in the Spatial Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_1}$ (Geographic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Direct impacts:				
$\ln s_i$	0.916	0.859	0.941	1.158
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.693	−1.636	−0.793	−1.635
	(0.005)	(0.013)	(0.269)	(0.043)
Indirect impacts:				
$\boldsymbol{W} \ln s_j$	1.030	0.960	0.915	3.012
	(0.118)	(0.198)	(0.057)	(0.004)
$\boldsymbol{W} \ln(n_j + 0.05)$	−2.008	−2.559	−1.458	−2.709
	(0.484)	(0.423)	(0.476)	(0.218)
Total impacts:				
$\ln s_i + \boldsymbol{W} \ln s_j$	1.945	1.820	1.856	4.170
	(0.007)	(0.023)	(0.000)	(0.000)
$\ln(n_i + 0.05) + \boldsymbol{W} \ln(n_j + 0.05)$	−3.701	−4.196	−2.251	−4.343
	(0.230)	(0.220)	(0.294)	(0.054)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

samples, the impact is only significant for the PWT 6.2 sample at the 10%-level. For the PWT 7.1 sample, this effect has tripled in size compared to the other estimates and is significant at the 1%-level. These findings indicate again that the results in Ertur and Koch are not robust with respect to changing to more recent versions of the Penn World Table. It is however interesting to note that at least for the first three samples the direct and indirect impacts from the investment rate contribute both approximately 50% to the total impact of this variable. Table 2.2 also shows that the results concerning the impacts of the population growth rate are not robust across samples.

Finally, the estimates in Table 2.2 illustrate that basing the interpretation of the model on the estimates in Table 2.1 would lead to incorrect conclusions. For instance, interpreting the coefficient associated with the spatial lag of the investment rate, $\boldsymbol{W} \ln s_j$ in Column 2 of Table 2.1, as an indicator for the indirect impact would lead to the inference that this effect is negative (−0.347), implying that an increase in the investment rate in neighboring countries would result in a decrease of per capita income in the country under consideration. The true impact estimate in Column 1 of Table 2.2, however, suggests that the spillover effect is positive, though marginally insignificant at the 10%-level.

Before turning to the estimation results for the interaction matrix based on genetic distance, it should be remembered that Ertur and Koch have also employed an interaction matrix based on the specification in Equation (2.22b) to assess the sensitivity of their results using their initial choice of weight matrix. The detailed results of the robustness analysis for this interaction matrix across all four samples are delegated to Appendix B.5. Tables B.1 and B.2 in this appendix demonstrate again that some of the original results fail to hold when estimating the model across the different samples.

Concerning interaction matrix $\boldsymbol{W_2}$, an important comment needs to be made. This matrix does not seem to correspond exactly to the specification Ertur and Koch (2007) actually use in their empirical analysis. From the Matlab code on the article's website,[98] it is clear that their estimation results are obtained by dividing the geographic distances d_{ij} by 1,000. A reason for this transformation is not given however, and it turns out that the estimation results are highly sensitive to this alternative specification (see the results in Tables B.3 and B.4 in Appendix B.5). For instance, without dividing the distances by 1,000, the estimation results imply highly significant negative values for the parameters ϕ and γ, and the implied capital share of income increases to an unreasonably, but highly significant value of 90%. With respect to the impact estimates, the values for the direct and total impacts are approximately comparable across both specifications, the indirect effects however turn from being not significant in the specification as implemented by Ertur and Koch to being strongly significant in the specification as claimed in the article (i.e. without the division by 1,000).

2.5.3 Results – Interaction Matrix Based on Genetic Distance

This section presents the estimation results for the model in which the interaction matrix is based on genetic distance. The general specification of the weights is given by the one in Equation (2.22a), and the analyses use weighted F_{ST} genetic distance.[99] As the results for the standard model do not depend on the interaction matrix, Table 2.3 shows only the results from the estimation of the spatial Durbin model. The results for the standard model are suppressed in order to avoid duplication.[100]

Column 1 provides the results for the full sample of 91 countries for data taken from PWT 6.1. In contrast to the benchmark, i.e. the original results in Ertur and Koch,

[98] See `http://qed.econ.queensu.ca/jae/datasets/ertur001/` (accessed: 23 July, 2014).

[99] For the results based on (weighted) Nei's distance see Tables B.5 and B.6 in Appendix B.5. The estimation results are comparable to the ones shown in Tables 2.3 and 2.4, which might be explained by the fact that the correlation between the two measures of genetic distance is 93.9% (Spolaore and Wacziarg, 2009, 482) and thus very high.

[100] It is worth pointing out however, that in all samples the standard model continues to be misspecified, as the values for Moran's I test suggest spatial autocorrelation in the error term (p-values are 0.000 in all four tests) also when using interaction matrix $\boldsymbol{W_3}$ in these tests.

Table 2.3: Estimation Results for the Spatial Durbin Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_3}$ (Weighted F_{ST} Genetic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Unconstrained estimation:				
Constant	8.654	8.246	5.941	−1.932
	(0.001)	(0.000)	(0.011)	(0.423)
$\ln s_i$	0.820	0.945	0.888	0.972
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.034	−0.871	−0.148	−0.930
	(0.054)	(0.099)	(0.790)	(0.153)
$\boldsymbol{W} \ln s_j$	0.901	0.665	0.725	−0.009
	(0.000)	(0.001)	(0.002)	(0.983)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.651	0.431	−1.625	−1.912
	(0.500)	(0.632)	(0.096)	(0.078)
$\boldsymbol{W} \ln y_j$	0.322	0.327	0.198	0.556
	(0.006)	(0.002)	(0.128)	(0.000)
Constrained estimation:				
Constant	5.520	5.452	6.013	2.452
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_i - \ln(n_i + 0.05)$	0.785	0.856	0.870	0.996
	(0.000)	(0.000)	(0.000)	(0.000)
$\boldsymbol{W}[\ln s_j - \ln(n_j + 0.05)]$	0.850	0.653	0.655	0.130
	(0.000)	(0.001)	(0.003)	(0.743)
$\boldsymbol{W} \ln y_j$	0.280	0.296	0.245	0.605
	(0.019)	(0.005)	(0.045)	(0.000)
Test of restriction	2.450	1.949	1.862	3.708
	(0.294)	(0.377)	(0.394)	(0.157)
Implied α	1.491	1.830	1.598	−0.273
	(0.003)	(0.052)	(0.038)	(0.803)
Implied ϕ	−1.052	−1.360	−1.133	0.772
	(0.039)	(0.155)	(0.147)	(0.474)
Implied γ	−0.319	−0.189	−0.219	0.238
	(0.098)	(0.196)	(0.212)	(0.288)
$\alpha + \frac{\phi}{1-\gamma}$	0.694	0.686	0.669	0.740
	(0.000)	(0.000)	(0.000)	(0.000)

Note: p-values are given in parentheses. The restriction for the spatially augmented model is tested with the likelihood ratio (LR) test.

the estimates based on genetic distance show, for instance, that the coefficient associated with the spatial lag of the investment rate is now positive and highly significant (p-value = 0.000). The results for the constrained estimation also differ from the ones with an interaction matrix using geographic distance, as the implied value for α is now implausibly large and highly significant. Moreover, the estimate for γ, measuring the degree of technological interdependence is now negative and marginally significant at the

10%-level, which seems implausible.[101] Similar results are also obtained for the other samples. When using data from PWT 7.1 for instance, the implied value for the capital share of income in Column 4 actually turns negative (although the p-value is 0.803). In neither sample, based on a likelihood ratio test, the joint parameter restriction $\beta_1 + \beta_2 = 0$ and $\theta_1 + \theta_2 = 0$ is rejected though.

Table 2.4: Estimation Results for the Direct, Indirect and Total Impacts in the Spatial Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_3}$ (Weighted F_{ST} Genetic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Direct impacts:				
$\ln s_i$	0.887	1.004	0.918	1.035
	(0.000)	(0.000)	(0.011)	(0.000)
$\ln(n_i + 0.05)$	−1.009	−0.856	−0.198	−1.229
	(0.058)	(0.101)	(0.722)	(0.060)
Indirect impacts:				
$\boldsymbol{W} \ln s_j$	1.673	1.401	1.098	1.116
	(0.000)	(0.000)	(0.000)	(0.156)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.506	0.260	−1.996	−5.246
	(0.701)	(0.830)	(0.062)	(0.001)
Total impacts:				
$\ln s_i + \boldsymbol{W} \ln s_j$	2.560	2.405	2.015	2.151
	(0.000)	(0.000)	(0.000)	(0.012)
$\ln(n_i + 0.05) + \boldsymbol{W} \ln(n_j + 0.05)$	−0.503	−0.560	−2.193	−6.475
	(0.707)	(0.627)	(0.035)	(0.003)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

Despite these results from the estimation of the constrained model across the four samples, the impact estimates in Table 2.4, which are calculated from the unconstrained estimation results will be briefly discussed. Across all samples, the estimates for the direct impact of a change in the investment rate on steady-state per capita income is comparable to the results for the model with interaction matrix $\boldsymbol{W_1}$. One important difference to the results in Table 2.2 concerns the spillovers from a change in the investment rate for the full sample of 91 countries. The estimated effects are now highly significant and imply that a change of 1% in the investment rate in all countries except country i would result in an increase of approximately 1.7% in per capita income in country i. Another interesting result is that these spillovers are not significant in the sample for PWT 7.1 in Column 4

[101] The implied values for α, ϕ, and γ are of approximately similar size in the estimation based on Nei's distance for this sample, though neither value is significant at the 10%-level. This is the exception from the claim about comparable estimation results for both measures of genetic distance made in Footnote 99.

of Table 2.4, whereas the reverse holds for this sample in the estimation with geographic distance. There (see Column 4 in Table 2.2), this estimate is not only highly significant, but also large in size. Table 2.4 furthermore clearly shows that the impacts with respect to the population growth rate are highly sensitive to the particular version of the PWT used in the estimation. Hence, it can be concluded that the original results by Ertur and Koch are challenged strongly both by changing the measure in the interaction matrix from geographic to genetic distance and also by substituting the data set from older versions of the Penn World Table for more recent versions.

2.6 Conclusion for Chapter 2

This chapter has presented the growth model with technological interdependence among countries developed by Ertur and Koch (2007) and subjected their empirical results to a series of robustness checks. In contrast to the original specification, which uses an interaction matrix based on geographic distance, here data on measures of genetic distance from Spolaore and Wacziarg (2009) has been used to construct an alternative interaction matrix. Furthermore, additional robustness checks have been conducted to assess the sensitivity of the original results across different versions of the Penn World Table for the same period and the same set of countries. The analyses show that the original results by Ertur and Koch are only robust to a certain extent. While the hypothesis that countries need to be analyzed in an interdependent system is supported, other results are highly sensitive to the version of the Penn World Table that is used in the empirical estimation. Ertur and Koch (2007) estimate, for instance, an implied capital share of income slightly below 30%, which is significant at the 5% level. However, this result is not robust when estimating the model with data from PWT 6.2 or 7.1 instead.

Furthermore, whereas Ertur and Koch only provide estimates of the direct impacts on per worker income associated with changes in the exogenous variables, in this chapter values for the indirect and total impacts have been calculated as well. The results again indicate non-robustness across different versions of the PWT, as, for example, the indirect impacts (or spillovers) associated with changes in the investment rate of physical capital on per worker income in steady state are not significant in the PWT 6.1 sample, but significant in the ones based on PWT 6.2 and 7.1, respectively. Results have also been shown to be highly sensitive to the precise specification of the weights in the interaction matrix based on geographical distance.

Based on theoretical and technical considerations, genetic distance has been introduced as an alternative measure to geographic distance, on which to base an interaction matrix. Concerning the empirical results for this alternative matrix, it is found that, whereas in the original model indirect spillovers from capital investment were insignificant in the PWT 6.1 sample, using a measure of genetic distance, these spillovers now have a significant effect on steady-state income per worker. However, the version of the model with an interaction matrix based on genetic distance implies an implausibly large capital share of income. In addition, also in the case of genetic distance, estimation results are sensitive to the version of the data set. It can thus be stated that the empirical results in Ertur and Koch are highly sensitive to the measure on which the weights in the interaction matrix are based (geographic or genetic distance) as well as to the concrete specification of the weights in the interaction matrix. This conclusion vividly demonstrates the importance of sound data when it comes to, for instance, giving policy advice.

In this chapter, as, for example, in Fischer (2011), only level regressions have been addressed. Future work will also investigate the sensitivity of the estimates for the growth regressions in Ertur and Koch (2007), as well as the impact of introducing human capital into this model. Results from Ertur and Koch (2006) suggest that this factor is not related to growth within this framework. However, as the results in this chapter clearly demonstrate, that this holds across different versions of the PWT need not necessarily be the case. It should also be pointed out that an endogenous version of the model framework exists (Ertur and Koch, 2011), which for a smaller set of countries and a shorter time period provides empirical support, based in part on data from PWT 6.2, in favor of the endogenous version. But again, this is no guarantee that this necessarily needs to hold across different versions of the PWT. Robustness should be assessed for this finding as well. As this chapter has also clearly demonstrated the sensitivity of the empirical results to the precise choice of the interaction matrix, further research will be devoted to this issue. In particular, the method of Bayesian Model Averaging will be used to address the uncertainty concerning the specification of the interaction matrix in this model.

Before concluding this chapter, a brief remark on policy concerning the role of geographic and genetic distance in determining the strength of knowledge spillovers: As humans have demonstrated numerous times over the course of history in often horrible ways, neither the geographic distances (through expansionary wars or state collapses) nor the genetic distances (via genocide or the slave trade) between countries are fixed in the long run.[102] Abstracting from these, policy can, however, still have an impact by e.g. fostering openness and thereby removing barriers to the diffusion of knowledge (Spolaore and Wacziarg, 2009, 524).

[102] Also another, maybe at first glance more innocuous possibility, comes to mind in light of the dramatically falling costs for gene sequencing. These are currently below \$4,500 for the whole human genome compared to nearly $\$100,000,000$ in 2001 (Wetterstrand, 2015). I do not think that it is completely unrealistic that personal genetic data will be used by states to influence, for instance, visa decisions.

Chapter 3: Schumpeterian Growth with Technological Interdependence: An Application to US States[103]

3.1 Introduction to Chapter 3

Interaction between countries or regions occurs in many forms. One particular dimension of this interaction concerns the diffusion of knowledge or knowledge spillovers. Given that knowledge is a key factor in economic development, this implies that the level of development, measured by, for example, income per capita, in one region depends on characteristics in the regions it interacts with.

The presence of these interdependent relationships motivates the need to incorporate these also in theoretic models. One example in the recent economic literature that takes into account interdependence between countries is the exogenous growth model developed in Ertur and Koch (2007). In this chapter, the transition to an endogenous growth model will be made by presenting a model by the same authors (Ertur and Koch, 2011), which builds heavily upon the contributions by Aghion and Howitt (1998) and Howitt (2000). The novelty of the model by Ertur and Koch (2011) is that it incorporates complex spatial interactions, modeled via technological interdependence between regions, in the context of an endogenous growth model, in which profit-driven investment in research and development (R&D) determines the rate of technological progress. In particular, the authors develop an integrated theoretical and empirical framework that nests a series of growth models.

This chapter fills a gap in the literature as, to the best of my knowledge, the model has not yet been investigated empirically for the US states. The shift in focus from a cross-country to a cross-regional setting is important for the following reasons. First, the United States is the global leader in investments in R&D. In 2011, R&D investments in the United States accounted for approximately 30% of the global total, far ahead of the next-ranked countries China, Japan, and Germany with shares of 15%, 10%, and 7%,

[103] Chapter 3 has been published in similar form as Working Paper No. 75 in the Working Paper Series in Economics at KIT (Deeken, 2015b).

respectively. The dominance was even higher in 2001, when the United States' global share was 37% (all figures are from National Science Board (2014, 4-17)). The second reason for choosing US states as the units of analysis addresses interdependence between these units. As Keller (2002) points out, the strength of technological knowledge spillovers declines with the geographic distance between the originating and the receiving country, implying that diffusion of technology is not a frictionless process. Geographic distance in this situation captures, for instance, socio-economic differences, but also those in institutions between countries (Ertur and Koch, 2007, 1036), which have been highlighted in the growth literature as a fundamental determinant of cross-country income differences.[104] The advantage of studying diffusion of technology within a single country is the common institutional setting,[105] which possibly reduces part of the frictions. The third reason for choosing the United States relates back to the first. Eaton and Kortum (1996) find that for the OECD countries the amount of a country's growth in productivity that depends on research efforts in the United States is larger than 50%, which points to substantial spillovers from the United States. In addition, Eaton and Kortum (1999, 558) estimate that in the past 60% of the United States' productivity growth originated from research conducted domestically. This figure is in stark contrast to the corresponding values for Japan or Germany, with figures of 16% or 35%, respectively, and it raises the question, if significant spillovers also exist between US states or only between the Unites States and other countries. The cited figure of 60% is silent about any spillovers between US states.

Indicative evidence for the potential existence of these spillovers is provided by the map in Figure 3.1, which shows the average R&D investment rate (or R&D intensity) over the period 1997-2007 in the 48 continental US states plus the District of Columbia.[106]

States with average R&D investment rates above 2% can be found predominantly on the western seaboard and in the south-west (with the exception of Nevada) as well as in the north-east and the region around the Great Lakes (the notable exception in these regions is Maine). In regions where states with high R&D intensities abound, the potential for spillovers and a subsequent impact on output is high.

Figure 3.2 illustrates the data on the R&D intensities in a different way. It shows a Moran scatterplot of the average R&D intensity (in standardized form) on the horizontal axis, and the vertical axis measures the standardized value of the spatial lag of this variable. The value of the spatial lag for a given state comprises the average of the R&D intensities of this states' direct neighbors.

[104] As a starting point, consider the seminal contribution by Acemoglu et al. (2001).

[105] However, US states have considerable autonomy in the United States' federalist system. For a comparison with the German system on this aspect, see Halberstam and Hills, Jr. (2001).

[106] A correspondence between state abbreviations and names is provided in Table C.6 in Appendix C.5.

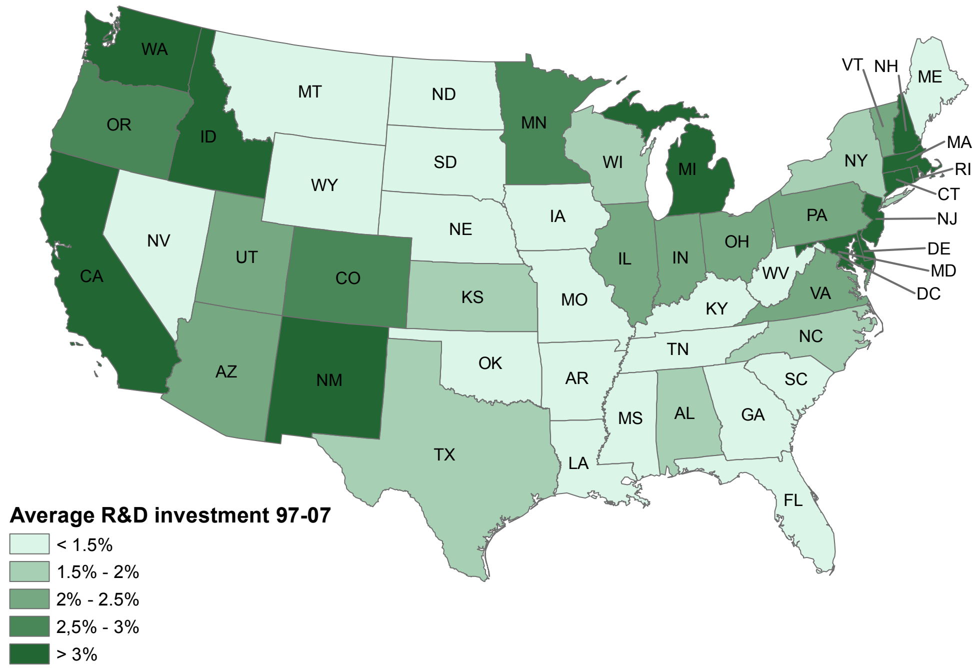


Figure 3.1: Average R&D Investment Rate for the 48 Continental US States plus Washington, D.C. over the period 1997-2007 (Data: OECD, 2015).

Each point in the scatterplot corresponds to a single state so that states in the upper right quadrant have average R&D intensities above the mean and are also surrounded by states for which the same holds. The reverse holds for states in the lower left quadrant, whereas in the upper left quadrant states can be found whose own average R&D investment rate is below the mean, but who are neighbors to states with above-average R&D intensities. The dot labeled "NM" in the figure denotes New Mexico, which has the highest R&D intensity in the sample.[107] However, its neighboring states fall below the average.

The chapter is organized as follows: Section 3.2 introduces the basic structure of the multi-region Schumpeterian before Section 3.3 specifies the nature of technological interdependence between regions and derives the equation for the income per worker in steady state. In Section 3.4, the focus is on the empirical specification of the model and the estimation strategy. The data for the empirical analysis is presented in detail in Section 3.5, which also discusses the estimation results. Finally, Section 3.6 concludes this chapter.

[107]The high value for New Mexico can be explained by the presence of Los Alamos National Laboratory and Sandia National Laboratories, which are federally funded research and development centers. Compare the information by the National Science Foundation available under `http://www.nsf.gov/statistics/infbrief/nsf02322/` (accessed: 9 August, 2015).

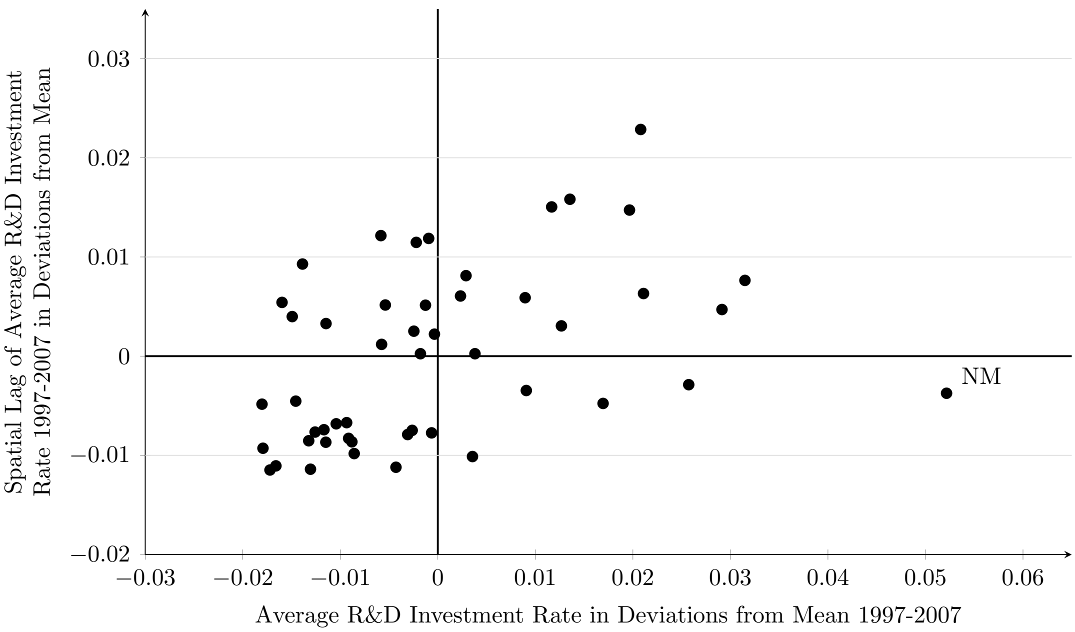


Figure 3.2: Moran Scatterplot of the Average R&D Investment Rate for the 48 Continental US States plus Washington, D.C. over the period 1997-2007 (Data from (OECD, 2015)).

Note: The variables are in the form of deviations from the mean so that the value 0 on the abscissa is equivalent to the mean value of 2.2%.

3.2 Multi-Region Schumpeterian Growth Model without Technological Interdependence

This section describes the multi-region Schumpeterian growth model in Ertur and Koch (2011), which builds upon work by Aghion and Howitt (1998, Chapter 3 and 12.2) and Howitt (2000). The expression "multi region" that is attached to this setup might be a slight misnomer though, as each region is assumed to develop independently from the other regions so that the term "single-region model" would be more appropriate for this section. However, to make the transition to the multi-region model in Section 3.3 easier, already here a single region in the economy will be indexed. Section 3.2.1 describes the production side of the region's final good sector and Section 3.2.2 illustrates its intermediate goods sector, before Section 3.2.3 clarifies the connections in the research and development (R&D) sector.

3.2.1 Final Good Sector

The economy under consideration consists of $i = 1, \dots, N$ regions. A single final good is produced in each region with labor and a continuum of intermediate goods (or varieties) as input factors. The final good sector operates under perfect competition, and the good is produced via the following production function, illustrated here for region i,

$$Y_i(t) = Q_i(t)^{\alpha-1} \int_0^{Q_i(t)} A_i(v,t) x_i(v,t)^{\alpha} L_i(t)^{1-\alpha}\, dv, \tag{3.1}$$

where $Y_i(t)$ is output in region i at time t. This output, besides its use as a consumption good, also functions as a capital good in the production of intermediates and as an input into research and development activities. The variable $x_i(v,t)$ measures the flow of intermediate good v used in the production of the final good, and $Q_i(t)$ indicates how many different intermediate goods exist in region i at time t. The continuum of intermediates is therefore measured on the interval $v \in [0, Q_i(t)]$. $A_i(v,t)$ is a productivity parameter, which reflects the quality of intermediate product v and thus increases with successive vintages of the good. Finally, $L_i(t) = L_i(0)e^{n_i t}$ is the flow of labor, and $n_i > 0$ is the constant growth rate of labor.[108] It is assumed that the population and labor force size coincide and that labor is supplied inelastically.

Following Acemoglu (2009, 435 and 461), the demand for intermediate good v can be calculated by maximizing the instantaneous profits of a representative final goods producer at time t.[109] The problem is

$$\begin{aligned} \max_{x_i(v,t)} \Pi_i(v,t) = {} & Q_i(t)^{\alpha-1} \int_0^{Q_i(t)} A_i(v,t) x_i(v,t)^{\alpha} L_i(t)^{1-\alpha}\, dv \\ & - \int_0^{Q_i(t)} p_i(v,t) x_i(v,t)\, dv - w_i(t) L_i(t). \end{aligned} \tag{3.2}$$

Applying the rule for differentiating under the integral sign, and solving the necessary condition for $p_i(v,t)$ leads to the inverse demand schedule for variety $v \in [0, Q_i(t)]$[110]

$$p_i(v,t) = \alpha A_i(v,t) l_i(t)^{1-\alpha} x_i(v,t)^{-(1-\alpha)}. \tag{3.3}$$

[108] The restriction that the labor growth rate is positive is not actually spelled out explicitly in Ertur and Koch (2011) though. However, since labor is an essential input in the production of the final good (for the proof, see, for example Barro and Sala-i-Martin (2004, 77-78)), the positive growth rate can be inferred. This assumption is maybe not as innocuous as it seems, in particular when it comes to testing the model's implications empirically. See also Footnote 149 in this context.

[109] Again this is not explicitly spelled out in Ertur and Koch (2011) either, but, in general, firms maximize the present discounted value of future profits. However, since firms rent the services of the input factors labor and capital (in the form of intermediates), and there are no adjustment costs, no dynamic constraints exist, and the intertemporal maximization problem becomes a static one (or more precisely, a sequence of static problems (see e.g. Barro and Sala-i-Martin (2004, 32) and Acemoglu (2009, 435))).

[110] See Appendix C.2.1 for the derivation.

Here, $l_i(t) \equiv \frac{L_i(t)}{Q_i(t)}$ denotes the number of workers per variety. With the help of results developed in Section 3.2.2.1, it can be shown that the production function in intensive form is given by[111]

$$\hat{y}_i(t) = \hat{k}_i(t)^{\alpha} \tag{3.4}$$

where $\hat{y}_i(t) \equiv \frac{Y_i(t)}{A_i(t)L_i(t)}$ is the output per effective worker, and $\hat{k}_i(v,t)$ is capital per effective worker.

Concerning the production function in Equation (3.1), it is important to note that the integral is multiplied by the factor $Q_i(t)^{\alpha-1}$. The factor is introduced in order to avoid that producers of the final good become increasingly more productive simply due to the availability of an increasing number of varieties. This effect, which can be interpreted as a form of technological progress has, for instance, been developed in the endogenous growth model by Romer (1990).[112]

The focus in the model described here is on technological interdependence (or more specifically, technology transfer) between regions. Hence, it is assumed that regions trade neither in goods nor in factors (Ertur and Koch, 2011, 220). Therefore, in general, the intermediate goods used and produced in region i as well as its final good are specific to this particular region. Nonetheless, due to technological interdependence, this is not the case for the process by which a specific intermediate good is produced. The respective idea for the production process might well have originated in a different region (Howitt, 2000, 831). The details of this idea will be provided in Section 3.3.

3.2.2 Intermediate Goods Sector

This section describes the production relations in the intermediate goods sector. It starts with the firms' optimization problem and illustrates the different assumptions underlying the generation of horizontal and vertical innovations. In general, horizontal innovations (or product innovations) increase the number of existing varieties, whereas vertical innovations (or process innovations) increase the productivity (quality) of an already existing variety.

3.2.2.1 Firms in the Intermediate Goods Sector

In the sector for intermediate goods, the production function for each monopolist in a given sector v is described by

$$x_i(v,t) = \frac{K_i(v,t)}{A_i(v,t)} \tag{3.5}$$

[111] See Appendix C.2.2 for the derivation.

[112] The absence of an effect of an increasing number of varieties on productivity in the model presented here is demonstrated at the end of Appendix C.2.2.

where $K_i(v,t)$ is the capital input in terms of the final good. From the functional form of the production function, it can be inferred that the production of varieties of higher quality becomes increasingly more capital intensive. This follows from the presence of the factor $A_i(v,t)$ in the denominator which rises with each new vintage of the good. In order to produce the intermediate good, the monopolist needs to rent capital at the price of $r_i(t)+\delta_i$ per unit, where $r_i(t)$ is the interest rate in region i and δ_i is the exogenously given region-specific depreciation rate. With this information, and, since $K_i(v,t) = A_i(v,t)x_i(v,t)$ from Equation (3.5), it follows that the monopolist's profit function is

$$\pi_i(v,t) = p_i(v,t)x_i(v,t) - [r_i(t)+\delta_i]\, A_i(v,t)x_i(v,t). \tag{3.6}$$

Solving the inverse demand function in Equation (3.3) for $x_i(v,t)$ leads to the direct demand function for intermediates

$$x_i(v,t) = [\alpha A_i(v,t)]^{\frac{1}{1-\alpha}}\, l_i(t)p_i(v,t)^{-\frac{1}{1-\alpha}}. \tag{3.7}$$

Hence, the profit maximization problem for the monopolist is given by the constrained optimization problem of maximizing the profits in Equation (3.6) subject to the demand function in Equation (3.7). Substituting the expression for $x_i(v,t)$ into the profit function above, leads to the unconstrained profit maximization problem of the monopolist

$$\begin{aligned}\max_{p_i(v,t)} \pi_i(v,t) = {} & p_i(v,t)\,[\alpha A_i(v,t)]^{\frac{1}{1-\alpha}}\, l_i(t)p_i(v,t)^{-\frac{1}{1-\alpha}} \\ & - [r_i(t)+\delta_i]\, A_i(v,t)\,[\alpha A_i(v,t)]^{\frac{1}{1-\alpha}}\, l_i(t)p_i(v,t)^{-\frac{1}{1-\alpha}}.\end{aligned}$$

Setting the derivative $\frac{\partial \pi_i(v,t)}{\partial p_i(v,t)}$ equal to zero, results in the necessary condition

$$\begin{aligned}& -\frac{\alpha}{1-\alpha}\,[\alpha A_i(v,t)]^{\frac{1}{1-\alpha}}\, l_i(t)p_i(v,t)^{-\frac{\alpha}{1-\alpha}-1} \\ & +\frac{1}{1-\alpha}\,[r_i(t)+\delta_i]\,\alpha^{\frac{1}{1-\alpha}} A_i(v,t)^{\frac{2-\alpha}{1-\alpha}} l_i(t)p_i(v,t)^{-\frac{1}{1-\alpha}-1} = 0\end{aligned}$$

and solving for the profit-maximizing price yields

$$p_i(v,t) = [r_i(t)+\delta_i]\,\frac{A_i(v,t)}{\alpha}. \tag{3.8}$$

Substituting this price into Equation (3.7) leads to

$$x_i(v,t) = \alpha^{\frac{2}{1-\alpha}} l_i(t)\,[r_i(t)+\delta_i]^{-\frac{1}{1-\alpha}}.$$

This result shows that the production of the intermediate good is independent of v (i.e. independent of the specific variety produced), and hence it holds that

$$x_i(v,t) = x_i(t), \tag{3.9}$$

implying that the equilibrium in the intermediate goods sector is symmetric so that independent of the specific variety v all monopolists produce the identical amount $x_i(t)$ of their respective variety.[113]

Noting that in equilibrium $x_i(t) = \hat{k}_i(t)l_i(t)$ holds,[114] it follows that the equilibrium interest rate is given by

$$r_i(t) = \alpha^2 \hat{k}_i(t)^{\alpha-1} - \delta_i. \tag{3.10}$$

Finally, using the profit-maximizing price in Equation (3.8), substituting the equilibrium interest rate and the expression for the quantity, $x_i(t) = \hat{k}_i(t)l_i(t)$, in the symmetric equilibrium into the profit function in Equation (3.6), implies that the monopolist's profits are given by

$$\pi_i(v,t) = A_i(v,t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t), \tag{3.11}$$

where the function $\tilde{\pi}_i\big(\hat{k}_i(t)\big)$ is defined as $\tilde{\pi}_i\big(\hat{k}_i(t)\big) \equiv \alpha(1-\alpha)\hat{k}_i(t)^{\alpha}$.

3.2.2.2 Horizontal Innovations in the Intermediate Goods Sector

The relevant assumption concerning horizontal innovations is that new varieties are created by imitation. Moreover, no resources are spent on this activity so that imitation is not a deliberate effort by individuals. As Aghion and Howitt (1998, 107) laconically put it: "imitation just happens". Hence, individuals in the economy can be sure that new varieties will enter the economy, but the specific point in time when a new intermediate good will be available for production of final output or when a new sector opens up in which to reap monopoly profits remains uncertain. Therefore, the occurrence of innovations is governed by a random process, and the specific random process assumed is a Poisson process. In more formal terms, each agent in region i imitates with equal likelihood, and

[113]Naturally, this also results, if one sets up the profit maximization problem with quantity as the decision variable (see, for example, Varian, 1992, 234) so that

$$\max_{x_i(v,t)} = \alpha A_i(v,t)x_i(v,t)^{\alpha-1}l_i(t)^{1-\alpha}x_i(v,t) - \left[r_i(t)+\delta_i\right]A_i(v,t)x_i(v,t).$$

Taking the derivative with respect to quantity, it follows that the marginal revenue and marginal cost function are proportional to $A_i(v,t)$, and, since this is the only difference between the firms producing an intermediate product, the symmetric equilibrium in Equation (3.9) follows (Howitt, 2000, 832).

[114]This result is derived in Appendix C.2.2 as an intermediate result in the derivation of the production function in intensive form.

her Poisson arrival rate[115] of imitation is given by $\xi > 0$, which is identical across regions. This implies that the aggregate flow of new intermediate goods is given by

$$\dot{Q}_i(t) = \xi L_i(t). \tag{3.12}$$

As Appendix C.2.3 demonstrates, the number of workers per variety $l_i(t)$ converges to

$$l_i = \frac{n_i}{\xi}, \tag{3.13}$$

which is independent of time t and thus constant.

3.2.3 Research and Development – Vertical Innovations

Apart from increases in the number of intermediate goods (horizontal innovations), a key characteristic of the model are quality, i.e. productivity, improvements of already existing intermediate products (vertical innovations). On a general level, quality improvements for a given variety result from investment in R&D in that particular sector.[116] Here, the final good is the relevant input factor. It is assumed that the inventor of a higher-quality variety in sector v at the same time also is the producer of this intermediate good.[117]

The mere fact of engaging in research activities naturally is no guarantee for success. As is standard in this type of models (see, for example, Aghion and Howitt, 1998, 54-55), the underlying random process for the occurrence of vertical innovations is also assumed to be a Poisson process. However, in this case, the Poisson arrival rate in any sector $v \in [0, Q_i(t)]$ is slightly more complicated as it is not given by a single parameter, but instead by the function

$$\phi_i(t) = \lambda_i \kappa_i(t)^{\phi}. \tag{3.14}$$

The variable $\kappa_i(t)$ denotes the sector-specific expenditures on vertical R&D adjusted for productivity, and the parameter ϕ, for which $0 \leq \phi \leq 1$ holds, gauges the strength of a given amount of R&D expenditures on λ_i (Ertur and Koch, 2011, 222). To be more precise with respect to R&D expenditures, these are given by $\kappa_i(t) = \frac{S_{A,i}(t)}{Q_i(t)A_i(t)^{max}}$, where $S_{A,i}(t)$ is the total input into R&D in region i, so that $\frac{S_{A,i}(t)}{Q_i(t)}$ reflects the total amount

[115] See Appendix C.1 for a primer on Poisson processes.

[116] The specific setup in the intermediate sector with imitation leading to new varieties and innovation to a higher quality of existing varieties was introduced by Young (1998), who formalized ideas expressed verbally in earlier work by Gilfillan (1935a,b). In this approach, a scale effect (i.e. a positive effect of population on the per capita growth rate), which was criticized by Jones (1995a,b) is not present. See also, Aghion and Howitt (1998, 106-110).

[117] This assumption is made for convenience. As Barro and Sala-i-Martin (2004, 290) state, results would be the same, if one alternatively assumed that inventors charged producers of intermediate goods a license fee for the use of the blueprint or process innovation.

invested in a given sector aggregated over all firms. $A_i(t)^{max}$ is the maximal value of $A_i(v,t)$ (or the leading-edge productivity parameter), and it is defined by

$$A_i(t)^{max} \equiv \max\{A_i(v,t); v \in [0, Q_i(t)]\}. \tag{3.15}$$

An important assumption is made concerning this parameter. Potential innovators all have immediate access to this technological knowledge and thus "all draw on the same pool" (Aghion and Howitt, 1998, 87-88).

Adjustment of the sector-specific resource investment by the leading-edge technology parameter captures the assumption of ever increasing complexity in the research process (Ertur and Koch, 2011, 222). With technology ever increasing, more and more resources need to be spent to prevent the rate of innovation from slowing down. In other words, "as technology advances, the resource cost of further advances increases proportionally" (Aghion and Howitt, 1998, 410). Note that, since the prospective payoffs from an innovation are identical across sectors, productivity-adjusted R&D investment, $\kappa_i(t)$, is also identical for each sector in region i.

Potential innovators face the questions of whether to conduct research at all, and if so how much to invest in R&D. Concerning these decisions, the value of an innovation to a successful innovator in a given sector is a critical variable. This value is given by

$$V_i(t) = \int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))ds} \pi_i(\tau)\, d\tau. \tag{3.16}$$

At some point in the future, a higher-quality variety will be invented in this sector, and the incumbent will be replaced by the successful innovator and lose his profits.[118] The equation above takes this into account and adjusts for it by including the Poisson arrival rate of new innovations in the discount factor.[119] Adjusted for productivity, the value of an innovation is defined as $v_i(t) \equiv \frac{V_i(t)}{A_i(t)^{max}}$ (Ertur and Koch, 2011, 222) so that

$$v_i(t) = \int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \frac{1}{A_i(\tau)^{max}} \pi_i(\tau)\, d\tau.$$

[118] Innovations will result from new entrants into the sector due to the Arrow replacement effect (Arrow, 1962). This effect states that incumbents who innovate would only replace part of their existing profits. On the other hand, researchers entering the sector have access to the leading-edge technology parameter, and, if they are successful, can reap the complete monopoly profits. Hence, these researchers have higher incentives to innovate than incumbents.

[119] A formal derivation of this value is provided in Appendix C.2.4.

Substituting for $\pi_i(\tau)$ from Equation (3.11) and noting that by assumption the productivity level of a firm that innovates at time t is at the leading edge, implies that $A_i(v,t) = A_i(t)^{max}$ in this case. Therefore[120]

$$\begin{aligned} v_i(t) &= \int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \frac{A_i(\tau)^{max}}{A_i(\tau)^{max}} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\, d\tau \\ &= \int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big) l_i(\tau)\, d\tau. \end{aligned}$$

Taking the derivative of this equation with respect to time, leads to the following research-arbitrage equation[121]

$$\frac{\dot{v}_i(t)}{v_i(t)} = r_i(t) + \phi_i(t) - \frac{l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)}{v_i(t)}. \tag{3.17}$$

Written in this form, the function in Equation (3.16) is also known as the (stationary) Hamilton-Jacobi-Bellman Equation (see, for example, Acemoglu, 2009, 245 and 462-463). Expressed equivalently as

$$r_i(t)v_i(t) = l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big) + \dot{v}_i(t) - \phi_i(t)v_i(t),$$

it shows that the required return on an innovation, $r_i(t)v_i(t)$, for a firm that engages in R&D, needs to equal its flow profits, $l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)$, plus any capital gains, $\dot{v}_i(t)$, adjusted for the fact that with positive probability $\phi_i(t)$ a new innovation occurs at some point in time, and the monopolist's product thus becomes obsolete from this point onwards.

An individual considering conducting R&D with the aim of improving a particular variety v has expected profits $\pi^e_{A,i}$. In particular,

$$\pi^e_{A,i} = \lambda_i \kappa_i(t)^\phi \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} \cdot V_i(t) + (1 - \lambda_i\kappa_i(t)^\phi) \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} \cdot 0 - S_{A,i}(v,t). \tag{3.18}$$

Here, $\lambda_i\kappa_i(t)^\phi$ is the probability of being successful in research, and $1 - \lambda_i\kappa_i(t)^\phi$ is the complementary probability of failure in research. $S_{A,i}(v,t)$ denotes how many resources the firm invests in R&D, and the division by $S_{A,i}(t)/Q_i(t)$ captures negative externalities in the research process. More precisely, overlap and duplication of research efforts are underlying this assumption (Ertur and Koch, 2011, 222). Hence, there is no linear increase in profits with resources invested in R&D. Note that the R&D technology requires only output as an input.[122] In other words, only laboratory equipment is required to engage

[120] In the article by Ertur and Koch the dependence of the number of workers per variety l_i on τ is missing.

[121] This derivation involves applying Leibniz's Formula, and the detailed steps are provided in Appendix C.2.5.

[122] The price of these resources is normalized to 1 as they are measured in units of the output good, which is the numéraire in this model.

in research activities, but no workers or scientists need to be employed. Therefore, this model is a variant of a "lab-equipment" model (see, for instance, Acemoglu, 2009, 433).

Incumbent firms in the R&D sector then face the following profit-maximization problem (this follows from simplifying Equation (3.18) and dropping the superscript for expectations to enhance readability)

$$\max_{S_{A,i}(v,t)} \pi_{A,i}(v) = \lambda_i \kappa_i(t)^{\phi} \frac{S_{A,i}(v,t)}{S_{A,i}(t)/Q_i(t)} V_i(t) - S_{A,i}(v,t).$$

The necessary condition therefore reads

$$\frac{\partial \pi_{A,i}(v)}{\partial S_{A,i}(v,t)} = 0 \qquad \Longleftrightarrow \qquad \frac{\lambda_i \kappa_i(t)^{\phi}}{S_{A,i}(t)/Q_i(t)} V_i(t) = 1.$$

By substituting $V_i(t) = v_i(t) A_i(t)^{max}$ and using the definition of $\kappa_i(t)$ in this condition, it follows that the value of an innovation is given by

$$v_i(t) = \frac{1}{\lambda_i} \kappa_i(t)^{1-\phi}.$$

Solving for $\kappa_i(t)$ and log-differentiating the resulting expression yields $\frac{\dot{\kappa}_i(t)}{\kappa_i(t)} = \frac{1}{1-\phi} \frac{\dot{v}_i(t)}{v_i(t)}$. Substituting thereafter from the research-arbitrage equation in (3.17) and then inserting the expression for the Poisson arrival rate from (3.14) leads to the following differential equation

$$\frac{\dot{\kappa}_i(t)}{\kappa_i(t)} = \frac{1}{1-\phi} \left[r_i(t) + \lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi-1} l_i(t) \tilde{\pi}_i\big(\hat{k}_i(t)\big) \right]. \tag{3.19}$$

This equation describes how the resources invested in R&D (measured in terms of the final good) evolve over time.

In the derivation of this expression, the leading-edge productivity parameter $A_i(t)^{max}$ has been used. As innovations result in knowledge spillovers, this parameter is not constant. In particular, its growth rate and thereby the growth rate of technological progress is equal to

$$g_i(t) \equiv \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} = \frac{\sigma}{Q_i(t)} Q_i(t) \lambda_i \kappa_i(t)^{\phi} = \sigma \lambda_i \kappa_i(t)^{\phi}. \tag{3.20}$$

Basically, therefore, $A_i(t)^{max}$ grows with the aggregate rate of innovations (i.e. the Poisson arrival rate from Equation (3.14) times the number of differentiated varieties $Q_i(t)$) multiplied by a factor of proportionality $\sigma/Q_i(t) > 0$. This factor captures by how much public knowledge increases as a result of an additional innovation or, expressed differently, it measures "the marginal impact of each innovation on the stock of public knowledge" (Aghion and Howitt, 1998, 411). However, this impact is diminishing in $Q_i(t)$. Over time, horizontal innovations lead to an increase in the number of intermediates, and the division of the factor of proportionality by this number ensures that innovations of a given size for a particular product, will have a diminishing impact (Ertur and Koch, 2011, 223).

Having determined the growth rate for the leading-edge productivity parameter, it is helpful for subsequent derivations to look at the corresponding growth rate for the average productivity parameter, $A_i(t)$. In general, a successful innovation for intermediate good v changes productivity for this good from $A_i(v,t)$ to $A_i(t)^{max}$.[123] Across innovating sectors, the average increase from a successful innovation is given by $A_i(t)^{max} - A_i(t)$. Taking into account that innovations are generated with rate $\lambda_i \kappa_i(t)^\phi$ uniformly across all sectors, and that average productivity remains unaffected by horizontal innovations, it follows that the change in average productivity can be expressed as

$$\dot{A}_i(t) = \lambda_i \kappa_i(t)^\phi \big(A_i(t)^{max} - A_i(t)\big).$$

Appendix C.2.7 demonstrates that the ratio of the leading-edge productivity parameter to the average productivity parameter converges to the constant $1+\sigma$ so that $A_i(t)^{max} = (1+\sigma)A_i(t) \, \forall \, t$, implying that the growth rates of both variables will be identical.

3.2.4 Physical Capital Accumulation and Steady State

As in a standard neoclassical Solow model, the accumulation of physical capital is governed by the general equation

$$\dot{\hat{k}}_i(t) = s_{K,i}\hat{k}_i(t)^\alpha - \big(n_i + g_i(t) + \delta_i\big)\hat{k}_i(t). \tag{3.21}$$

Here, $s_{K,i}$ denotes the investment rate for physical capital in region i and δ_i signifies the depreciation rate for physical capital, which is region-specific. The evolution of the economy can then be described by the following system of differential equations:

$$\begin{aligned} \dot{\hat{k}}_i(t) &= s_{K,i}\hat{k}_i(t)^\alpha - \big(n_i + g_i(t) + \delta_i\big)\hat{k}_i(t) \\ \dot{\kappa}_i(t) &= \frac{\kappa_i(t)}{1-\phi}\Big[r_i(t) + \lambda_i\kappa_i(t)^\phi - \lambda_i\kappa_i(t)^{\phi-1} l_i(t)\tilde{\pi}_i\big(\hat{k}_i(t)\big)\Big] \end{aligned}$$

where the first equation follows from Equation (3.21) by inserting for the growth rate from Equation (3.20), and the second equation above is just Equation (3.19) multiplied by $\kappa_i(t)$. In steady state, capital in efficiency units and productivity-adjusted R&D investment are constant so that $\dot{\hat{k}}_i(t) = \dot{\kappa}_i(t) = 0$. Imposing this condition and denoting steady-state values with an asterisk, implies that the steady-state rate of technological progress in

[123] One might wonder about the distribution of productivities across sectors in this model. Appendix C.2.6 demonstrates that the relative productivities $a_i(v,t) = A_i(v,t)/A_i(t)^{max}$ converge to an invariant distribution, meaning that even though $A_i(t)^{max}$ increases over time and sectors change position in the distribution, its shape remains constant in the long run (Aghion and Howitt, 1992, 88).

region i is given by $g_i^* = \sigma\lambda_i\left(\kappa_i^*\right)^\phi$ and that the steady-state value for $\hat{k}_i^*$ is defined by the $\dot{\hat{k}}_i(t) = 0$-isocline as[124]

$$\hat{k}_i^* = \left(\frac{s_{K,i}}{n_i + \sigma\lambda_i\left(\kappa_i^*\right)^\phi + \delta_i}\right)^{\frac{1}{1-\alpha}}. \tag{3.22}$$

This isocline is depicted as the downward-sloping curve (I) in $(\kappa_i(t) - \hat{k}_i(t))$-space in the upper right hand in Figure 3.3. From setting $\dot{\kappa}_i(t) = 0$, it follows that the $\dot{\hat{k}}_i(t) = 0$-isocline is given by

$$1 = \lambda_i\big(\kappa_i^*\big)^{\phi-1}\frac{\tilde{\pi}_i(\hat{k}_i^*)l_i}{r_i^* + \lambda_i\big(\kappa_i^*\big)^\phi}.$$

This relation is the upward-sloping schedule labeled (II) in Figure 3.3.

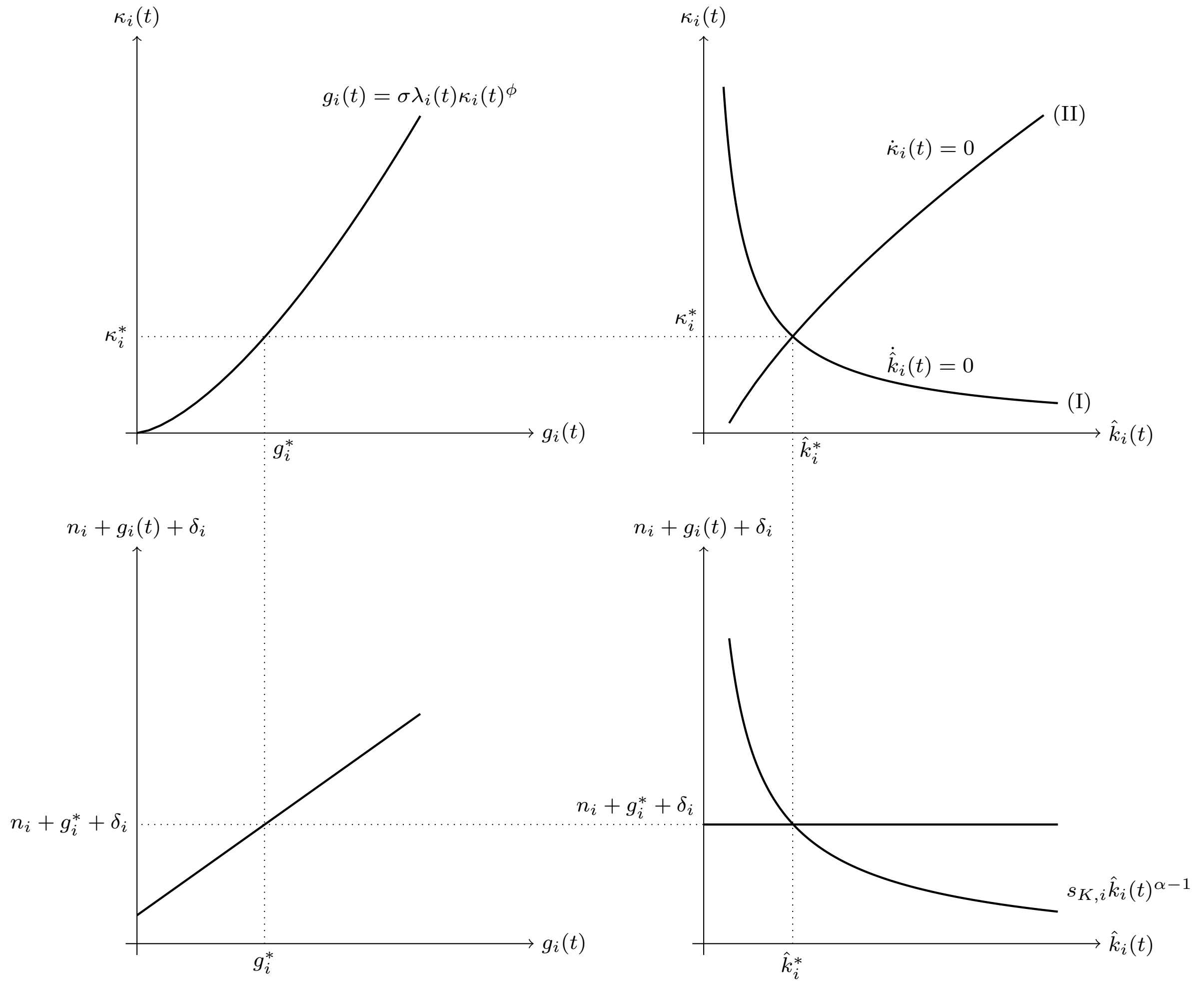


Figure 3.3: Illustration of the Steady State (Adapted from Ertur and Koch (2011)).

Curve (I) is downward sloping as in steady state an increase in R&D investment leads to an increase in the growth rate g_i^*. From Equation (3.22), it then follows that for

[124]There seems to be a typo in the corresponding Equation (22) in Ertur and Koch (2011), where the left-hand side should read $(\hat{k}_i^*)^{1-\alpha}$ instead of $(\hat{k}_i^*)^\alpha$.

equilibrium to be maintained the capital-output ratio, $\frac{\hat{k}^*}{\hat{y}^*} = (\hat{k}^*)^{1-\alpha}$, needs to fall. On the other hand, Curve (II) is upward sloping, since when $\hat{k}_i^*$ increases, the interest rate in steady state falls (compare Equation (3.10)) and profits increase (see Equation (3.11)). Hence, in equilibrium R&D expenditures need to rise.

Turning now to the remaining parts of Figure 3.3, the lower right one shows the Solow diagram as, for example, in Barro and Sala-i-Martin (2004, 56). The main difference to the standard version is that here the effective depreciation rate, $n_i + g_i(t) + \delta_i$, through its dependence on the rate of technological progress, $g_i(t)$, is endogenously determined by investment in R&D and thus moves up until the steady state is reached (Ertur and Koch, 2011, 224). This determination of $g_i(t)$ through $\kappa_i(t)$ is depicted in the upper left part of the figure, whereas the positive dependence of the effective depreciation rate on technological progress is depicted in the lower left part of the figure. In steady state, with $g_i(t) = g_i^*$, the effective depreciation rate is constant, which allows for determining the level of physical capital per effective worker and the level of R&D investment via the dotted lines.

3.3 Multi-Region Schumpeterian Growth Model with Technological Interdependence

This section introduces the analytical setup in which diffusion of knowledge depends on a region's gap to its own technological frontier. In addition, the steady-state equation on which the estimation will be based, is derived.

3.3.1 Research Productivity, Knowledge Spillovers, and Technology Gap

Turning now to the case of multiple regions, the assumption that all regions develop independently from each other is abandoned. Interdependence enters the model via the assumption that the productivity in the research sector, λ_i, in region i depends on its own level of technology relative to the level of other regions as well as on the way the connection between regions is modeled. In formal terms, the region-specific research productivity is given by

$$\lambda_i = \lambda \prod_{j=1}^{N} \left(\frac{A_j(t)}{A_i(t)} \right)^{\gamma_i v_{ij}} . \tag{3.23}$$

Note that the technology frontier is specific to each region due to the presence of the parameters v_{ij}. Concerning these, it is assumed that they are non-negative, finite and non-stochastic. Moreover, $\sum_{j=1}^{N} v_{ij} = 1$ is assumed. In general, not all regions necessarily are equally able to increase their research productivity due to a given increase in knowledge in the regions it is connected to. In this regard, the absorptive capacity of a region plays an important role.[125] This notion is picked up by Ertur and Koch (2011) in the parameter γ_i, as it is assumed that the absorption capacity depends on the human capital stock, H_i, in region i in the following way: $\gamma_i = \gamma H_i$, where $\gamma < 1$ is a measure of the amount of knowledge spilling over from other regions. At this point, the derivation in Appendix C.2.7, which demonstrates that the growth rates of the leading-edge productivity parameter and the average productivity parameter are identical, becomes helpful. Substituting the expression for λ_i into Equation (3.20), and using that $g_i(t) \equiv \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} = \frac{\dot{A}_i(t)}{A_i(t)}$ leads to

$$g_i(t) \equiv \sigma\lambda\kappa_i(t)^{\phi} \prod_{j=1}^{N} \left(\frac{A_j(t)}{A_i(t)}\right)^{\gamma_i v_{ij}}.$$

The last term in this equation represents the distance to the technological frontier for region i. This implies that the further away a region is from its own technology frontier, i.e. the larger is the average technological level in the regions it is connected to or the lower is its own level of technology, the higher is its productivity in the research sector. The intuition is that there exists a large pool of knowledge in the region's environment into which it has not yet tapped into. Spillovers from other regions or equivalently spatial externalities are comparatively large in this case.[126] Conversely, a region close to its technological frontier cannot benefit from spillovers or technology diffusion from connected regions in the same extent as the pool of knowledge has been largely tapped out and copying "foreign" technology becomes more difficult (Ertur and Koch, 2011, 226).[127]

Since in steady state $\hat{k}_i$ and κ_i grow at constant rates in each region, it follows that a region's distance to its own technological frontier remains constant. However, for steady state to occur this requires that all regions grow at identical rates or, expressed differently, converge to parallel growth paths in the long run. This steady state growth rate for regions $i = 1, \ldots, N$ is given by

$$g^w \equiv \sigma\lambda\kappa_i^{\phi} \prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}} \tag{3.24}$$

where time dependence t as well as the asterisks indicating steady-state values have been dropped to enhance readability. Regions converge to the same growth rate in the long run due to the inverse relation between how many resources are invested in the research sector

[125] This corresponds to ideas developed in Nelson and Phelps (1966), although the specific word "absorptive capacity" is not mentioned by them.

[126] As Ertur and Koch (2011, 217-218) point out, this is the concept of the "advantage of backwardness" by Gerschenkron (1962).

[127] These effects are similar to the effects of "standing on the shoulders of giants" (compare Caballero and Jaffe, 1993) and "fishing out" (see Jones, 1995a, 765) mentioned in the literature on endogenous growth models with respect to the research productivity in a single country.

and this sector's productivity in steady state. Investing a comparatively large amount of resources in the research sector so that κ_i is relatively high, implies that the level of technology will in turn also be relatively high. From Equation (3.23) it then follows that the ratio of the average level of technology to the own level of technology will be comparatively low, i.e. a region is close to its own technology frontier, which implies that research productivity λ_i in turn will be relatively low, too. A region with comparatively low R&D expenditures has a relatively high research productivity due to the large distance to its own technology frontier and as Ertur and Koch (2011, 226) note, due to technology diffusion and its impact on research productivity, convergence to the steady state growth rate occurs.

In order to test the model empirically in the following section, Equation (3.24) will be rewritten. As an intermediate step note that the productivity-adjusted sector-specific expenditures into R&D are given by $\kappa_i(t) = \frac{S_{A,i}(t)}{Q_i(t)A_i(t)^{max}}$. Multiplying and dividing this expression by $\frac{Y_i}{L_i}$ and using $A_i(t)^{max} = (1+\sigma)A_i(t)$ from Appendix C.2.7 leads to $\kappa_i = \frac{S_{A,i}}{Y_i}\frac{Y_i}{L_i}\frac{L_i}{Q_i}\frac{1}{(1+\sigma)A_i}$. With the result from Equation (3.13), this can be equivalently expressed as

$$\kappa_i = s_{A,i} y_i \frac{n_i}{\xi} \frac{1}{(1+\sigma)A_i}. \tag{3.25}$$

Here, the definition $s_{A,i} \equiv \frac{S_{A,i}}{Y_i}$ for the investment rate in the research sector has been applied. The global technology growth rate can then be shown to be given by the expression[128]

$$g^w = \frac{\sigma\lambda}{[(1+\sigma)\xi]^{\phi}} s_{A,i}^{\phi} y_i^{\phi} n_i^{\phi} A_i^{-\phi-1} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}}. \tag{3.26}$$

Applying the natural logarithm to this equation and then solving for $\ln A_i$ yields

$$\ln A_i = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} + \frac{\phi}{1+\phi} (\ln s_{A,i} + \ln n_i + \ln y_i) + \frac{\gamma H_i}{1+\phi} \sum_{j\neq i}^{N} v_{ij} \ln A_j.$$

[128] The derivation might not be immediately obvious and is therefore given in Appendix C.2.8.

Stacking the equations for regions $i = 1, \ldots, N$, the level of technology is given by

$$\underbrace{\begin{pmatrix} \ln A_{1t} \\ \vdots \\ \ln A_{Nt} \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}} = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota} + \frac{\phi}{1+\phi} \underbrace{\begin{pmatrix} \ln s_{A,1} + \ln n_1 + \ln y_1 \\ \vdots \\ \ln s_{A,N} + \ln n_N + \ln y_N \end{pmatrix}}_{\substack{=\boldsymbol{s_A}+\boldsymbol{n}+\boldsymbol{y} \\ (N\times 1)}}$$

$$+ \frac{\gamma}{1+\phi} \underbrace{\begin{pmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & H_N \end{pmatrix}}_{\substack{=\boldsymbol{H}=diag(H_i) \\ (N\times N)}} \underbrace{\begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{pmatrix}}_{\substack{=\boldsymbol{V} \\ (N\times N)}} \underbrace{\begin{pmatrix} \ln A_1 \\ \vdots \\ \ln A_N \end{pmatrix}}_{\substack{=\boldsymbol{A} \\ (N\times 1)}}$$

Defining $\boldsymbol{W} \equiv \boldsymbol{H} \cdot \boldsymbol{V}$ with entries $v_{ii} = 0$ if $i = j$ in $\boldsymbol{V}$, the equivalent matrix expression for the level of technology is

$$\boldsymbol{A} = \frac{1}{1+\phi} \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota} + \frac{\phi}{1+\phi} (\boldsymbol{s_A} + \boldsymbol{n} + \boldsymbol{y}) + \frac{\gamma}{1+\phi} \boldsymbol{WA}. \tag{3.27}$$

Given that the matrix $\left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)$ is non singular and thus has an inverse,[129] Equation (3.27) can be solved for $\boldsymbol{A}$ to yield a matrix equation for the level of technology

$$\begin{aligned} \boldsymbol{A} = & \frac{1}{1+\phi} \left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)^{-1} \left(\ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} \boldsymbol{\iota}\right) \\ & + \frac{\phi}{1+\phi} \left(\boldsymbol{I} - \frac{\gamma}{1+\phi}\boldsymbol{W}\right)^{-1} (\boldsymbol{s_A} + \boldsymbol{n} + \boldsymbol{y}). \end{aligned} \tag{3.28}$$

3.3.2 Income per Worker in Steady State

In this section an expression that determines the income per worker in steady state will be derived. From Equation (3.4) it follows that the production function per worker in steady state for region i is given by $y_i^* = A(\hat{k}_i^*)^{\alpha}$. Substituting for the steady-state level of capital in efficiency units leads to

$$y_i^* = A \left(\frac{s_{K,i}}{n_i + \sigma\lambda_i \left(\kappa_i^*\right)^{\phi}} + \delta_i \right)^{\frac{1}{1-\alpha}}.$$

[129] An application of Gerschgorin's Theorem (see Gerschgorin, 1931) ensures that. See Appendix C.2.9 for a similar case.

After taking the natural logarithm and stacking the expressions for regions $i = 1, \ldots, N$, the steady-state incomes in per worker terms can be expressed in the following matrix equation: $\boldsymbol{y} = \boldsymbol{A} + \frac{\alpha}{1-\alpha}\boldsymbol{s_K}$ in which the matrix $\boldsymbol{s_K}$ is an $N \times 1$ matrix with the terms $\frac{s_{K,i}}{n_i+g^w+\delta_i}$ for the respective regions. Inserting the result for $\boldsymbol{A}$ from Equation (3.28) into the expression above, yields

$$\boldsymbol{y} = \left(\ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}}\right)\boldsymbol{\iota} + \phi(\boldsymbol{s_A} + \boldsymbol{n}) + \frac{\alpha(1+\phi)}{1-\alpha}\boldsymbol{s_K} - \frac{\alpha\gamma}{1-\alpha}\boldsymbol{W}\boldsymbol{s_K} + \gamma\boldsymbol{W}\boldsymbol{y}. \tag{3.29}$$

Writing this equation for an individual region i clarifies the determinants of the level of per worker income in steady state

$$\begin{aligned}\ln y_i =\ & \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^{\phi}} + \phi(\ln s_{A,i} + \ln n_i) + \frac{\alpha(1+\phi)}{1-\alpha}\ln \frac{s_{K,i}}{n_i+g^w+\delta_i} \\ & - \frac{\alpha\gamma H_i}{1-\alpha}\sum_{j\neq i}^{N} v_{ij}\ln \frac{s_{K,j}}{n_j+g^w+\delta_j} + \gamma H_i \sum_{j\neq i}^{N} v_{ij}\ln y_j .\end{aligned} \tag{3.30}$$

It is important to note here that a change in the independent variables in region i affects the steady-state levels in the regions to which it is connected, and the steady-state levels in neighboring regions in turn have an influence on the respective level in region i. Therefore, studying the effect of, for example, a change in the investment rate in physical capital requires an analysis of the complete interdependent system in Equation (3.29). In general, the impact of a change in one of the independent variables can be divided into two parts. The first one represents the impact on the income per worker in steady state in region i due to a change in the independent variable in this region, and the second one details the effect of an identical change in the same variable in all regions $j = 1, \ldots, N$ with $j \neq i$ that region i is connected to. For example, the $N \times N$ matrix of income per worker elasticities with respect to the R&D investment rate $\boldsymbol{s_A}$, is given by

$$\boldsymbol{\eta}^{\boldsymbol{s_A}} \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{s_A}} = \phi(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1} = \phi\boldsymbol{I} + \phi\sum_{r=1}^{\infty}\gamma^r\boldsymbol{W}^r. \tag{3.31}$$

This result is obtained by solving Equation (3.29) for $\boldsymbol{y}$ and then differentiating the result with respect to $\boldsymbol{s_A}$.[130] Concerning the last equality, it follows as the inverse $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$ is given by the Neumann series $\sum_{r=0}^{\infty}\gamma^r\boldsymbol{W}^r$ (see, for instance, Meyer (2000, 126 and 618) for this result) so that

$$(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1} = \boldsymbol{I} + \gamma\boldsymbol{W} + \gamma^2\boldsymbol{W}^2 + \cdots + \gamma^r\boldsymbol{W}^r + \cdots = \sum_{r=0}^{\infty}\gamma^r\boldsymbol{W}^r. \tag{3.32}$$

[130] Naturally, this derivation is only valid given that the inverse $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$ exists. Appendix C.2.9 provides the conditions under which this inverse exists.

This series is also called the spatial multiplier.[131] With respect to the elasticity in Equation (3.31), it highlights that changes in R&D investment in a given region i will have an impact on income per worker in all other locations. Hence, the total effect can be decomposed into the two impacts described above.

The first effect is given by[132]

$$\eta_i^{s_{A,i}} = \phi + \phi \sum_{r=1}^{\infty} \gamma_i^r v_{ii}^{(r)} > 0 \tag{3.33}$$

where $v_{ii}^{(r)}$ denotes the element i in row i and column i of the matrix $\boldsymbol{V}$ taken to the power of r. The second effect, the impact on region i of a change in R&D expenditures in the regions it is connected to, is

$$\eta_i^{s_{A,j}} = \phi \sum_{r=1}^{\infty} \gamma_i^r v_{ij}^{(r)} > 0. \tag{3.34}$$

In a similar manner, the aggregate effect of changes in the physical capital investment rate can be derived to yield

$$\boldsymbol{\eta}^{s_K} \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{s_K}} = \frac{\alpha}{1-\alpha}\boldsymbol{I} + \frac{\alpha\phi}{1-\alpha}(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1} = \frac{\alpha(1+\phi)}{1-\alpha}\boldsymbol{I} + \frac{\alpha\phi}{1-\alpha}\sum_{r=1}^{\infty}\gamma^r \boldsymbol{W}^r \tag{3.35}$$

which is positive, as knowledge diffuses across regions. For the employment growth rate, the corresponding elasticity is given by

$$\begin{aligned}\boldsymbol{\eta}^{n} \equiv \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{n}} = & -\frac{\alpha}{1-\alpha}\mathbf{diag}\left(\frac{\boldsymbol{n}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right) + \frac{\alpha\phi}{1-\alpha}\mathbf{diag}\left(\frac{\boldsymbol{g}+\boldsymbol{\delta}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right) \\ & + \frac{\alpha\phi}{1-\alpha}\sum_{r=1}^{\infty}\gamma^r \boldsymbol{W}^r \mathbf{diag}\left(\frac{\boldsymbol{g}+\boldsymbol{\delta}}{\boldsymbol{n}+\boldsymbol{g}+\boldsymbol{\delta}}\right).\end{aligned} \tag{3.36}$$

This elasticity captures that on the one hand, per worker income is positively influenced by increases in the employment growth rate, as this leads to a larger number of horizontally

[131] See e.g. Ertur and Koch (2011, 232), Elhorst (2010, 21-22), or LeSage and Pace (2014) on this expression.

[132] Note that even though the entries v_{ii} in the matrix $\boldsymbol{V}$ might be zero, this is not necessarily the case for entries in the corresponding matrix raised to a higher order as the following counterexample shows:

$$\boldsymbol{V}\boldsymbol{V} = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} = \begin{pmatrix} 4/9 & 1/9 & 3/9 & 1/9 \\ 1/6 & 1/3 & 1/6 & 1/3 \\ 1/3 & 1/9 & 4/9 & 1/9 \\ 1/6 & 1/3 & 1/6 & 1/3 \end{pmatrix} = \boldsymbol{V}^2$$

Therefore, the matrix $\boldsymbol{V}$ is not idempotent. Economically, this effect can be understood as knowledge spilling over from region i to region j from where a spillover originates back to region i. In other words, feedback effects exist in this model.

differentiated products on which R&D can be conducted, and it captures that on the other hand, a negative impact exists, which results from the dilution of physical capital (Ertur and Koch, 2011, 250).[133]

3.4 Empirical Specification and Estimation Method

This section describes the empirical specification of the model and details the econometric estimation method. In particular, the derivation of the log-likelihood function and its concentrated version will be discussed in detail.

3.4.1 Empirical Specification

From the expression for the steady-state level of income per worker in Equation (3.30), the following empirical counterpart in reduced form can be derived[134]

$$\begin{aligned} \ln y_i = {} & \beta_0 + \beta_1 \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \beta_2 \ln s_{A,i} + \beta_3 \ln n_i \\ & + \theta H_i \sum_{j \neq i}^{N} v_{ij} \ln \frac{s_{K,j}}{n_j + g^w + \delta_j} + \gamma H_i \sum_{j \neq i}^{N} v_{ij} \ln y_j + \varepsilon_i. \end{aligned} \tag{3.37}$$

In this equation, the parameters are given by the following expressions $\beta_0 \equiv \ln \frac{\sigma\lambda}{g^w[(1+\sigma)\xi]^\phi} > 0$, $\beta_1 = \frac{\alpha(1+\phi)}{1-\alpha} > 0$, $\beta_2 = \beta_3 = \phi > 0$, and $\theta = -\frac{\alpha\gamma}{1-\alpha} < 0$. The error term or region-specific shock, ε_i, is assumed to be identically and independently distributed (iid) for $i = 1, \ldots, N$. Accounting for the interdependence between regions, the equation above can be rewritten in matrix form as

$$\boldsymbol{y} = \boldsymbol{\iota}\beta_0 + \boldsymbol{X\beta} + \theta \boldsymbol{WZ} + \gamma \boldsymbol{Wy} + \boldsymbol{\varepsilon}. \tag{3.38}$$

This specification is a Spatial Durbin Model (SDM) as it includes spatial lags of the exogenous as well as endogenous variables (LeSage and Pace, 2009).[135] The list below provides an overview of variable definitions in this specification:

$\boldsymbol{y}$ is an $N \times 1$ vector of the natural logarithm of real income per worker,

$\boldsymbol{\iota}$ is an $N \times 1$ vector of ones,

β_0 is a scalar,

[133] The two different diagonal matrices in Equation (3.36) are both of dimension $N \times N$, and their general terms are given by $\frac{n_i}{n_i+g^w+\delta_i}$ and $\frac{g^w+\delta_i}{n_i+g^w+\delta_i}$, respectively, with $i = 1, \ldots, N$ (Ertur and Koch, 2011, 250).

[134] For simplicity, the time index has been set to $t = 0$ and is omitted.

[135] To be more precise, Equation (3.38) is a constrained version of the standard Spatial Durbin Model, since in this case only a subset of the potential spatial lags of the exogenous variables is included.

$\boldsymbol{X}$ is an $N \times 3$ matrix of the explanatory variables (the investment rate in physical capital, $s_{K,i}$, divided by the effective depreciation rate, $n_i + g^w + \delta_i$, the growth rate of the number of workers, n_i, and the investment rate in R&D, $s_{A,i}$ – all in logs),

$\boldsymbol{\beta}$ is a 3×1 vector $[\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)']$ of the regression parameters for the explanatory variables,

θ is a scalar,

$\boldsymbol{W}$ is the $N \times N$ interaction matrix (or spatial weight matrix) in non row-normalized form,

$\boldsymbol{Z}$ is the $N \times 1$ vector of the investment rate in physical capital divided by the effective depreciation rate,

$\boldsymbol{WZ}$ is the $N \times 1$ vector of the spatial lag of the investment rate in physical capital divided by the effective depreciation rate,

γ is the spatial autoregressive coefficient,

$\boldsymbol{Wy}$ is an $N \times 1$ vector denoting the spatial lag of the endogenous variable,

$\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of errors with mean zero and variance $\sigma^2 \boldsymbol{I}$ so that $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ holds.

The model specified in Equation (3.38) nests a series of growth models as special cases of the multi-region Schumpeterian growth model. For instance, the familiar Solow model (see, for example, the original contributions by Solow (1956) and Swan (1956)) is a special case of Equation (3.37). It results when no interaction (or technological interdependence) between regions exists and consequently $\gamma = 0$ (compare Equation (3.23)). Furthermore, in the standard Solow model, R&D expenditures are not present, which implies $\phi = 0$.[136] With these conditions, it follows from Equation (3.37) that in this case steady-state income per worker is given by

$$\ln y_i = \beta_0^S + \beta_1^S \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \varepsilon_i^S. \tag{3.39}$$

[136] There exist extensions of the model, which include this variable. See, for example, Nonneman and Vanhoudt (1996) or Keller and Poutvaara (2005). Additional augmentations of the standard Solow model have been developed, too. These include extending the model by human capital (Mankiw et al., 1992), by health (Knowles and Owen, 1995), by IQ and longevity Ram (2007), or by history (Dalgaard and Strulik, 2013). These models are, however, not nested in the multi-region Schumpeterian model discussed here and hence not estimated.

Written in matrix form, this is equivalent to $\boldsymbol{y} = \beta_0 \boldsymbol{\iota} + \beta_1^S \boldsymbol{X^S} + \boldsymbol{\varepsilon^S}$ with $\boldsymbol{X^S}$ an $N \times 1$ vector of the investment rate in physical capital divided by the effective depreciation rate, β_1^S the corresponding regression parameter, and $\boldsymbol{\varepsilon^S}$ an iid vector for the error terms.

Next, the Schumpeterian model by Howitt (2000) and Aghion and Howitt (1998) is also a special case of the multi-region Schumpeterian model as these authors abstain from modeling spillovers due to investment in physical capital (implying $\theta = 0$) and assume that the amount of knowledge that diffuses to other regions is identical for all regions (Howitt, 2000, 838). Hence, if the amount of knowledge diffusion is independent of the specific region, the term $\gamma H_i \sum_{j \neq i}^N v_{ij} \ln y_j$ in Equation (3.37) can be subsumed into the constant of the empirical specification. The result then is

$$\ln y_i = \beta_0^H + \beta_1^H \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \beta_2^H \ln s_{A,i} + \beta_3^H \ln n_i + \varepsilon_i^H \tag{3.40}$$

which in matrix form reads $\boldsymbol{y} = \beta_0^H \boldsymbol{\iota} + \boldsymbol{X^H \beta^H} + \boldsymbol{\varepsilon^H}$ with $\boldsymbol{X^H}$ an $N \times 3$ matrix of the regressors specified in the equation above and $\boldsymbol{\beta^H}$ the 3×1 vector of corresponding coefficients. The error in this specifications is also iid.

Finally, given that R&D investment has no impact on the Poisson arrival rate and thus $\phi = 0$, it follows that $\beta_2 = 0 = \beta_3$ in Equation (3.37), and the resulting model is the spatially augmented Solow model developed in Ertur and Koch (2007).[137] Formally, this specification is given by

$$\begin{aligned} \ln y_i = \beta_0^{EK} + \beta_1^{EK} \ln \frac{s_{K,i}}{n_i + g^w + \delta_i} + \theta^{EK} H_i \sum_{j \neq i}^N v_{ij} \ln \frac{s_{K,j}}{n_j + g^w + \delta_j} \\ + \gamma^{EK} H_i \sum_{j \neq i}^N v_{ij} \ln y_j + \varepsilon_i^{EK}. \end{aligned} \tag{3.41}$$

which is a Spatial Durbin Model. In matrix notation, it is given as $\boldsymbol{y} = \beta_0^{EK} \boldsymbol{\iota} + \beta^{EK} \boldsymbol{X^{EK}} + \theta^{EK} \boldsymbol{W X^{EK}} + \gamma^{EK} \boldsymbol{W y} + \boldsymbol{\varepsilon^{EK}}$ with $\boldsymbol{X}^{EK}$ an $N \times 1$ vector of the values for the investment rate in physical capital divided by the effective depreciation rate, $\boldsymbol{W X^{EK}}$ the spatial lag of this variable, $\boldsymbol{W y}$ the spatial lag of the dependent variable, and $\boldsymbol{\varepsilon^{EK}}$ the iid error term.

[137] The model presented in Equation (3.41) differs from from the one in Ertur and Koch (2007) with respect to the interaction matrix $\boldsymbol{W}$, as in their contribution the matrix of the human capital stock, $\boldsymbol{H}$, is absent.

3.4.2 Estimation Strategy

As LeSage and Pace (2009) point out, a Spatial Durbin Model can be equivalently expressed as a Spatial Autoregressive Model (SAR). Rewriting Equation (3.38) accordingly, leads to

$$\boldsymbol{y} = \gamma \boldsymbol{W}\boldsymbol{y} + \tilde{\boldsymbol{X}}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{3.42}$$

with $\tilde{\boldsymbol{X}} = [\boldsymbol{\iota}\, \boldsymbol{X}\, \boldsymbol{W}\boldsymbol{Z}]$ an $N \times 5$ matrix and $\boldsymbol{\delta} = [\beta_0\, \boldsymbol{\beta}\, \theta]'$ a 5×1 vector. In reduced form, this model is therefore given by[138]

$$\boldsymbol{y} = (\boldsymbol{I} - \gamma \boldsymbol{W})^{-1} \tilde{\boldsymbol{X}}\boldsymbol{\delta} + (\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}\boldsymbol{\varepsilon}.$$

Note that this reduced-form specification implies that the spatial lag of the endogenous variable is correlated with the error term, i.e.

$$Cov[(\boldsymbol{W}\boldsymbol{y}), \boldsymbol{\varepsilon}] = E[(\boldsymbol{W}\boldsymbol{y})\boldsymbol{\varepsilon}'] - E[\boldsymbol{W}\boldsymbol{y}] = \boldsymbol{W}(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}\sigma^2.$$

Hence, ordinary least squares (OLS) estimators will not be consistent.

An alternative to using OLS to estimate the model is provided by Maximum Likelihood (ML) estimation (compare e.g. Lee, 2004). This requires making a distributional assumption for the error terms. Above, it was assumed that the error terms follow a normal distribution, and in this case the log-likelihood function reads

$$\begin{aligned} \ln L(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = &- \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I} - \gamma \boldsymbol{W}| \\ &- \frac{1}{2\sigma^2}\left[(\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]' \left[(\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]. \end{aligned} \tag{3.43}$$

In particular, the presence of the determinant $\ln|\boldsymbol{I} - \gamma \boldsymbol{W}|$ in this expression might not be immediately obvious. The following derivation of the function above therefore sheds some light on this term.

3.4.2.1 Derivation of the Log-likelihood Function

Given the distributional assumption made above for the error (or disturbance) terms, ε_i, in a given region, these have the following probability density function

$$f(\varepsilon_i; 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right)$$

[138] On the existence of the inverse, see Appendix C.2.9.

so that the joint density function of the error terms reads

$$
\begin{aligned}
f(\varepsilon_1, \ldots, \varepsilon_N; \mathbf{0}, \sigma^2 \boldsymbol{I}) &= \prod_{i=1}^{N} f(\varepsilon_i; 0, \sigma^2) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right) \\
f(\boldsymbol{\varepsilon}; \mathbf{0}, \sigma^2 \boldsymbol{I}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right)
\end{aligned}
$$

where the last line follows from $\sum_{i=1}^{N} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$. However, the disturbance terms cannot be observed, and therefore the likelihood function needs to be based on $\boldsymbol{y}$, which is observable (Anselin, 1988b, 62). Hence, the vector of random variables $\boldsymbol{\varepsilon}$ needs to be transformed into the vector of random variables $\boldsymbol{y}$. This works with the help of a general result on the transformation of variables. It holds that the joint density function $g(\cdot)$ for $\boldsymbol{y}$ is given by (Davidson and MacKinnon, 2004, 430-431)

$$
g(\boldsymbol{y}) = f(\boldsymbol{\varepsilon}) \cdot \left|\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{y}}\right|.
$$

Due to this result, the determinant will enter the likelihood function. This determinant is also called the Jacobian (determinant) of the transformation (see, for example, Greene, 2003, 844-45). From Equation (3.42) it follows that the vector of disturbances is given by

$$
\boldsymbol{\varepsilon} = (\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y} - \tilde{\boldsymbol{X}}\boldsymbol{\delta}. \tag{3.44}
$$

Therefore, the Jacobian determinant for this case reads $\left|\frac{\partial \boldsymbol{\varepsilon}}{\partial \boldsymbol{y}}\right| = |\boldsymbol{I} - \gamma \boldsymbol{W}|$. Accordingly, the joint density function for $\boldsymbol{y}$ is

$$
g(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) \cdot |\boldsymbol{I} - \gamma \boldsymbol{W}|.
$$

As the likelihood function coincides with the joint density function (Verbeek, 2004, 164), it can be expressed as

$$
L(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \cdot \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) \cdot |\boldsymbol{I} - \gamma \boldsymbol{W}|.
$$

Inserting for $\boldsymbol{\varepsilon}$ from Equation (3.44), and taking the natural logarithm of this expression results in

$$\begin{aligned}\ln L(\boldsymbol{y};\boldsymbol{\delta},\gamma,\sigma^2) = &-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I}-\gamma\boldsymbol{W}| \\ &- \frac{1}{2\sigma^2}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]\end{aligned}$$

which is identical to Equation (3.43) above.[139]

Finding the ML estimator, requires maximizing the log-likelihood function with respect to the parameters $\boldsymbol{\delta}, \gamma$, and σ^2, i.e. setting the 5×1 score vector equal to the corresponding zero vector (Verbeek, 2004, 166-167). This multivariate optimization problem can be transformed into a univariate one by concentrating the log-likelihood function with respect to $\boldsymbol{\delta}$ and σ^2. The approach (see Pace and Barry, 1997, 235-236) is to substitute closed-form solutions for the estimators, $\hat{\boldsymbol{\delta}}(\gamma)$ and $\hat{\sigma}^2(\gamma)$, that depend only on the data and the unknown parameter γ, into Equation (3.43). These solutions can be derived from the first-order conditions for $\boldsymbol{\delta}$ and σ^2 (LeSage and Pace, 2009, 47). The resulting concentrated log-likelihood function can then be maximized with respect to the parameter γ to obtain an estimate, $\hat{\gamma}$, for this parameter. This estimate can in turn be used to back out estimates for the other parameters from the expressions for $\hat{\boldsymbol{\delta}}(\hat{\gamma})$ and $\hat{\sigma}^2(\hat{\gamma})$ (LeSage and Pace, 2009, 47).

3.4.2.2 Derivation of the Concentrated Log-likelihood Function

Following the approach outlined above, the derivative of the log-likelihood function with respect to σ^2 yields[140]

$$\frac{\partial \ln L(\cdot)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]. \tag{3.45}$$

Setting this derivative equal to zero, leads to the maximum likelihood estimator for σ^2, i.e.

$$\hat{\sigma}^2(\gamma) = \frac{1}{N}\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}\right]'\left[(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{y}-\tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}\right]. \tag{3.46}$$

Taking the derivative of Equation (3.43) with respect to $\boldsymbol{\delta}$ and solving for the maximum likelihood estimator $\hat{\boldsymbol{\delta}}$ is a little more involved so that at this point only the result is

[139] The log-likelihood function in the standard regression model might be more familiar, but no determinant occurs in that expression. The reason for the difference is that in the standard case where $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$ the Jacobian is equal to 1. In general, the presence of the determinant in the formula for the transformation ensures that after the transformation the volume under the joint probability density function is still equal to unity (LeSage and Pace, 2009, 80).

[140] Alternatively, taking σ as the parameter in the log-likelihood function instead of σ^2 would lead to identical results in the end.

presented, while the detailed derivation is delegated to Appendix C.3.1. The estimator is given by

$$\hat{\boldsymbol{\delta}} = \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}'(\boldsymbol{I} - \gamma \boldsymbol{W})\boldsymbol{y}. \tag{3.47}$$

Defining $\hat{\boldsymbol{\delta}}_{\boldsymbol{O}} \equiv \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}'\boldsymbol{y}$ and $\hat{\boldsymbol{\delta}}_{\boldsymbol{L}} \equiv \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1} \tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y}$, the estimator can be equivalently expressed as[141]

$$\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{\delta}}_{\boldsymbol{L}}.$$

Defining furthermore the estimated residuals of a regression of $\boldsymbol{y}$ on $\tilde{\boldsymbol{X}}$ as $\hat{\boldsymbol{e}}_{\boldsymbol{O}} \equiv \boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{O}}$ and the estimated residuals of a regression of $\boldsymbol{W}\boldsymbol{y}$ on $\tilde{\boldsymbol{X}}$ as[142] $\hat{\boldsymbol{e}}_{\boldsymbol{L}} \equiv \boldsymbol{W}\boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{L}}$, the maximum likelihood estimator $\hat{\sigma}^2$ can be expressed as

$$\hat{\sigma}^2(\gamma) = \left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right].$$

Substituting this estimator into the log-likelihood function in Equation (3.43), yields

$$\begin{aligned}
\ln L(\boldsymbol{y};\gamma) = & -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right] + \ln|\boldsymbol{I} - \gamma\boldsymbol{W}| \\
& - \frac{1}{2\frac{1}{N}} \cdot \frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})}{(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})} \\
= & -\frac{N}{2}\left[\ln(2\pi) + 1\right] + \ln|\boldsymbol{I} - \gamma\boldsymbol{W}| - \frac{N}{2}\ln\left[\frac{(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})'(\hat{\boldsymbol{e}}_{\boldsymbol{O}} - \gamma \hat{\boldsymbol{e}}_{\boldsymbol{L}})}{N}\right] \qquad (3.48)
\end{aligned}$$

which now only depends on the parameter γ. Computation of the maximum likelihood estimator $\hat{\gamma}$ is facilitated by taking recourse to a result by Ord (1975, 121). This result states that the determinant $|\boldsymbol{I} - \gamma\boldsymbol{W}|$ can be expressed in a simpler way via the eigenvalues $\lambda_i, \ldots, \lambda_N$ of the interaction matrix. In particular, it holds that $|\boldsymbol{I} - \gamma\boldsymbol{W}| = \Pi_{i=1}^{N}(1 - \gamma\lambda_i)$ or, after taking the natural logarithm: $\ln|\boldsymbol{I} - \gamma\boldsymbol{W}| = \sum_{i=1}^{N}\ln(1 - \gamma\lambda_i)$. This latter result is substituted into the log-likelihood in Equation (3.48). The advantage of employing this expression is that in the numerical optimization procedure for the determination of $\hat{\gamma}$, the eigenvalues need only be determined once (Ertur and Koch, 2011, 233). Having determined $\hat{\gamma}$ numerically, the value can be substituted into the closed-form solutions for $\hat{\sigma}^2$ and $\hat{\boldsymbol{\delta}}$ in Equations (3.46) and (3.47) to obtain the estimates for these parameters.

[141] This is an unbiased estimate. See Keilbach (2000, 153) for the proof.

[142] In Ertur and Koch (2011, 233) there is a slight mistake as they state (converted to the notation used here) that $\hat{\boldsymbol{e}}_{\boldsymbol{L}} = \boldsymbol{y} - \tilde{\boldsymbol{X}}\hat{\boldsymbol{\delta}}_{\boldsymbol{L}}$, whereas the expression given in the main text above is the correct one.

3.5 Data, Estimation Results, and Interpretation of Model Parameters

This section first provides a detailed overview of the data and the construction of the variables for the empirical analysis. Afterwards estimation results of the models specified in Section 3.4.1 will be presented and discussed. Estimates for the direct, indirect and total impacts of the variables in the spatial models will also be presented.

3.5.1 Data

The empirical analysis focusses on the US federal states. As is common practice in studies analyzing US economic development on a state level, Alaska, Hawaii, and (by definition) Washington, D.C. are dropped from the sample so that only the 48 contiguous (or continental) states are included.[143] In addition, following the approach by Bode et al. (2012, 27), Delaware is also excluded so that the baseline sample consists of 47 states. The state of Delaware is home to a large financial industry, and it might be the case that this characteristic influences the estimation results. Also, as Hanushek et al. (2015, 16) note, gross state product (GSP) in Delaware might not be well described by a standard production function, as more than 35% of its GSP in 2007 is accounted for by finance and industry, whereas the remaining states only reach less than half this value.[144]

The sample period in the empirical analysis below covers the 11 years from 1997-2007. This period is rather short, but still in line with the studies (on different units of observation) by, for instance, Ertur and Koch (2011, 235), who analyze a period of 14 years or Fischer (2011, 430) and Fischer et al. (2010, 592), who have data for 10 years.[145] For the present analysis, data for more recent years is available in the case of a subset of the variables used in the analysis. The reason 2007 is chosen as the final year is twofold: On the one hand, it is chosen to avoid the financial crisis starting in 2008 influencing the results, and, on the other hand, data for the investment in physical capital is only available up to 2007. For years prior to 1997 data is available for many variables pertaining to the analysis. However, 1997 is chosen as a cutoff, since the time series for the dependent variable has a structural break in that year.[146]

[143]Compare, for instance, Holtz-Eakin (1993), Barro and Sala-i-Martin (1992) or, more recently, Yamarik (2011) for this composition of the sample.

[144]Hanushek et al. (2015, 16) quote figures from an article in The Economist (2013) stating that Delaware is a tax haven where companies outnumber people (945,000 vs. 917,092).

[145]As Fischer (2011, 429) notes with reference to Durlauf and Quah (1999) and Islam (1995), steady-state regressions are valid for relatively short time periods.

[146]The break occurs as the US shifted from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS) and the Bureau of Economic Analysis on their website strongly "cautions against appending the two data series" (compare `http://www.bea.gov/regional/docs/product/` (accessed: 11 August, 2015)).

Output, y_i, is measured as real chained-weighted gross state product generated in the private sector measured in 2000 dollars, and the data stems from the Bureau of Economic Analysis' (BEA) regional accounts data (BEA, 2015b). The variable is constructed by dividing nominal gross state product generated in the private sector by the implicit price deflators for the gross domestic product (GDP), which is taken from the national accounts data of the BEA (2015a).[147] In more detail, the following approach is employed (compare Peri, 2012, 350): The time series for the GDP deflator is from the BEA (2015a) and has 2009 as its reference year. Therefore, the reference year for this series is first changed to the year 2000 before using these values to convert GSP in nominal dollars to GSP in 2000 real dollars.[148]

Labor is measured as total employment on private payrolls, as in, for example, Yamarik (2013). This data is reported by the Bureau of Labor Statistics (2015) in its Current Employment Statistics, and n_i is the average annual growth rate of total employment.[149] Values for the state-level real investment rate, $s_{K,i}$, are not available from official US agencies. However, Yamarik (2013), updating a previous contribution by Garofalo and Yamarik (2002), provides values for state-level real investment in 2000 dollars. Dividing those by the real GSP values then leads to the values for the state-level real investment rate. Furthermore, in Yamarik (2013), annual values for state-specific depreciation rates of physical capital, δ_i, are also provided so that here, in contrast to other studies, it is possible to deviate from the assumption of identical depreciation rates across all units of observation and use the average state-specific annual depreciation rate of physical capital in the empirical study instead.[150] The growth rate, g_w, is set to 0.02, which is in line with the value chosen by Howitt (2000, 841) and also is similar to the approach by Yamarik

[147] This is similar to, for instance, Yamarik (2006) and Barro and Sala-i-Martin (2004), who use the national consumer price index to deflate nominal personal income.

[148] As Barro and Sala-i-Martin (2004, 497) note: "As long as the same index is used at each date for each state, the particular index chosen does not affect the relative levels and growth rates across states".

[149] The values for this variable pose a slight problem for the estimation in the next section. The model is specified in logs, but, as the summary statistics in Table 3.1 show, the minimum value for the employment growth rate is -0.5%, for which the logarithmic transformation is not defined. Besides this value for Michigan, also Ohio has a negative employment growth rate over the period (-0.01%). Several approaches exist to deal with this issue. The one preferred here for its simplicity, follows Sarel (1996, 203), who encounters this problem in the context of inflation rates. He sets the negative values equal to the smallest positive observed value in the sample. For the present analysis, the respective value is 0.23% (for Mississippi). Alternatively, for comparable situations, it is suggested to add a constant to the variable before applying the logarithmic transformation to ensure that all values are positive (see, for example, Dowdy et al. (2004, 329) or Wooldridge (2013, 193)). Adding, for instance, 0.006 to all employment growth rates before taking logs ensures that the growth rate for Michigan is positive but small (0.1%). Thirdly, the observations for Michigan and Ohio could be dropped from the sample. All three methods of handling this problem lead to only minor quantitative changes in the estimated parameters. Note that the adjustment is only necessary for the employment growth rate variable, but not for the effective depreciation rate, $n_i + g^w + \delta_i$, which is positive for all observations.

[150] The data for $s_{K,i}$ and δ_i is available on Steven Yamarik's website under: `https://web.csulb.edu/~syamarik/` (accessed: 11 August, 2015).

(2006) considering that he obtains a mean value of 9% for $n_i + g_w + \delta_i$ where the sample covers the time period 1950-2000.[151] Investment in R&D, $s_{A,i}$, is measured as the average real research and development expenditure as a percentage of real gross state product. Data for this variable is provided by the Organisation for Economic Co-operation and Development's (OECD) Regional Database (OECD, 2015). Total R&D expenditures are given by summing up expenditures in the business, government, higher education, and private non-profit sectors (OECD, 2015). Concerning the human capital stock, H_i, this variable is measured by the average share of individuals above the age of 24 with four or more years of college (more specifically, a Bachelor's degree or higher). This is in accordance with the measure used by, for example, Bode et al. (2012) or Yamarik (2006). The data is supplied by the Current Population Survey of the United States Census Bureau (2015). For this variable, no state-level data is available for 2007 so that this year is omitted in calculating the average values.[152]

With respect to the interaction matrix $\boldsymbol{W}$, it is important to highlight that the weights should be exogenous to the variables in the model (Ertur and Koch, 2007, 1042). This restricts the choice of variables that might be considered to model connectivity between states considerably. In general, studies have relied on geographic distance to specify the weights in the interaction (or spatial weight) matrix. This measure allowed researchers to capture that effects between units of observations diminish with geographic distance (see, for example, Eaton and Kortum (1996) or Keller (2002)).[153] The distance-decay effect can be formalized in a variety of ways. Here, three different interaction matrices of the form $\boldsymbol{W} = \boldsymbol{HV}$ with general weights given by $w_{ij} = H_i v_{ij}$ will be considered to assess the robustness of the empirical results. As will be clear from the functional forms specified in Equations (3.49), (3.50), and (3.51) below, the matrices $\boldsymbol{V_1}, \boldsymbol{V_2}$ and $\boldsymbol{V_3}$ are row standardized, whereas the matrices $\boldsymbol{W_1}, \boldsymbol{W_2}$ and $\boldsymbol{W_3}$ are not, as they are multiplied by the matrix $\boldsymbol{H}$.

The first interaction matrix, $\boldsymbol{W_1}$, is based on a binary first-order contiguity matrix as in Fischer (2011, 430) or Rey and Montouri (1999, 146). States are considered contiguous (or, more simply, neighbors), if they share a common border (i.e. Montana and North

[151] Assuming $g_w = 0.02$, the average value of $n_i + g_w + \delta_i$ in the present sample is approximately 8% (see Table 3.1).

[152] See the user note at the following link: `http://www.census.gov/hhes/socdemo/education/data/cps/2007/usernote.html` (accessed: 11 August, 2015).

[153] Measures based on geographic distance are, however, not the only possibility. Another exogenous measure is genetic distance between units of observation. In the present analysis, it is unfortunately not possible to use this alternative measure, since the necessary bilateral distance measures are not available on the level of US states (see Spolaore and Wacziarg (2009) for the relevant country-level data). A similar issue arises for another potential candidate measure, linguistic distance, that has been used in studies applying spatial econometric methods (compare, for example, Isphording and Otten (2013) or Melitz and Toubal (2014)).

Dakota) and the modifier "first-order" refers to the fact that only direct neighbors are relevant[154] so that Minnesota is a first-order neighbor of North Dakota, but a second-order neighbor of Montana (see the map in Figure 3.1).[155] In formal terms, the weights in matrix $\boldsymbol{W_1}$ are therefore described by $w_{ij}(1) = H_i v_{ij}(1)$, with

$$v_{ij}(1) = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{\sum_{j \neq i}^{N} v_{ij}(1)} & \text{if } i \text{ and } j \text{ are neighbors.} \end{cases} \tag{3.49}$$

A second possibility to model the distance-decay effect abstracts from the binary option adopted above and connects all states directly with each other. The weights in matrix $\boldsymbol{W_2}$ are given by $w_{ij}(2) = H_i v_{ij}(2)$, and, as, for example, in Ertur and Koch (2011), the following continuous functional form is assumed for these weights

$$v_{ij}(2) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-d_{ij}}}{\sum_{j \neq i}^{N} e^{-d_{ij}}} & \text{otherwise.} \end{cases} \tag{3.50}$$

Here, d_{ij} is the great circle distance – the shortest path between two points on the surface of a sphere – between the geographic centroids of the US states. These centroids are illustrated by the black dots in Figure 3.4.

The third interaction matrix, $\boldsymbol{W_3}$, has weights $w_{ij}(3) = H_i v_{ij}(3)$ and has a similar form to the matrix in, for example, Bode et al. (2012) and Basile (2014). It adopts the negative exponential form of matrix $\boldsymbol{W_2}$, but scales it with a factor τ. In addition, a distance cutoff is introduced. If the distance between the centroids of the two states is larger than this threshold, the corresponding matrix entry is set to zero, implying that direct spillovers between these states are non-existent. Formally, the matrix entries are calculated by

$$v_{ij}(3) = \begin{cases} 0 & \text{if } i = j \text{ or if } d_{ij} > 512\text{km} \\ \frac{e^{-\tau d_{ij}}}{\sum_{j \neq i}^{N} e^{-\tau d_{ij}}} & \text{if } d_{ij} < 512\text{km.} \end{cases} \tag{3.51}$$

As in Bode et al. (2012), τ is set to 0.02, and the distance cutoff is chosen following Basile (2014, 12) as the minimum distance ensuring that all states have at least one neighbor. For the present sample, this distance is slightly below 512km (the distance between the centroids of Arizona and New Mexico). Concerning τ, Bode et al. (2012) supply a helpful illustration: They argue that the weights in the interaction matrices can be understood

[154]One might wonder about the quadripoint where the borders of Colorado, New Mexico, Arizona and Utah meet (see Figure 3.1). In the present analysis, the pairs Arizona/Colorado and Utah/New Mexico are considered neighbors.

[155]Note that this specification does not rule out spillovers from Minnesota to Montana, as all states are connected via the spatial multiplier (compare Equation (3.32)).

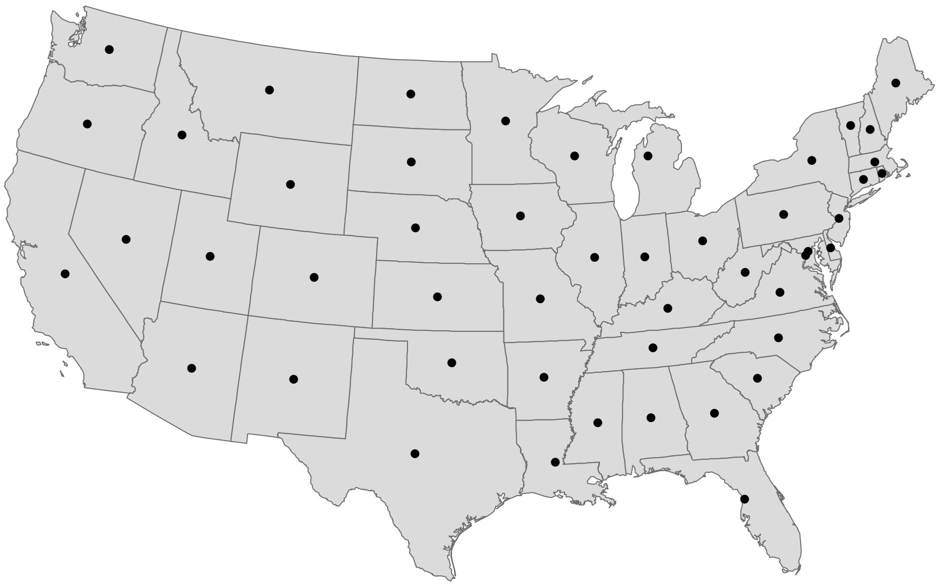


Figure 3.4: Geographic Centroids of US States.

similar to iceberg transportation costs[156] where the parameter τ indicates the percentage of knowledge diffusion that is lost per kilometer. For $\tau = 0.02$ this implies that after 50 kilometers $1 - e^{-0.02*50\text{km}} \approx 63.2\%$ of the iceberg "has melted away" and approximately 86.5% after 100 kilometers.

Before presenting the estimation results in the following section, Table 3.1 provides summary statistics for the variables used in the empirical analyses.

3.5.2 Estimation Results

Table 3.2 shows the estimation results[157] for the series of models described in Section 3.4.1. In Column 1, the standard Solow model from Equation (3.39) is estimated by ordinary least squares (OLS), and the results show that, in line with the predictions of this model, the investment rate in physical capital divided by the effective depreciation rate has a positive and significant impact on steady-state income per worker (p-value = 0.033). As the

[156] These are familiar from new economic geography (see, e.g. Krugman, 1991, 489). Samuelson described the general concept in the following way: "To carry [a] good across the ocean you must pay some of the good itself" and illustrated it more specifically by continuing that "only a fraction of ice exported reaches its destination as unmelted ice" (Samuelson, 1954, 268). However, the general idea goes back almost two centuries to von Thünen, who noted with respect to the transport of grain by horse-drawn carriage that if the distance between farm and city (and back to the farm) is large enough (50 miles in the specific example he describes), then "ist also der Transport des Korns auf 50 Meilen unmöglich, weil die ganze Ladung oder deren Werth auf der Hin- und Zurückreise von den Pferden und den dabei angestellten Menschen verzehrt wird" (von Thünen, 1826, 9).

[157] All estimations have been conducted in Matlab with the Spatial Econometrics Toolbox provided by LeSage. The toolbox is available under: `http://www.spatial-econometrics.com/` (accessed: 11 August, 2015).

Table 3.1: Summary Statistics – Baseline Sample.

Variable	Mean	Median	Standard deviation	Minimum	Maximum
y_i	85,012.07	80,896.36	13,921.60	66,616.49	123,281.63
$s_{K,i}$	0.085	0.081	0.017	0.063	0.139
n_i	0.012	0.011	0.008	−0.005	0.037
δ_i	0.048	0.048	0.002	0.044	0.051
$n_i + g^w + \delta_i$	0.080	0.079	0.008	0.062	0.101
$s_{A,i}$	0.022	0.019	0.015	0.005	0.075
H_i	0.255	0.244	0.046	0.158	0.351
$\frac{s_{K,i}}{n_i+g^w+\delta_i}$	1.064	1.034	0.172	0.862	1.706
$\boldsymbol{W_1 s_K}$	0.264	0.261	0.051	0.159	0.416
$\boldsymbol{W_2 s_K}$	0.274	0.262	0.070	0.162	0.494
$\boldsymbol{W_3 s_K}$	0.271	0.260	0.061	0.159	0.438
$\boldsymbol{W_1 y}$	21,073.61	20,214.21	4,628.53	12,977.69	32,609.39
$\boldsymbol{W_2 y}$	21,302.32	20,192.91	5,446.95	12,398.25	42,844.25
$\boldsymbol{W_3 y}$	21,616.61	19,961.00	5,454.27	13,399.71	38,037.59

Note: The given values are the original values (i.e. not in logs) for the benchmark sample of 47 states and the period 1997-2007 with y_i the income per worker in 2007.

model is specified in logs, the estimated coefficient points to an increase of approximately 3.3% due to a 10% increase in the investment rate in physical capital.

Table 3.2: Estimation Results for Three Different Models for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.322	11.690	11.019	10.966	10.965	11.253	10.977	10.977
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.326	0.362	0.310	0.325	0.282	0.329	0.321	0.278
	(0.033)	(0.017)	(0.023)	(0.023)	(0.053)	(0.018)	(0.029)	(0.064)
$\ln s_{A,i}$	—	0.065	—	—	—	0.023	−0.006	−0.005
		(0.041)				(0.494)	(0.868)	(0.906)
$\ln n_i$	—	0.023	—	—	—	0.018	0.010	0.008
		(0.467)				(0.523)	(0.717)	(0.769)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	−1.728	−0.567	−0.237	−1.641	−0.606	−0.268
			(0.096)	(0.263)	(0.719)	(0.118)	(0.239)	(0.687)
γ	—	—	0.117	0.126	0.126	0.096	0.131	0.129
			(0.004)	(0.000)	(0.001)	(0.048)	(0.009)	(0.011)
AIC	−3.794	−3.813	−3.896	−3.965	−3.942	−3.829	−3.885	−3.860
BIC	−3.715	−3.655	−3.738	−3.807	−3.785	−3.593	−3.648	−3.624
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Column 2 shows the estimation results from the Howitt model, specified in Equation (3.40). The coefficient for the investment rate over the effective depreciation rate has increased slightly, quantitatively as well as in significance, compared to the estimation of the Solow model. The newly added variable, investment in R&D, is estimated to have a positive and significant effect (p-value = 0.041) on per worker income in steady state. However, the effect is smaller than for investments in physical capital, as a 10% increase in R&D investment would result in a 0.65% increase in per worker income in steady state. Regarding the remaining variable, the employment growth rate, its effect is not significant with a p-value of 0.467.[158]

The next 3 columns estimate the spatially-augmented Solow model from Equation (3.41) by maximum likelihood as described in Section 3.4.2. Here, the approach differs slightly from Ertur and Koch (2011), as the approach from Basile (2014) is followed to estimate the Spatial Durbin Model instead of the Spatial Error Model (SEM) to obtain estimates for the coefficient of the spatial lag of the investment variable as well.[159] For all three interaction matrices the effect of the investment variable accords with implications derived from the theoretic model, and, with the exception of the matrix $\boldsymbol{W_3}$, is also significant at the 5%-level. The estimated coefficients for the spatial lag of the investment variable are not significant in either case. Note, however, that the estimate for the spatial autoregressive coefficient is highly significant (at the 1%-level) for all three matrices.

As this model, by definition, contains interaction between regions, and an interdependent system is estimated, a direct interpretation of the estimated parameters as in the case of the models estimated by OLS is not feasible and might lead to invalid conclusions. The next section presents a method developed by LeSage and Pace (2009) to disentangle the direct and indirect impacts in spatial models.

In the remaining three columns, the multi-region Schumpeterian growth model from Equation (3.37) is estimated. Similar to the case of the spatially-augmented Solow model, the impact of the investment rate in physical capital is positive for all three interaction matrices and significant at the 5%-level for matrices $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$. Also, the spatial lag of this variable is significant in neither case at standard significance levels. Concerning the newly added variables, the investment rate in R&D and the employment growth rate, the estimated coefficients are not significant in either case. However, again the estimate for the spatial autoregressive coefficient, γ, is estimated to be positive and significant

[158]The non-significance of this variable in the Howitt specification is also found by Ertur and Koch (2011) in their cross-country sample.

[159]In contrast to the SDM model, the SEM model contains spatial autocorrelation only in the error term, but not in the regressors. See Ertur and Koch (2011, 234 and 240-241) for the specific model addressed here and, for instance, LeSage and Pace (2009) on the spatial error model in general.

for all three matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$, implying that the states cannot be treated as independent observations.

These estimation results do not provide a clear picture, as, for example, the impact of the R&D investment rate, seems to affect income per worker in steady state in the non-spatial model, but not in the spatial model, although the information criteria point to the latter one.[160] Nonetheless, due to the estimates for the parameter γ, it emerges from these results that interaction effects between observations seem to be relevant.[161]

In Section 3.5.1 it has been mentioned that the state of Delaware has been excluded from the baseline sample due to the presence of a large financial and insurance sector. Appendix C.4.1 presents the estimation results when Delaware is included in the sample (see Table C.1). As it turns out, including the state in the sample, results in the coefficient on the physical investment rate over the effective depreciation rate losing its significance, thereby lending credence to the conjecture that this state might not be well described by the model considered here.[162]

3.5.3 Interpretation of the Model Parameters

Due to the interaction effects contained in the spatial models via the inclusion of the spatial lags, the coefficient estimates in Table 3.2 cannot be interpreted directly. At this point, it is helpful to refer back to the elasticities calculated in Equations (3.31), (3.35),

[160] Akaike's Information Criterion (AIC) and the Schwarz or Bayesian Information Criterion (BIC) are calculated according to the formulae (see, for example, Greene (2003, 160)):

$$\text{AIC} = \log\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}\right) + 2\frac{K}{N} \qquad \text{and} \qquad \text{BIC} = \log\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}\right) + \frac{K}{N}\log N,$$

where $\hat{\varepsilon}$ denotes the residuals of the estimation and K signifies the number of parameters (for the original contributions regarding these information criteria, see Akaike (1973) and Schwarz (1978)). It should be kept in mind here that only nested models can be compared according to these criteria. Accordingly, comparisons are possible across model with the same interaction matrix, but not between, for example, the models in the last two columns.

[161] The goal here is to test empirically the four different types of models that are contained in a "completely integrated theoretical and empirical framework" (Ertur and Koch, 2011, 216). Hence, the subject of model comparison as traditionally understood, is assigned a reduced role here. In the context of comparison of (spatial) econometric models, the two ends of the spectrum are the specific-to-general approach and the general-to-specific approach (see, for example, Le Gallo (2014, 1528-1529) on these approaches). The former strategy has been found to outperform the latter strategy in a specific context not including the SDM as a possible specification (Florax et al., 2003). On the other hand, LeSage and Pace (2009) suggest to start with the SDM model, whereas an approach outlined by Elhorst can be seen as a combination of the two search strategies that chooses as a starting point, however, the specific model. A further reason these approaches have not been adhered to strictly here is that they rely on tests which have been specified for row-standardized interaction matrices (compare, for example, Anselin (1988a) and Anselin et al. (1996)). It is not clear, if these can be applied in the given context in a straightforward manner for models in which the interaction matrix is not described by this characteristic.

[162] Also, Washington D.C. has been omitted from the sample. Including it does not lead to qualitative changes in the results compared to the baseline estimates. Detailed estimation results are omitted though.

and (3.36). These $N \times N$ matrices describe the effects of changes in the explanatory variables on the dependent variables. The individual entries in these matrices denote, for instance, the effect of an increase in the investment rate in R&D in Maine on the per worker income in steady state in North Dakota. It becomes clear that the effects will differ depending on the pairs of states chosen and thus reporting all individual effects is rather unwieldy. LeSage and Pace (2009) helpfully provide a method to summarize in a clear manner the estimation results on the direct and indirect effects (or spillovers).[163] The direct effects are the partial derivatives measuring the change in the dependent variable in region i due to a change in the explanatory variable in region i. These effects are measured on the diagonal of the matrix of elasticities (compare Equation (3.33)). LeSage and Pace (2009) suggest to summarize the direct impact with the average value of the diagonal matrix elements.

A change in the explanatory variable in region i also affects the dependent variable in the other regions, and these indirect impacts are captured by the off-diagonal entries in the matrix (compare Equation (3.34)). With regard to this effect, the proposed summary measure is the average of the row sum of these off-diagonal matrix entries. This row sum measures the effect on the dependent variable in region i due to a change in the explanatory variables in the remaining regions. Straightforwardly, the average of these row sums is then chosen as the summary measure for the indirect effects.[164] Summing up the direct and indirect effects (i.e. all the elements in a row) gives then a measure for the total impact. The average of these sums is chosen as the corresponding summary measure.

Table 3.3 presents the estimates for the direct, indirect, and total impacts for the multi-region Schumpeterian model calculated in the way just described. The results show that the indirect effects are not significant for any of the variables included in the regression, independently of the specific interaction matrix.[165] Concerning the estimates for the direct and total impacts, these are positive and significant for the investment rate over the effective depreciation rate in the case of interaction matrices $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$, but not if matrix $\boldsymbol{W_3}$ that contains a distance cutoff is included. Quantitatively, the significant

[163] A very lucid exposition of their approach can be found in Section 6 of Elhorst (2010).

[164] As LeSage and Pace (2009, 37) demonstrate, an identical value for the indirect effect is obtained by summing up the off-diagonal column elements and calculating the average of these sums. The interpretation is however different, as, for instance, the latter measure captures the impact of a change in the exogenous variable in region i on the dependent variable in all other regions. In the context, of the present model this measure reports, for example, the impact of an increase in R&D investment in Massachusetts on the per worker income in the remaining US states, whereas the sum of the off-diagonal row elements would report the change in, for example, the per worker income in Massachusetts due to a change in the R&D investment rate by an identical amount in the remaining states.

[165] Inference on the statistical significance of the parameters is based on p-values which have been obtained from simulating the distribution of the respective effects with the help of the variance-covariance matrix derived in Appendix C.3.2.

Table 3.3: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.330	0.321	0.278
	(0.022)	(0.034)	(0.070)
$\ln s_{A,i}$	0.023	−0.006	−0.005
	(0.497)	(0.867)	(0.905)
$\ln n_i$	0.018	0.010	0.008
	(0.525)	(0.718)	(0.772)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.034	0.047	0.039
	(0.149)	(0.105)	(0.150)
$\boldsymbol{W} \ln s_{A,j}$	0.001	−0.003	−0.002
	(0.724)	(0.699)	(0.724)
$\boldsymbol{W} \ln n_j$	0.002	0.001	0.001
	(0.623)	(0.797)	(0.851)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W}\frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.364	0.368	0.318
	(0.020)	(0.031)	(0.066)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	0.025	−0.009	−0.007
	(0.509)	(0.837)	(0.875)
$\ln n_i + \boldsymbol{W} \ln n_j$	0.020	0.012	0.009
	(0.528)	(0.726)	(0.780)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

estimates point to an increase of approximately 3.6% in per worker income due to a 10% increase in investment in physical capital.[166]

Regarding the sample that includes Delaware, estimates for the impacts are given in Table C.2 in Appendix C.4.1. They show that, in contrast to the baseline sample, the direct and total impacts are not significant no matter the interaction matrix included.

Summarizing the empirical results with respect to the multi-region Schumpeterian growth model, it needs to be stated that even though the model's implications are borne out for a particular sample in a cross-country analysis in Ertur and Koch (2011), these results are not readily transferable to the sample of US states analyzed here. Whereas R&D investments have a positive impact on income per worker in the Howitt model, in which the amount of knowledge that diffuses between regions is identical (see Ertur and Koch (2011, 238)), this is not the case in its version with more complex spatial interactions.

[166]For the spatial Solow model, i.e. the SDM model in Columns 3-5 in Table 3.2, the respective impacts are not significant in any case and detailed results are omitted here.

It might be that the inclusion of the spatial lags in the SDM model is not warranted. As Greene (2003, 151) notes, in such a situation the estimates become less precise and therefore are less likely to be significant. Indeed, the results from testing for the presence of spatial autocorrelation in the residuals of the Howitt model with Moran's I test[167] do not point to estimating a spatial version of the model. However, as the results in Table 3.2 show, the estimate for the spatial autoregressive coefficient is highly significant.[168] The estimation of this model therefore provides new information; in particular, when compared to the results by Basile (2014) for 248 European NUTS 2 regions. He estimates a growth version of the multi-region Schumpeterian model for these regions for the period 1991-2011 and finds that the estimates have the signs implied by theory and are significant.[169] For the US states, it might be that spatial interaction between states exists, which is, however, only captured by the variable income per worker and not by, for instance, R&D investments. The Moran scatterplot in Figure 3.2 only hinted at potential spillovers from R&D investments, but the econometric analysis finds no support for these.

Before concluding, a final series of estimation results for the baseline sample of observations will be briefly discussed. Despite the warning against appending the two data series for the GSP variable mentioned in Section 3.5.1, Appendix C.4.2 ignores this. The results are qualitatively similar to the ones for the shorter sample. However, notable differences in the significance of variables exist, for example, in the Howitt model where the R&D investment rate no longer has a statistically significant impact. In the multi-region Schumpeterian model when matrix $\boldsymbol{W_1}$ is used, the spatial autoregressive coefficient is not significant (compare Table C.4 for these results). A further difference concerns the direct and total impact estimates for the variable investment rate in physical capital over the effective depreciation rate. These have increased in size to values larger than 0.5 and are highly significant with p-values below 0.003 (see Table C.5).[170]

[167]The test statistic is given by

$$I = \frac{N}{S_0}\left(\frac{\hat{\varepsilon}'\boldsymbol{W}\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}\right)$$

where $\hat{\varepsilon}$ are the residuals from the OLS regression and S_0 is a standardization factor, that equals 1 in the case of a row-standardized interaction matrix, as it is given by the sum of all the elements in $\boldsymbol{W}$ (see, for example, Le Gallo (2014, 1524), who also provides the expressions for the expectation and variance of I, derived by Cliff and Ord (1972) under the null hypothesis).

[168]Tentative evidence from more specific Lagrange Multiplier tests, which in contrast to Moran's I test, specify a particular alternative hypothesis also point to including a spatial lag in the Howitt model. Results of these tests are omitted here though due to the possible issues regarding non-standardized interaction matrices mentioned at the end of Footnote 161.

[169]He also provides estimates for the direct, indirect, and total impacts, but no information on their significance is given.

[170]Also, in the case of interaction matrix $\boldsymbol{W_2}$, the indirect effect is marginally significant now at the 5%-level.

3.6 Conclusion for Chapter 3

In this chapter, the multi-region Schumpeterian growth model developed by Ertur and Koch (2011) has been presented in detail. A characteristic feature of this model is that regions are not considered to develop in isolation from each other, but rather interdependence between regions via knowledge spillovers is explicitly included. Technological progress in the model results from purposeful investments in R&D. It has been shown that how much a region can benefit from a given amount of knowledge spillovers depends on the way a region is connected to other regions and on the distance to its own technological frontier.

Also, the econometric strategy to estimate the equation for the steady-state income per worker that results from the theoretical model has been thoroughly outlined. In contrast to the original contribution, the level of aggregation in the empirical analysis has been reduced, and the model's implications have been tested for a sample consisting of states within a single country (the United States) instead of across countries.

The estimation results presented here do not provide full support for all implications derived from the theoretic model. For instance, the hypothesis of technological interdependence between the regions receives support, as the parameter gauging this characteristic is estimated to be positive and statistically significant for all three interaction matrices considered. However, a statistically significant impact of, for example, R&D investment on per worker income could not be detected in this model, even though it was present in the model with a simplified interaction structure (i.e. the Howitt (2000) model). Despite this result, the more nuanced way interaction is modeled in Ertur and Koch (2011) may seem more plausible, as these authors assume that the net effect of the knowledge spillovers depends on the absorptive capacity, i.e. the level of human capital in the receiving region, which is in contrast to the more basic assumption that the amount of knowledge diffused by each region to the other regions is identical.

This distinction may also point to an explanation for the differing estimation results. As the OECD notes in its Science, Technology and Industry Outlook: “US firms are at or near the forefront of technological advances in a number of areas” (OECD, 2010, 232), and the “United States has long been, and still is, at the forefront of cutting-edge science, technology and innovation” (OECD, 2014, 444). Moreover, the various US states are heterogeneous. Hence, it might be the case that potential knowledge spillovers from investment in R&D and physical capital arise in the form of highly-specialized knowledge in a given state, and this knowledge might only diffuse to a very low extent, as it cannot be productively used in the states the originating state is connected to. The receiving states might lack the absorptive capacity to benefit from inter-industry spillovers.

From a different perspective, the model does not differentiate between, for example, various types of workers and an identical level of human capital in two states might hide a large diversity in the composition of human capital. If one assumes identical human capital levels in two states that have different industry structures whose requirements are mirrored in the diversity of the respective state's human capital, then the model's mechanics would imply that spillovers originating in the state with a strong presence in e.g. nanotechnology to the state with a large presence in, for example, car manufacturing would necessarily be reflected in an increase in per worker income. However, the knowledge generated in nanotechnology might not be readily applicable in the car manufacturing sector, since human capital in this sector lacks the necessary complementarity to benefit from the knowledge spillovers. The implied impact on per worker income might then not show up in the data.

This chapter has deliberately chosen to stay in a similar framework as Ertur and Koch (2011), both theoretically as well as econometrically, to obtain results that are comparable to a certain extent. Naturally, other estimation approaches exist, and future research will focus on estimating, for instance, a spatial panel model for this sample. Also, the specific choice of the interaction matrix is an interesting topic for further study. In the present analysis, even though the estimation results for the three interaction matrices were similar, they were not identical. The method of Bayesian Model Averaging may be a fruitful avenue for finding a matrix that fits the data more closely. Furthermore, as knowledge spillovers decrease with distance, conducting the analysis at e.g. the level of the county or metropolitan area might lead to additional insights. However, data availability is the restricting factor in this case.

Conclusion and Outlook

A starting point of this thesis has been the observation that economic activity does not appear to be distributed randomly across space and that location matters for people's (economic) well being. Though not the only determinant, spatial interaction between, for example, countries or regions is of central importance for this result as the present thesis has demonstrated in a variety of ways.

After presenting key empirical facts and theories on the aspects of agglomeration (urbanization), economic growth, and integration, Chapter 1 has looked more closely at integration, since it provides the linkage between agglomeration and growth. It was argued that integration needs to be understood in a more nuanced way than in standard new economic geography models, in which integration is mostly captured by a decrease in transport costs. Given knowledge's role in determining economic growth, the cost of sharing information, which has an impact on the strength of knowledge spillovers, needs to be considered as well. Embedded in a discussion on the notions of size and scale (effects), it was pointed out that via integration mere size, for example, the absolute number of workers in a region, might be transformed into density, i.e. the number of workers in a region of given size. In contrast to size, density might then be able to activate latent agglomeration economies (or external economies of scale) in a region. Of central importance for integration to assume this transformative role is on the one hand, the specific design of integration by policy makers and on the other hand, the institutional foundation in the concerned regions. To quote a key sentence from Chapter 1: "integration pins down the impact of institutions to a spatial dimension" (see page 27).

In addition, no "one size fits all"-policy exists: Integration policies depend not only on the level of aggregation under consideration (for instance, integrating countries vs. integration of regions within a country), but also on the level of development. Chapter 1 has demonstrated that, for example, taking advantage of agglomeration economies in highly urbanized countries clearly requires different policies than trying to achieve the same goal in countries at the other end of the spectrum. The discussion in Chapter 1 highlighted the key concepts in a mostly verbal manner even though numerous empirical facts and a model have been presented as well.

Chapter 2 shifted the methodological focus by investigating knowledge spillovers in a fully integrated theoretic and empiric spatial growth framework due to Ertur and Koch (2007).

Their model includes technological interdependence working via spatial externalities between countries and provides an important addition to the standard growth models used in the literature.

A series of robustness checks, divided into two dimensions, were conducted in Chapter 2. As exogeneity of the weights in the interaction matrix is important, the choice of possible measures is restricted. However, based on a suggestion by Ertur and Koch (2011) and considerations put forward in Spolaore and Wacziarg (2009), genetic distance provides a plausible measure, since it captures barriers to the diffusion of development. The first check therefore concerned choosing genetic distance as an alternative to geographic distance for the choice of the bilateral weights in the model's interaction matrix. Estimating the reduced-form empirical specification determining income per capita confirms the finding that countries need to be investigated in an interdependent context, also if genetic distance is used. The estimated spatial autoregressive coefficient is positive and highly significant in both cases. Nonetheless, the findings are not identical for both distance measures. To highlight one example, Ertur and Koch (2007) estimate an implied capital share of slightly below one third, which is in line with values predicted by e.g. Gollin (2002). This result cannot be confirmed with genetic distance. In this case, the respective value is implausibly high.

Next, robustness was analyzed with respect to the underlying data source for the variables not depending on geographic or genetic distance. Taking the data from Penn World Table Version 6.2 or 7.1 instead of 6.1 as in Ertur and Koch (2007) and basing the interaction matrix on geographic weights, supports the hypothesis that countries need to be analyzed in an interdependent system and that models neglecting this characteristic are misspecified. However, continuing with the example of the capital share of income, this estimate is not robust across different versions of the PWT. The implied coefficient is not statistically different from zero for more recent versions of the data. This even holds for the estimation based on PWT 6.1, if the sample size is adjusted to produce a balanced sample across the three different versions to account for missing data in PWT 6.2.

With the help of a methodology devised by LeSage and Pace (2009), it was possible in Chapter 2 to quantify the strength of cross-country knowledge spillovers that result from spatial interaction. Ertur and Koch (2007) provide no direct detailed information on these knowledge spillovers, but it turns out that those spillovers arising from the investment in physical capital are not significant in their original sample. With one exception this result is robust across the three different versions of the PWT that were considered. This is in stark contrast to the results that were obtained using genetic distance. In this case, only one version (PWT 7.1) implied an insignificant estimate. For the other versions, the corresponding estimates were highly significant.

A last result from Chapter 2 concerns an inconsistency in Ertur and Koch (2007). They specify two different functional forms for the weights in the interaction matrix, but their published results actually use a form that differs by a scaling factor from the form specified in the publication. This has non-trivial implications for the results as, for example, the estimate for the spatial autoregressive parameter changes sign between the two specifications.

Chapter 1 was intensively concerned with integration, but also pointed out the importance of institutions, whereas Chapter 2 investigated the strength of knowledge spillovers across countries. It was noted that knowledge diffusion is not a frictionless process as, for example, institutional differences between countries play a role.

Chapter 3 picked up this train of thoughts again by staying in an extended framework of the model from the previous chapter, while highlighting the removal of some of the institutional differences by studying knowledge spillovers within the integrated institutional framework of the United States. Within the context of this extended framework, the multi-region Schumpeterian growth model by Ertur and Koch (2011), the hypothesis of technological interdependence between the US states receives empirical support. Nonetheless, neither a direct nor an indirect impact of R&D investments on per-worker income on the state level could be detected for either of the three alternative interaction matrices considered. The first was based on contiguity, more specifically, first-order neighborhood, the second one used the negative exponential of geographic distance for the weights and the third one combined this functional form with a distance cutoff.

Results differ, if the equation is estimated for a version of the model that is nested within the multi-region Schumpeterian growth framework and assumes a simplified interaction structure between states, in which the amount of knowledge spillovers does not depend on the absorptive capacity of the receiving state. In this simplified version of the model, R&D investments have a direct impact on per worker income. It was hypothesized that the reason for these contrasting results might be due to the US being a global technological leader. In order to take advantage of possible knowledge spillovers from one state, the receiving state would therefore need to have a workforce that is equipped with human capital that is complementary to these knowledge spillovers. Only a comparatively high level, as implicitly assumed in the simpler model, might not be enough.

In the following, concluding remarks with a brief outlook on further research and open questions will be given. I think the considerations from Chapter 1 clearly point to the fact that integration is a multidimensional issue and that policies aimed at closer integration need to be tailored to the scale at which they are meant to have an impact as well as to the institutional framework into which they are introduced. For instance, if the policy goal is to increase density in cities with the goal of benefiting from agglomeration economies like the ones implied by Marshall (the "mysteries of the trade become no mysteries; but are as it were in the air" (1890, IV.X.7)), then merely removing restrictions on internal migration or migration across country borders is not sufficient. People flocking to a city often will have no adequate places to live and work, and frequently NIMBYism is the problem,[171] which then needs to be addressed at the local level.

Also, trying to represent the dynamic properties of integration over the course of development and arriving at an endogenization of the concept in theoretic models could lead to new insights. In this regard, possible threshold effects might merit further attention, both when it comes to e.g. increasing density in the course of economic development, but also when loss of spatial concentration of economic activity is the relevant issue, thereby capturing possibly structural change. Ideally, such an approach would be combined with a mapping of the theoretical model to stylized facts.

The integrated theoretical and empirical frameworks discussed in Chapters 2 and 3 have the advantage of flexibility in the sense that their implications can be tested empirically for different levels of aggregation. They are not exclusively applicable to study cross-country interdependence. The model from Chapter 2 has been applied, for example, to European NUTS 2 regions and Brazilian microregions (see Fischer (2011) and Rodrigues Júnior et al. (2010), respectively).

In many empirical studies data availability restricts the sample size and time horizon. Apart from this factor, this thesis has clearly shown that data quality is another issue that cannot be neglected. Data on GDP or investment rates at various levels of aggregation are not as precise as one would like. In the same sense as data on economic activity is inferred from data on illumination at night captured by satellites (see Figure 1), values for the variables used in the regressions are to a certain extent also only indicators and therefore uncertain. I believe, it is important that researchers relying on data from the PWT check the sensitivity of their estimation results across different versions. Currently, research is being conducted to arrive at a "consensus estimate" for the income variable from the various versions of the PWT (see Crespo Cuaresma et al., 2015). This methodology could be adapted to arrive at similar consensus estimates for the remaining variables that depend on data from the PWT, too.

Also, with the World Development Indicators (WDI), a further data set exists that could be used in cross-country regression analyses. However, as Ram and Ural (2014) note in a short comparison study of these two data sets, no general conclusion can be made about

[171] The acronym stands for "not in my backyard" and pejoratively captures the opposition of residents to new zoning rules allowing new development projects in their neighborhood. See e.g. Avent (2011) on examples for the US concerning this issue.

the differences in results that would occur if the same model was estimated with first taking data from the WDI and then from the PWT.

Finally, in the models in the last two chapters of this thesis, the frictions in the process of knowledge diffusion were proxied by geographic and genetic distances. However, that does not imply that policy is ineffective in influencing the strength of knowledge spillovers, even though the two distances itself are, at least in my opinion, definitely outside the range of morally acceptable policy variables. Policy can nonetheless still influence the net effect of geographic or genetic distance on knowledge diffusion by reducing barriers across societies. As Spolaore and Wacziarg (2009, 524) point out, translation and adaptation of, for instance, technological innovations from one tradition to another might be effective in this situation. This is also the case for the promotion of openness and exchanges across societies. Hence, in the end, a better understanding of interaction is key.

A Appendix to Chapter 1

List of Geographic Entities

A list of countries and entities sorted into the various regions in Figures 1.1 and 1.2 according to the classification in the New Maddison Project Database is given below.

Table A.1: Countries Aggregated into Major Geographical Regions According to the Classification in the New Maddison Project Database.

Africa:	Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoro Islands, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Morocco, Mozambique, Namibia, Niger, Nigeria, Republic of the Congo, Rwanda, Saint Helena, Sao Tomé & Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Western Sahara, Zambia, Zimbabwe
Asia:	Afghanistan, American Samoa, Bahrain, Bangladesh, Bhutan, Brunei, Burma, Cambodia, China, Cook Islands, East Timor, Fiji, French Polynesia, Guam, Hong Kong, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kiribati, Kuwait, Laos, Lebanon, Macao, Malaysia, Maldives, Marshall Islands, Micronesia, Mongolia, Nauru, Nepal, New Caledonia, North Korea, Northern Mariana Islands, Oman, Pakistan, Palau, Papua New Guinea, Philippines, Qatar, South Korea, Samoa, Saudi Arabia, Singapore, Solomon Islands, Sri Lanka, Syria, Taiwan, Thailand, Tonga, Turkey, Tuvalu, United Arab Emirates, Vanuatu, Vietnam, Wallis and Fortuna, West Bank and Gaza, Yemen
Latin America:	Anguilla, Antigua and Barbuda, Argentina, Aruba, Bahamas, Barbados, Belize, Bermuda, Bolivia, Brazil, British Virgin Islands, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Montserrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, St. Kitts and Nevis, St. Lucia, St. Pierre and Miquelon, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Turks and Caicos Islands, United States Virgin Islands, Uruguay, Venezuela
Western Europe:	Andorra, Austria, Belgium, Cyprus, Denmark, Faeroe Islands, Finland, France, Germany, Gibraltar, Greece, Greenland, Guernsey, Iceland, Ireland, Isle of Man, Italy, Jersey, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, United Kingdom

B Appendix to Chapter 2

B.1 Definition of F_{ST} Genetic Distance

This appendix provides a formal definition of the concept of F_{ST} genetic distance which was developed by Wright (1951).[1] Consider as an example the case of two populations, A and B, that are of equal size and a single gene which can either have the form of allele 1 or of allele 2.[2] Denote the gene frequency of allele 1 in population A by p_A and the one for allele 2 by q_A. The probability that two randomly selected alleles at the locus under consideration will be identical (i.e. homozygosity occurs) is given by $p_A^2 + q_A^2$. The case of heterozygosity (i.e. two randomly selected alleles will differ) then is $het_A = 1-(p_A^2+q_A^2) = 2p_Aq_A$. This follows as $p_A+q_A = 1$ and hence $(p_A+q_A)^2 = p_A^2+q_A^2+2p_Aq_A = 1$ holds. For population B, the equivalent expressions for homozygosity and heterozygosity are given by $p_B^2 + q_B^2$ and $het_B = 1 - (p_B^2 + q_B^2) = 2p_Bq_B$, respectively.

Denoting the average gene frequencies of the two alleles in the two populations as $\overline{p} = \frac{1}{2}(p_A + p_B)$ and $\overline{q} = \frac{1}{2}(q_A + q_B)$, it follows that in the sum of the two populations heterozygosity is given by $het_{AB} = 1 - (\overline{p}^2 + \overline{q}^2) = 2\overline{pq}$. The average heterozygosity in the two populations is $het_{mean} = \frac{1}{2}(het_A + het_B)$. By comparing het_{mean} to het_{AB}, F_{ST} is a measure for the “variation in gene frequencies of populations” (Spolaore and Wacziarg, 2009, 525)

$$F_{\text{ST}} = \frac{het_{AB} - het_{mean}}{het_{AB}} = 1 - \frac{p_Aq_A + p_Bq_B}{2\bar{p}\bar{q}} = \frac{1}{4}\frac{(p_A - p_B)^2}{\overline{p}(1-\overline{p})}. \tag{B.1}$$

It now follows that the genetic distance between two populations is zero, if their allele frequencies at the given locus are identical (i.e. $p_A = p_B$) and that F_{ST} equals one if the respective frequencies are completely different (i.e. $p_A = 1$ and $p_B = 0$ or vice versa).[3]

[1] For a compact review of F_{ST} that covers additional details, see Holsinger and Weir (2009).

[2] Compare Spolaore and Wacziarg (2009, 524-525) for this approach and Cavalli-Sforza et al. (1994, 26-27) for extensions to more than two alleles and two populations.

[3] Note that Cavalli-Sforza et al. (1994, 29), for instance, provide the following formula for F_{ST} genetic distance

$$F_{\text{ST}} = \frac{V_p}{\overline{p}(1-\overline{p})} \tag{B.2}$$

where $\overline{p}$ are the average gene frequencies across the populations under consideration, and V_p indicates the variance between gene frequencies across these populations. If now $p_A \equiv \overline{p} + \sigma$ and $p_B \equiv \overline{p} - \sigma$ with $\sigma \geq 0$ and the variance is denoted by σ^2, then the formula in Equation (B.1) is equivalent to the one provided in Equation (B.2) (see Spolaore and Wacziarg (2015, 6-7) for this derivation).

In the construction of the interaction matrix based on genetic distance in Section 2.5, weighted F_{ST} distances are used to account for the fact that populations in, for instance, the United States or the United Kingdom consist of many subpopulations. If now the United States contains the populations $i = 1, \ldots, I$ and the United Kingdom the populations $j = 1, \ldots, J$ and s_{1i} is the share of population i in the United States and s_{2j} the share of population j in the United Kingdom, then the weighted F_{ST} genetic distance between these two countries is given by (see, for example, Spolaore and Wacziarg (2009, 484-485))

$$F_{\mathrm{ST}}^{\mathrm{W}} = \sum_{i=1}^{I} \sum_{j=1}^{J} (s_{1i} \times s_{2j} \times d_{ij}) \qquad \text{(B.3)}$$

where d_{ij} denotes the F_{ST} genetic distance between populations i and j.

B.2 Spatial Weight Matrices and Great Circle Distances

Numerous possibilities exist to model spatial connectivity via a spatial weight matrix. This appendix illustrates two possibilities and provides details on the calculation of great circle distances that are commonly used in empirical work. As an example, for how to model the spatial relationship between geographic regions, consider the four NUTS 2 regions[4] Schleswig-Holstein (SH), Hamburg (HH), Lüneburg (LÜ), and Mecklenburg-Vorpommern (MP) depicted in the map in Figure B.1.

Figure B.1: NUTS 2 Regions in Northern Germany.

[4]NUTS is an acronym of the French Nomenclature des Unités territoriales statistiques, i.e. the Nomenclature of territorial units for statistics of the EU, and the NUTS 2 level comprises government regions.

Defining now those regions that share are common border as neighboring regions, the following spatial weight matrix of first-order neighbors can be constructed

$$\boldsymbol{W_A} = \begin{pmatrix} & \text{SH} & \text{HH} & \text{LÜ} & \text{MV} \\ \text{SH} & 0 & 1 & 1 & 1 \\ \text{HH} & 1 & 0 & 1 & 0 \\ \text{LÜ} & 1 & 1 & 0 & 1 \\ \text{MV} & 1 & 0 & 1 & 0 \end{pmatrix}.$$

In row-standardized form, the result is

$$\boldsymbol{W_B} = \begin{pmatrix} & \text{SH} & \text{HH} & \text{LÜ} & \text{MV} \\ \text{SH} & 0 & 1/3 & 1/3 & 1/3 \\ \text{HH} & 1/2 & 0 & 1/2 & 0 \\ \text{LÜ} & 1/3 & 1/3 & 0 & 1/3 \\ \text{MV} & 1/2 & 0 & 1/2 & 0 \end{pmatrix}. \tag{B.4}$$

Here, queen contiguity is used as a concept to determine the spatial weights.[5] However, a spatial weight matrix can also be constructed based on geographic distance between units of observations. One relatively straightforward alternative in this vein is the geographic distance between country capitals. Consider, for example, the four capitals Berlin, Buenos Aires, Moscow and Seoul and take as spatial weights the shortest distance between them. Since the earth is (approximately) a sphere, this distance is not a straight line, but the shortest path between the cities along the surface of the earth, i.e. an arc of a great circle. Figure B.2 shows the great circle distances between Berlin and Buenos Aires and between Moscow and Seoul on a Plate Carée projection of the earth, and Panels (a) and (b) in Figure B.3 show the same distances on the surface of a spherical earth.[6]

In general, the great circle distance between two points i and j can be calculated by using the spherical law of cosines (Shekhar and Xiong, 2008, 639-642)

$$d_{ij} = R_{\oplus} \times \arccos[\cos lat_i \cos lat_j \cos(long_i - long_j) + \sin lat_i \sin lat_j] \tag{B.5}$$

where $R_{\oplus} = 6{,}378.1$km is the (rounded value of the) earth's equatorial radius (see Ahrens (1995, 36)), and the values for latitude and longitude are in decimal degrees.

[5]The name follows from the movement of the queen on the chessboard. Other criteria for determining spatial weights, whose names have the same origin are bishop and rook contiguity (see, for instance, (Anselin, 1988b, 18)).

[6]These figures were drawn with ArcGIS.

Figure B.2: Great Circle Distances between Berlin and Buenos Aires and between Moscow and Seoul on a Plate Carée Projection of the Earth.

(a) Great Circle Distance between Berlin and Buenos Aires on the Surface of a Spherical Earth.

(b) Great Circle Distance between Moscow and Seoul on the Surface of a Spherical Earth.

Figure B.3: Great Circle Distances on the Surface of a Spherical Earth.

The geographic coordinates for the four capitals under consideration are as follows:

Berlin: N52°31′, E13°24′

Buenos Aires: S34°35′, W58°40′

Moscow: N55°45′, E37°36′

Seoul: N37°33′, E126°59′

These coordinates are taken from the CIA's World Factbook (Central Intelligence Agency, 2013) and before inputting them into the formula above they need to be converted from degree-minute format to decimal degrees by dividing the entry for minutes by 60 and adding it to the value for the degree (see Peterson and Smith (2012, 458)). In decimal-degree form south and west are denoted with negative values (Shekhar and Xiong, 2008, 639) so that the coordinates from above now read[7]

Berlin: $52.52°$, $13.4°$

Buenos Aires: $-34.58°$, $-58.67°$

Moscow: $55.75°$, $37.6°$

Seoul: $37.55°$, $126.98°$

The distance between Berlin and Buenos Aires, for instance, can then be calculated as

$$\begin{aligned} & d_{B,BA} = R_{\oplus} \times \arccos\left[\cos(52.52°)\cos(-34.58°)\cos(13.4° - (-58.68°))\right. \\ & \qquad\qquad \left. + \sin(52.52°)\sin(-34.58°)\right] \\ \iff \quad & d_{B,BA} = R_{\oplus} \times 107.2356°. \end{aligned}$$

Converting now from degrees to radians by multiplying the angle with $\pi/180$ gives the distance between Berlin and Buenos Aires as

$$d_{B,BA} = 6{,}378.1\text{km} \cdot 107.2356° \cdot \frac{\pi}{180} = 6{,}378.1\text{km} \cdot 1.8716 = 11{,}937.25\text{km}. \tag{B.6}$$

The complete spatial weight matrix for the four capitals thus reads

$$\boldsymbol{W_C} = \begin{array}{c} \\ \text{Berlin} \\ \text{Buenos Aires} \\ \text{Moscow} \\ \text{Seoul} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Berlin} & \text{Buenos Aires} & \text{Moscow} & \text{Seoul} \end{array} \\ \begin{pmatrix} 0 & 11\,936 & 1{,}610 & 8{,}138 \\ 11{,}936 & 0 & 13{,}505 & 19{,}431 \\ 1{,}610 & 13{,}505 & 0 & 6{,}616 \\ 8{,}138 & 19{,}431 & 6{,}616 & 0 \end{pmatrix} \end{array}. \tag{B.7}$$

The difference between the value in Equation (B.6) and the corresponding value in Equation (B.7) stems from rounding the results of the trigonometric functions in the calculation above. In the matrix, the values have been calculated by implementing the formula in

[7] These conversions can easily be done in Mathematica or Matlab using the functions FromDMS and dm2degrees, respectively. The values here are rounded to two decimal points.

Matlab directly. After row-standardizing the matrix, it could be used in a straightforward manner in the econometric exercise in the main text.

Note that as an alternative to the equatorial radius the mean radius of the earth could have been used in the calculation. In its Earth Fact Sheet[8] the National Aeronautic and Space Administration (NASA) gives a value of 6,371km for the mean radius. Substituting this value in Equation (B.5) would not change the relative distances between capitals, however.

It should as well be kept in mind that the distances have been calculated by assuming the earth is a sphere, although it is better described by an oblate spheroid and hence, for instance, Vincenty's formulae would be more accurate (Vincenty, 1975). For the distances considered here, the gain in accuracy is negligible though and using a spatial weights matrix based on Vincenty's formulae would not change the qualitative results in the main text.

B.3 Proof that $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$ exists

This appendix demonstrates that the inverse of $(\boldsymbol{I}-\gamma\boldsymbol{W})$ exists for the assumed parameter space of γ, given that $\gamma \neq 0$ and that $\frac{1}{\gamma}$ is not an eigenvalue of $\boldsymbol{W}$. The first condition is obvious as it simply posits the existence of spatial externalities.[9] In general, $(\boldsymbol{I} - \gamma\boldsymbol{W})$ will have an inverse, if it is non-singular, implying that $|\boldsymbol{I} - \gamma\boldsymbol{W}| \neq 0$. The matrix will thus be singular and have no inverse if $|\boldsymbol{I}-\gamma\boldsymbol{W}| = 0$. Applying the rules for determinants, (see, for example, Sydsæter et al., 2008, 5) this expression is equivalent to

$$\left|\frac{1}{\gamma}\boldsymbol{I} - \boldsymbol{W}\right| = 0 \qquad \Longleftrightarrow \qquad \left|\boldsymbol{W} - \frac{1}{\gamma}\boldsymbol{I}\right| = 0.$$

The second equation is the characteristic or eigenvalue equation of $\boldsymbol{W}$, demonstrating that $(\boldsymbol{I} - \gamma\boldsymbol{W})$ will not have an inverse if $\frac{1}{\gamma}$ is an eigenvalue of $\boldsymbol{W}$.

Having established the general conditions under which the inverse exists, it will now be shown that it exists for $0 \leq \gamma < 1$.[10]

The last step in the proof will use the result that the eigenvalues of $\boldsymbol{W}$ will be less than or equal to 1 in absolute value. This result is now proved via Gerschgorin's Circle Theorem.[11]

[8] This is available under: `http://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html` (accessed: 11 August, 2015).

[9] For $\gamma = 0$, the model would reduce to a standard Solow model with physical externalities.

[10] It will actually be shown that the inverse exists for $|\gamma| < 1$. This naturally includes the parameter space described by the inequality in the main text, where one could exclude $\gamma = 0$.

[11] The original statement is due to Gerschgorin (1931). Here, I rely on the presentations in Meyer (2000, 498) and Cheney and Kincaid (2008, 347-349).

This theorem states that the eigenvalues of a matrix $\boldsymbol{B} \in \mathbb{C}^{n \times n}$ lie in the complex plane in the area that is given by the intersection of the union of all Gerschgorin circles associated with the rows of $\boldsymbol{B}$ and the union of all Gerschgorin cirlces associated with the columns of $\boldsymbol{B}$. Formally, the Gerschgorin circles (or discs), $\mathcal{G}_i^r$, associated with the rows are given by

$$\mathcal{G}_i^r = \{z \in \mathbb{C} : |z - b_{ii}| \leq r_i\}, \quad \text{where} \quad r_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}| \quad \text{for} \quad i = 1, 2, \ldots, n$$

which means that the circles have the entry b_{ii} of the matrix $\boldsymbol{B}$ as their center and the sum of the absolute values of the off-diagonal entries of the respective row as their radius. The eigenvalues of the matrix are then contained in the union of these n Gerschgorin circles associated with the rows of $\boldsymbol{B}$, i.e. in $\bigcup_{i=1}^{N} \mathcal{G}_i^r$.

The Gerschgorin circles associated with the columns, $\mathcal{G}_i^c$, are given by[12]

$$\mathcal{G}_j^c = \{z \in \mathbb{C} : |z - b_{ii}| \leq c_j\}, \quad \text{where} \quad c_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} |b_{ij}| \quad \text{for} \quad j = 1, 2, \ldots, n$$

and the union of these n Gerschgorin circles is denoted by $\bigcup_{i=1}^{N} \mathcal{G}_i^c$. Hence, the eigenvalues of $\boldsymbol{B}$ will be contained in the following intersection

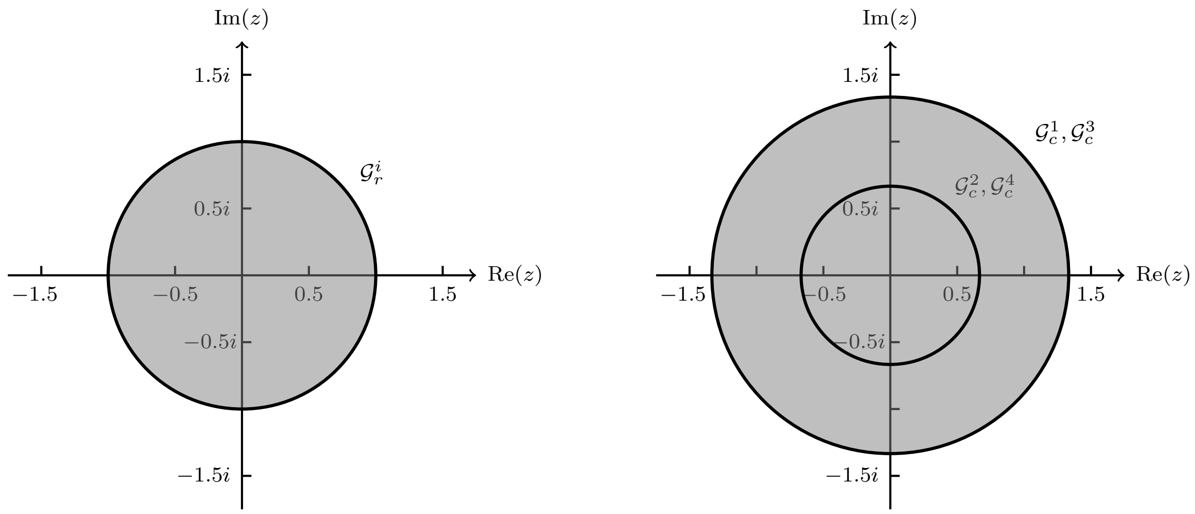
$$\left(\bigcup_{i=1}^{N} \mathcal{G}_i^r \right) \bigcap \left(\bigcup_{i=1}^{N} \mathcal{G}_i^c \right).$$

Before applying this theorem to the spatial weight matrix $\boldsymbol{W}$ from the main text, a brief graphical illustration for the matrix in (B.4) from Appendix B.2 will be provided to deepen the understanding of the theorem.

For the matrix $\boldsymbol{W_B}$, all Gerschgorin circles are centered around the point $(0, 0)$ in the complex plane, and since the matrix is row standardized all circles associated with the rows have a radius of 1. The union of these circles is shown in Panel (a) of Figure B.4. Naturally, the circles associated with the columns are also centered around $(0, 0)$ and, since $c_1 = c_3 = 4/3$ and $c_2 = c_4 = 2/3$, there are in effect only two circles for the columns, which have radii of 4/3 and 2/3, respectively. These circles and their union are depicted in Panel (b) of Figure B.4. Finally, Panel (c) of Figure B.4 overlays the two results, showing that all eigenvalues[13] (which are depicted with a white circle in the figure) will be contained within the unit circle.

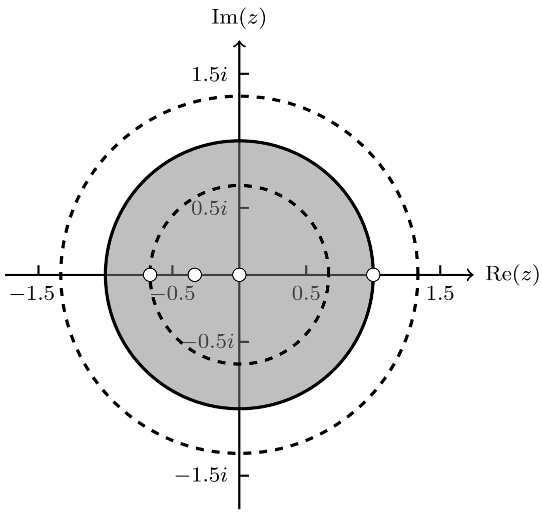
[12]That the eigenvalues of $\boldsymbol{B}$ also are contained in the circles associated with the columns follows, since the calculation of the eigenvalues involves the determinant, which is identical for a matrix and its transpose (Meyer, 2000, 463).

[13]The eigenvalues of $\boldsymbol{W_B}$ are $\lambda_1 = 1, \lambda_2 = -1/3, \lambda_3 = -2/3$, and $\lambda_4 = 0$.



(a) Union of Gerschgorin Circles Associated with the Rows of $\boldsymbol{W_B}$, i.e. $\bigcup_{i=1}^{N} \mathcal{G}_i^r$.

(b) Union of Gerschgorin Circles Associated with the Columns of $\boldsymbol{W_B}$, i.e. $\bigcup_{i=1}^{N} \mathcal{G}_i^c$.



(c) Area in which the Eigenvalues of the Matrix $\boldsymbol{W_B}$ are Contained, i.e. $\left(\bigcup_{i=1}^{N} \mathcal{G}_i^r\right) \cap \left(\bigcup_{i=1}^{N} \mathcal{G}_i^c\right)$.

Figure B.4: Illustration of Gerschgorin's Circle Theorem for the Matrix $\boldsymbol{W_B}$.

This brief illustration provides an insight into why the eigenvalues λ_i of the matrix $\boldsymbol{W}$ from the main text will be equal to or less than 1 in absolute value. The result hinges on the assumption that the spatial weight matrix is row standardized so that $r_i = 1$ for $i = 1, \ldots, N$, implying that $|\lambda_i| \leq 1$ will hold for the eigenvalues.

In order to finally show that $(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}$ exists a last intermediate result is helpful. From Schur's Triangularization Theorem it follows that via a similarity transformation[14] every

[14]Two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar "whenever there exists a nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{B}$. The product $\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{B}$ is called a **similarity transformation** on $\boldsymbol{A}$" (Meyer, 2000, 506, emphasis in the original).

square matrix can be made upper triangular (Meyer, 2000, 508). Hence, an invertible $N \times N$ matrix $\boldsymbol{P}$ exists so that

$$\boldsymbol{P}^{-1}\boldsymbol{W}\boldsymbol{P} = \boldsymbol{T}, \qquad \text{with} \qquad \boldsymbol{T} = \begin{pmatrix} t_{11} & \cdots & t_{1N} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_{NN} \end{pmatrix}$$

where as an implication of Schur's Triangularization Theorem the eigenvalues of $\boldsymbol{W}$ are the diagonal entries of the matrix $\boldsymbol{T}$. This matrix will now be used to prove that $\boldsymbol{I} - \gamma\boldsymbol{W}$ is non-singular.

Proof.

$$\begin{aligned} & |\boldsymbol{I}-\gamma\boldsymbol{W}| = |\boldsymbol{P}\boldsymbol{P}^{-1}(\boldsymbol{I}-\gamma\boldsymbol{W})| = |\boldsymbol{P}(\boldsymbol{I}-\gamma\boldsymbol{W})\boldsymbol{P}^{-1}| \\ \Longleftrightarrow \quad & |\boldsymbol{I}-\gamma\boldsymbol{W}| = |\boldsymbol{P}\boldsymbol{I}\boldsymbol{P}^{-1} - \gamma\boldsymbol{P}\boldsymbol{W}\boldsymbol{P}^{-1}| = |\boldsymbol{I}-\gamma\boldsymbol{T}| \\ \Longleftrightarrow \quad & |\boldsymbol{I}-\gamma\boldsymbol{W}| = (1-\gamma t_{11})\cdots(1-\gamma t_{NN}) \qquad \text{(B.8)} \\ \Longleftrightarrow \quad & |\boldsymbol{I}-\gamma\boldsymbol{W}| \neq 0 \end{aligned}$$

The last line follows if $|\gamma\lambda_{ii}| \neq 1$, which holds since $t_{ii} = \lambda_i$ and $|\lambda_{ii}| \leq 1$ from Gerschgorin's Theorem, and also $|\gamma| < 1$ holds. □

Here, the product rule on determinants as well as the fact that $|\boldsymbol{P}\boldsymbol{P}^{-1} = 1|$ (see Meyer, 2000, 508) has been used, and to obtain Equation (B.8) the rule for the determinant of a triangular matrix has been employed (see Meyer, 2000, 462).

An implication of applying Gerschgorin's Theorem in this case is that it rules out that $\frac{1}{\gamma}$ is an eigenvalue of $\boldsymbol{W}$. This follows since it has been established that the eigenvalues of $\boldsymbol{W}$ are in the interval $[-1, 1]$ and $|\gamma| < 1$ so that the hypothetical eigenvalue $\frac{1}{\gamma}$ would be larger than 1.

B.4 Derivations and Proofs of Selected Model Results

In this appendix the expressions for the spatial multiplier and the elasticities are derived.

B.4.1 Derivation of Equation (2.6)

This appendix proves that $(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}$ equals $\sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r$ and uses the result to derive the expression for the level of technology in country i in Equation (2.6) in Section 2.3.2. The crucial step in this proof will be to show that $\lim_{n\to\infty}(\gamma\boldsymbol{W})^n = \boldsymbol{0}$. If this holds, then $\boldsymbol{I} - \gamma\boldsymbol{W}$ has an inverse, thereby providing an alternative proof for its existence, and it follows that this inverse can be written as $\sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r$.[15]

Proof. As a first step, the following result is helpful

$$(\boldsymbol{I} - \gamma\boldsymbol{W})\left(\boldsymbol{I} + \gamma\boldsymbol{W} + (\gamma\boldsymbol{W})^2 + \cdots + (\gamma\boldsymbol{W})^n\right) = \boldsymbol{I} - (\gamma\boldsymbol{W})^n$$

where, if $\lim_{n\to\infty}(\gamma\boldsymbol{W})^n = \boldsymbol{0}$ the right-hand side tends to $\boldsymbol{I}$ as $n \to \infty$. Left multiplying this equation by $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}$ then leads to the Neumann series

$$\boldsymbol{I} + \gamma\boldsymbol{W} + (\gamma\boldsymbol{W})^2 + \cdots + (\gamma\boldsymbol{W})^n = \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r = (\boldsymbol{I} - \gamma\boldsymbol{W})^{-1}.$$

It remains to be shown that $\lim_{n\to\infty}(\gamma\boldsymbol{W})^n = \boldsymbol{0}$ indeed holds. This is equivalent to the statement that the spectral radius of the matrix $\gamma\boldsymbol{W}$ is strictly smaller than 1 (see Meyer, 2000, 618). Since the spectral radius of a matrix is given by its largest eigenvalue in absolute value (Meyer, 2000, 497), a straightforward application of Gerschgorin's Circle Theorem to the matrix $\gamma\boldsymbol{W}$ shows that its spectral radius is smaller than 1. This follows, since for the matrix $\boldsymbol{W}$ the largest eigenvalue is 1, and multiplying each matrix entry by $|\gamma| < 1$ would reduce the radii of the Gerschgorin circles. □

The expression $(\boldsymbol{I} - \gamma\boldsymbol{W})^{-1} = \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r$ is also referred to as the spatial multiplier (Ertur and Koch, 2007, 1044) and using this result in Equation (2.5) leads to

$$\boldsymbol{A} = \sum_{r=0}^{\infty} \lambda^r \boldsymbol{W}^r \boldsymbol{\Omega} + \phi \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r \boldsymbol{k}.$$

[15] This is called a Neumann Series and can be used to approximate the inverse (Meyer, 2000, 126).

This equation can be simplified by repeatedly substituting the result $\boldsymbol{W\Omega} = \boldsymbol{\Omega}$.[16] The equation then reads

$$\boldsymbol{A} = \frac{1}{1-\gamma} \cdot \boldsymbol{\Omega} + \phi \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r \boldsymbol{k}.$$

The last term on the right-hand side can be shown to equal

$$\phi \sum_{r=0}^{\infty} \gamma^r \boldsymbol{W}^r \boldsymbol{k} = \phi \begin{pmatrix} \ln k_1(t) \\ \vdots \\ \ln k_N(t) \end{pmatrix} + \begin{pmatrix} \prod_{j=1}^{N} \ln k_j(t)^{\phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W}^r)_{1j}} \\ \vdots \\ \prod_{j=1}^{N} \ln k_j(t)^{\phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W}^r)_{Nj}} \end{pmatrix}$$

so that after first collecting the terms in logarithms and then applying the exponential transformation, the level of technology for a given country i is given by

$$A_i(t) = \Omega(t)^{\frac{1}{1-\gamma}} \cdot k_i(t)^{\phi} \cdot \prod_{j=1}^{N} k_j(t)^{\phi \sum_{r=1}^{\infty} \gamma^r (\boldsymbol{W}^r)_{ij}}$$

which is Equation (2.6) in Section 2.3.2.

B.4.2 Derivation of the Elasticities

Define $\boldsymbol{S}$ as the $N \times 1$ vector of investment rates, s_i, in logarithms and $\boldsymbol{N}$ as the $N \times 1$ vector of the effective depreciation rates, $n_i+g+\delta$, also in logarithms, then Equation (2.14) can be rewritten in matrix form as

$$\begin{aligned} \boldsymbol{y} = & \frac{1}{1-\alpha-\phi}\boldsymbol{\Omega} + \frac{\alpha+\phi}{1-\alpha-\phi}\boldsymbol{S} - \frac{\alpha+\phi}{1-\alpha-\phi}\boldsymbol{N} \\ & - \frac{\alpha\gamma}{1-\alpha-\phi}\boldsymbol{WS} - \frac{\alpha\gamma}{1-\alpha-\phi}\boldsymbol{WN} + \frac{\gamma(1-\alpha)}{1-\alpha-\phi}\boldsymbol{Wy}. \end{aligned}$$

Solving this equation for $\boldsymbol{y}$, yields

$$\begin{aligned} \boldsymbol{y} = & \frac{1}{1-\alpha-\phi}\left[\boldsymbol{I} - \frac{\gamma(1-\alpha)}{1-\alpha-\phi}\boldsymbol{W}\right]^{-1}\boldsymbol{\Omega} \\ & + \left[\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\boldsymbol{W}\right]^{-1}\left(\frac{\alpha+\phi}{1-\alpha-\phi}\boldsymbol{I} - \frac{\alpha\gamma}{1-\alpha-\phi}\boldsymbol{W}\right)\boldsymbol{S} \\ & + \left[\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\boldsymbol{W}\right]^{-1}\left(\frac{\alpha+\phi}{1-\alpha-\phi}\boldsymbol{I} - \frac{\alpha\gamma}{1-\alpha-\phi}\boldsymbol{W}\right)\boldsymbol{N} \end{aligned}$$

[16]That this holds can be seen by writing out the details of the matrix multiplication and then using the assumption that $\boldsymbol{W}$ is row standardized so that $\sum_{j=1}^{N} w_{1j} = 1$.

and taking now the derivative with respect to $\boldsymbol{S}$ leads to a matrix for the elasticities of steady-state income with respect to the investment rate

$$\boldsymbol{\eta}_s = \frac{\alpha + \phi}{1 - \alpha - \phi}\boldsymbol{I} + \left[\frac{\phi}{(1-\alpha)(1-\alpha-\phi)}\right]\sum_{r=1}^{\infty}\boldsymbol{W}^r\left(\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\right)^r = -\boldsymbol{\eta}_n$$

where $\boldsymbol{\eta}_n$ denotes the corresponding matrix of elasticities with respect to the vector of effective depreciation rates $\boldsymbol{N}$. The elasticities given in Section 2.3.3 in Equations (2.15) and (2.16) for a country i then follow directly from the equation above.

B.5 Further Robustness Checks

This appendix gathers detailed estimation results for a series of specifications mentioned in the main text. The results in Tables B.1 and B.2 demonstrate that the original results by Ertur and Koch are not robust across different versions of the Penn World Table based on the specification for the interaction matrix $\boldsymbol{W_2}$ as actually implemented in their estimation. Tables B.3 and B.4 show that the estimation results are highly sensitive to division of the geographic distances between country capitals by 1,000 in the weights of interaction matrix $\boldsymbol{W_2}$. Finally, Tables B.5 and B.6 depict the results when the weights in the interaction matrix using genetic distances between countries are based on weighted Nei's genetic distance. In this case, the estimation results are not robust across the different samples, but comparable to the ones based on weighted F_{ST} distance in the main text with the exception mentioned in Footnote 101.

Table B.1: Estimation Results for the Standard and Spatially Augmented Solow Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_2}$ (Geographic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Unconstrained estimation:				
Constant	0.546	0.214	3.042	1.139
	(0.771)	(0.911)	(0.135)	(0.586)
$\ln s_i$	0.804	0.757	0.836	0.936
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.471	−1.263	−0.496	−1.094
	(0.08)	(0.030)	(0.479)	(0.146)
$\boldsymbol{W} \ln s_j$	−0.381	−0.370	−0.132	0.152
	(0.021)	(0.031)	(0.530)	(0.669)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.158	−0.145	−0.595	−0.623
	(0.840)	(0.856)	(0.500)	(0.520)
$\boldsymbol{W} \ln y_j$	0.657	0.659	0.516	0.577
	(0.000)	(0.000)	(0.000)	(0.000)
Constrained estimation:				
Constant	2.769	3.022	4.086	2.649
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_i - \ln n_i$	0.826	0.791	0.839	0.970
	(0.000)	(0.000)	(0.000)	(0.000)
$\boldsymbol{W}[\ln s_j - \ln(n_j + 0.05)]$	−0.318	−0.232	−0.085	0.252
	(0.045)	(0.174)	(0.679)	(0.457)
$\boldsymbol{W} \ln y_j$	0.665	0.635	0.510	0.578
	(0.000)	(0.000)	(0.000)	(0.000)
Test of restriction	2.378	2.090	0.652	0.543
	(0.305)	(0.352)	(0.722)	(0.762)
Implied α	0.323	0.268	0.143	−0.773
	(0.001)	(0.042)	(0.612)	(0.695)
Implied ϕ	0.129	0.174	0.313	1.265
	(0.126)	(0.138)	(0.235)	(0.519)
Implied γ	0.538	0.484	0.324	0.166
	(0.000)	(0.000)	(0.026)	(0.406)
$\alpha + \frac{\phi}{1-\gamma}$	0.603	0.605	0.606	0.743
	(0.000)	(0.000)	(0.000)	(0.000)

Note: p-values are given in parentheses. The restriction for the spatially augmented model is tested with the likelihood ratio (LR) test.

Table B.2: Estimation Results for the Direct, Indirect and Total Impacts in the Spatial Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_2}$ (Geographic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Direct impacts:				
$\ln s_i$	0.847	0.799	0.897	1.091
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.709	−1.583	−0.667	−1.378
	(0.004)	(0.012)	(0.332)	(0.064)
Indirect impacts:				
$\boldsymbol{W} \ln s_j$	0.388	0.337	0.562	1.487
	(0.286)	(0.386)	(0.049)	(0.037)
$\boldsymbol{W} \ln(n_j + 0.05)$	−2.128	−2.561	−1.588	−2.696
	(0.244)	(0.168)	(0.252)	(0.110)
Total impacts:				
$\ln s_i + \boldsymbol{W} \ln s_j$	1.236	1.136	1.459	2.578
	(0.004)	(0.014)	(0.000)	(0.003)
$\ln(n_i + 0.05) + \boldsymbol{W} \ln(n_j + 0.05)$	−3.837	−4.143	−2.255	−4.073
	(0.066)	(0.055)	(0.143)	(0.029)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

Table B.3: Estimation Results for the Spatially Augmented Solow Model According to the Specification of $\boldsymbol{W_2}$ as Implemented (Column 1) vs. as Claimed (Column 2) in Ertur and Koch (2007).

Data set	PWT 6.1	
Weight in the interaction matrix $\boldsymbol{W_2}$	$e^{-2d_{ij}/1{,}000}$	$e^{-2d_{ij}}$
Number of observations	91	91
Unconstrained estimation:		
Constant	0.546	5.140
	(0.771)	(0.005)
$\ln s_i$	0.804	1.238
	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.471	−2.475
	(0.008)	(0.000)
$\boldsymbol{W} \ln s_j$	−0.381	0.826
	(0.021)	(0.004)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.158	−1.570
	(0.840)	(0.012)
$\boldsymbol{W} \ln y_j$	0.657	−0.236
	(0.000)	(0.073)
Constrained estimation:		
Constant	2.769	8.884
	(0.000)	(0.000)
$\ln s_i - \ln n_i$	0.826	1.093
	(0.000)	(0.000)
$\boldsymbol{W}[\ln s_j - \ln(n_j + 0.05)]$	−0.318	2.021
	(0.045)	(0.000)
$\boldsymbol{W} \ln y_j$	0.665	−0.236
	(0.000)	(0.000)
Implied α	0.323	0.895
	(0.001)	(0.000)
Implied ϕ	0.129	−0.373
	(0.126)	(0.000)
Implied γ	0.538	−1.078
	(0.000)	(0.000)
$\alpha + \frac{\phi}{1-\gamma}$	0.603	0.716
	(0.000)	(0.000)

Note: p-values are given in parentheses. The likelihood ratio (LR) could not be performed for this matrix, as no value for the log-likelihood was returned in this model.

Table B.4: Estimation Results for the Direct, Indirect and Total Impacts in the Spatial Model According to the Specification of $\boldsymbol{W_2}$ as Implemented (Column 1) vs. as Claimed (Column 2) in Ertur and Koch (2007).

Data set	PWT 6.1	
Weight in the interaction matrix $\boldsymbol{W_2}$	$e^{-2d_{ij}/1{,}000}$	$e^{-2d_{ij}}$
Number of observations	91	91
Direct impacts:		
$\ln s_i$	0.847	1.210
	(0.000)	(0.000)
$\ln(n_i + 0.05)$	-1.709	-2.412
	(0.004)	(0.001)
Indirect impacts:		
$\boldsymbol{W} \ln s_j$	0.388	0.463
	(0.286)	(0.033)
$\boldsymbol{W} \ln(n_j + 0.05)$	-2.128	-0.850
	(0.244)	(0.001)
Total impacts:		
$\ln s_i + \boldsymbol{W} \ln s_j$	1.236	1.673
	(0.004)	(0.000)
$\ln(n_i + 0.05) + \boldsymbol{W} \ln(n_j + 0.05)$	-3.837	-3.262
	(0.066)	(0.000)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

Table B.5: Estimation Results for the Spatial Durbin Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_4}$ (Weighted Nei's Genetic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Unconstrained estimation:				
Constant	8.023	8.745	5.864	−2.162
	(0.003)	(0.000)	(0.013)	(0.373)
$\ln s_i$	0.847	0.962	0.901	1.005
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.091	−0.857	−0.137	−0.892
	(0.047)	(0.115)	(0.806)	(0.172)
$\boldsymbol{W} \ln s_j$	0.725	0.728	0.739	−0.027
	(0.003)	(0.001)	(0.003)	(0.949)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.595	0.396	−1.747	−2.041
	(0.550)	(0.676)	(0.074)	(0.065)
$\boldsymbol{W} \ln y_j$	0.330	0.285	0.181	0.557
	(0.006)	(0.015)	(0.184)	(0.000)
Constrained estimation:				
Constant	5.407	5.750	6.119	2.325
	(0.000)	(0.000)	(0.000)	(0.001)
$\ln s_i - \ln n_i$	0.812	0.898	0.884	1.031
	(0.000)	(0.000)	(0.000)	(0.000)
$\boldsymbol{W}[\ln s_j - \ln(n_j + 0.05)]$	0.668	0.688	0.671	0.047
	(0.005)	(0.001)	(0.004)	(0.905)
$\boldsymbol{W} \ln y_j$	0.308	0.262	0.232	0.624
	(0.019)	(0.024)	(0.070)	(0.000)
Test of restriction	1.710	1.857	2.082	3.945
	(0.425)	(0.395)	(0.353)	(0.139)
Implied α	1.855	1.615	1.529	−0.081
	(0.105)	(0.023)	(0.030)	(0.913)
Implied ϕ	−1.407	−1.141	−1.060	0.588
	(0.221)	(0.115)	(0.138)	(0.417)
Implied γ	−0.199	−0.225	−0.233	0.284
	(0.294)	(0.158)	(0.207)	(0.188)
$\alpha + \frac{\phi}{1-\gamma}$	0.681	0.683	0.669	0.741
	(0.000)	(0.000)	(0.000)	(0.000)

Note: p-values are given in parentheses. The restriction for the spatially augmented model is tested with the likelihood ratio (LR) test.

Table B.6: Estimation Results for the Direct, Indirect and Total Impacts in the Spatial Model According to Three Different Versions of the Penn World Table Based on Interaction Matrix $\boldsymbol{W_4}$ (Weighted Nei's Genetic Distance).

Data set	PWT 6.1		PWT 6.2	PWT 7.1
Number of observations	91	83	83	83
Direct impacts:				
$\ln s_i$	0.909	1.011	0.923	1.074
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(n_i + 0.05)$	−1.071	−0.843	−0.185	−1.227
	(0.050)	(0.118)	(0.741)	(0.063)
Indirect impacts:				
$\boldsymbol{W} \ln s_j$	1.454	1.365	1.080	1.122
	(0.000)	(0.000)	(0.000)	(0.152)
$\boldsymbol{W} \ln(n_j + 0.05)$	0.390	0.254	−2.101	−5.462
	(0.776)	(0.834)	(0.046)	(0.008)
Total impacts:				
$\ln s_i + \boldsymbol{W} \ln s_j$	2.363	2.376	2.001	2.195
	(0.000)	(0.000)	(0.000)	(0.011)
$\ln(n_i + 0.05) + \boldsymbol{W} \ln(n_j + 0.05)$	−0.681	−0.589	−2.286	−6.689
	(0.629)	(0.630)	(0.026)	(0.003)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

B.6 List of Countries Included in the Empirical Analyses

This appendix lists the countries that are included in the empirical analyses. Country codes are given as well. In the analyses with 83 countries, Angola, Bangladesh, Botswana, Central African Republic, Mauritania, Papua New Guinea, Sierra Leone, and Democratic Republic of Congo have been dropped to achieve a balanced sample over PWT Versions 6.1, 6.2, and 7.1.

Table B.7: Alphabetical List of the 91 Countries from PWT 6.1 Included in the Empirical Analyses.

Country	Code	Country	Code
Angola	AGO	Mali	MLI
Argentina	ARG	Mauritania	MRT
Australia	AUS	Mauritius	MUS
Austria	AUT	Mexico	MEX
Bangladesh	BGD	Morocco	MAR
Belgium	BEL	Mozambique	MOZ
Benin	BEN	Nepal	NPL
Bolivia	BOL	Netherlands	NLD
Botswana	BWA	New Zealand	NZL
Brazil	BRA	Nicaragua	NIC
Burkina Faso	BFA	Niger	NER
Burundi	BDI	Nigeria	NGA
Cameroon	CMR	Norway	NOR
Canada	CAN	Pakistan	PAK
Central African Republic	CAF	Panama	PAN
Chad	TCD	Papua New Guinea	PNG
Chile	CHL	Paraguay	PRY
Colombia	COL	Peru	PER
Costa Rica	CRI	Philippines	PHL
Côte d'Ivoire	CIV	Portugal	PRT
Democratic Republic of the Congo	ZAR	Republic of the Congo	COG
Denmark	DNK	Republic of Korea	KOR
Dominican Republic	DOM	Rwanda	RWA
Ecuador	ECU	Senegal	SEN
Egypt	EGY	Sierra Leone	SLE
El Salvador	SLV	Singapore	SGP
Ethiopia	ETH	South Africa	ZAF
Finland	FIN	Spain	ESP
France	FRA	Sri Lanka	LKA
Ghana	GHA	Sweden	SWE
Greece	GRC	Switzerland	CHE
Guatemala	GTM	Syria	SYR
Honduras	HND	Tanzania	TZA

Table B.7: *(Continued)*

Country	Code	Country	Code
Hong Kong	HKG	Thailand	THA
India	IND	Togo	TGO
Indonesia	IDN	Trinidad and Tobago	TTO
Ireland	IRL	Tunisia	TUN
Israel	ISR	Turkey	TUR
Italy	ITA	Uganda	UGA
Jamaica	JAM	United Kingdom	GBR
Japan	JPN	United States	USA
Jordan	JOR	Uruguay	URY
Kenya	KEN	Venezuela	VEN
Madagascar	MDG	Zambia	ZMB
Malawi	MWI	Zimbabwe	ZWE
Malaysia	MYS		

C Appendix to Chapter 3

C.1 Poisson Processes

In the literature on Schumpeterian growth models it is standard to model the occurrence of an innovation via a Poisson arrival rate or to read about Poisson processes (compare, for instance, Aghion and Howitt (1992) and Aghion et al. (2014)).[1] A detailed exposition of these notions is however seldom provided so that these concepts from statistical theory may pose some difficulties at first glance. This appendix therefore serves as a brief review of the general concepts concerning Poisson processes.

An important step towards understanding Poisson processes concerns the exponential distribution. If a continuous random variable X is exponentially distributed, then its probability density function (PDF) is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{C.1}$$

Here, λ is a parameter for which $\lambda > 0$, and x is a particular value of the random variable. This function is illustrated for two different values of λ in Figure C.1.

The cumulative distribution function, which gives the probability that $X \leq x$, with $x \geq 0$ is then given by

$$P\{X \leq x\} = F(x) = \int_0^x f(\tau)\, d\tau = \int_0^x \lambda e^{-\lambda \tau} d\tau = \Big|_0^x - e^{-\lambda \tau} = 1 - e^{-\lambda x}. \tag{C.2}$$

The exponential distribution has the property of being memoryless (Ross, 2010, 294). This property can formally be stated as

$$P\{X > s + t | X > t\} = P\{X > s\} \qquad \forall\, s, t \geq 0.$$

Interpreting the random variable X, for instance, as the lifetime of a certain machine, instrument, or device like a traffic light, the equation above states that the probability that a traffic light functions $s + t$ units of time (i.e. days), given that it already has

[1]Poisson processes also play an important role in other areas of economics like labor or monetary economics. Wälde (2011, 261) provides a brief list of applications in economics.

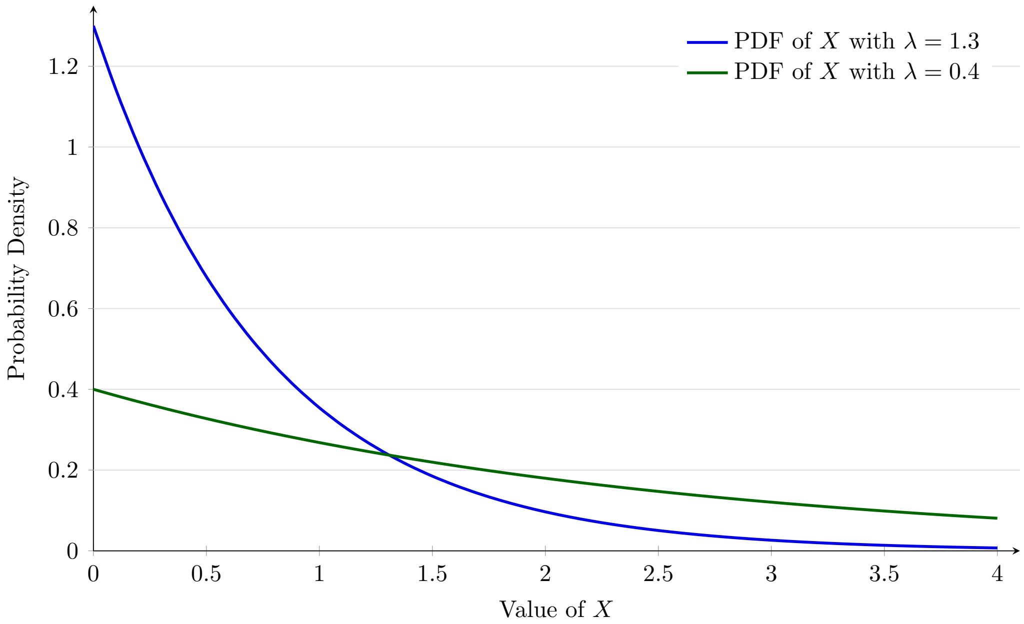


Figure C.1: Probability Density Functions (PDF) of an Exponentially Distributed Random Variable X with Parameter $\lambda = 1.3$ and $\lambda = 0.4$, respectively.

Note: The underlying data was generated in Mathematica.

worked for t days is the same as the unconditional probability that it works for s days. Put differently, the traffic light "does not remember" it has already worked for t days. A concept, known as the hazard rate or failure rate, helps to illustrate this property. For the given example, it is defined as the conditional probability that a traffic light, having survived t days, will fail. Formally, the failure rate, $r(t)$, is thus given by (Ross, 2010, 299)

$$r(t) = \frac{f(t)}{1 - F(t)}.$$

Inserting from Equations (C.1) and (C.2), one immediately sees that the failure rate is constant in the case of an exponentially distributed random variable[2]

$$r(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

The notion of failure rates will be picked up again after the following exposition on Poisson processes. These are a specific form of a counting process. Generally speaking, a counting process $\{N(t), t \geq 0\}$ is a stochastic process, which counts the number of events, $N(t)$, that have happened up until time t (like, for example, the number of cyclists who have crossed a certain bridge until noon). For a Poisson process, the following definition holds (Ross, 2010, 313): A Poisson process is a counting process with rate $\lambda > 0$, if (i) $N(0) = 0$,

[2]In fact, as Ross (2010, 299-300) demonstrates, the property of memorylessness exists only for random variables that are exponentially distributed.

(ii) the process has independent increments, and (iii) the number of events in any interval of length t has a Poisson distribution with mean λt, i.e. $\forall\, s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad n = 0, 1, \ldots$$

In the context of the model in the main text, for an individual about to engage in research and calculating the value of an innovation, information on the absolute number of innovations in a given sector is of minor interest compared to the time span over which she will be able to earn monopoly profits. Hence, information on the time between innovations is of central interest. Denoting the point of time of the first innovation as T_1 and defining T_n as the time span or interarrival time between event (or innovation) $n-1$ and n, implies that if, for example, innovation number 5 occurred at time 33, innovation 6 at time 34, and the next innovation at time 38, then one would have $T_6 = 34 - 33 = 1$ and $T_7 = 38 - 34 = 4$ as the values for the interarrival times. Information on the distribution of this sequence of random variables can now be derived by noting that the probability that the first event or innovation occurs after time t is given by the expression (Ross, 2010, 317)

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}.$$

Hence, T_1 is exponentially distributed. This result follows from using property (i) of the Poisson process and by noting that during the interval $[0, t]$ by definition no event occurs so that the number of events in this particular interval is $N(t+0) = N(t) = 0$.[3] Next, the probability of T_2, i.e. the probability that the time between events 1 and 2 is larger than t is the probability of T_2 given that T_1 already happened (which necessarily needs to be the case given the definition of T_2) and had an interarrival time of e.g. s. Then, it holds that

$$\begin{aligned} P\{T_2 > t | T_1 = s\} &= P\{0 \text{ events in } (s, s+t] | T_1 = s\} \\ &= P\{0 \text{ events in } (s, s+t]\} \\ &= e^{-\lambda t}, \end{aligned}$$

where the second line follows from the fact that the Poisson process has independent increments (i.e. the number of events that occur in non-overlapping intervals are independent from each other) so that the conditional and unconditional probabilities are identical. The third line holds, as the process has stationary increments,[4] implying that the distribution for the number of events in $(s, s+t)$ is identical for all s (Ross, 2010, 313).

[3]In general, it holds that $P\{X > x\} = 1 - P\{x \leq X\}$. Hence, with reference to the cumulative distribution function in Equation (C.2), the claim that T_1 has an exponential distribution is valid.

[4]This is implied by an alternative definition of a Poisson process to the one provided above. See Ross (2010, 314) for the details concerning this definition.

Having demonstrated that the interarrival times are independently and identically distributed, the failure rate for the interarrival times is therefore given by the parameter or intensity of the Poisson process, λ. Translated into the context of the model, a failure is equivalent to a new innovation, and the probability that a new innovation comes into existence during the interval dt is given by $r(t)dt = \lambda dt$ (Ross, 2010, 299).

Figures C.2 and C.3 illustrate important characteristics of Poisson processes and the exponential distribution with different values for λ. One clearly sees from the length of the horizontal lines in Figure C.2, that the time interval between innovations (or "failures") is not constant. Interpreting the units of time as years, it takes, for instance, only approximately three months to go from quality level 7 to level 10 (or come up with 3 additional products in that time span), whereas making the three steps from 3 to 6 takes approximately 5 years. Also, the number of absolute innovations is higher for the process with a higher value for λ (17 versus 4).

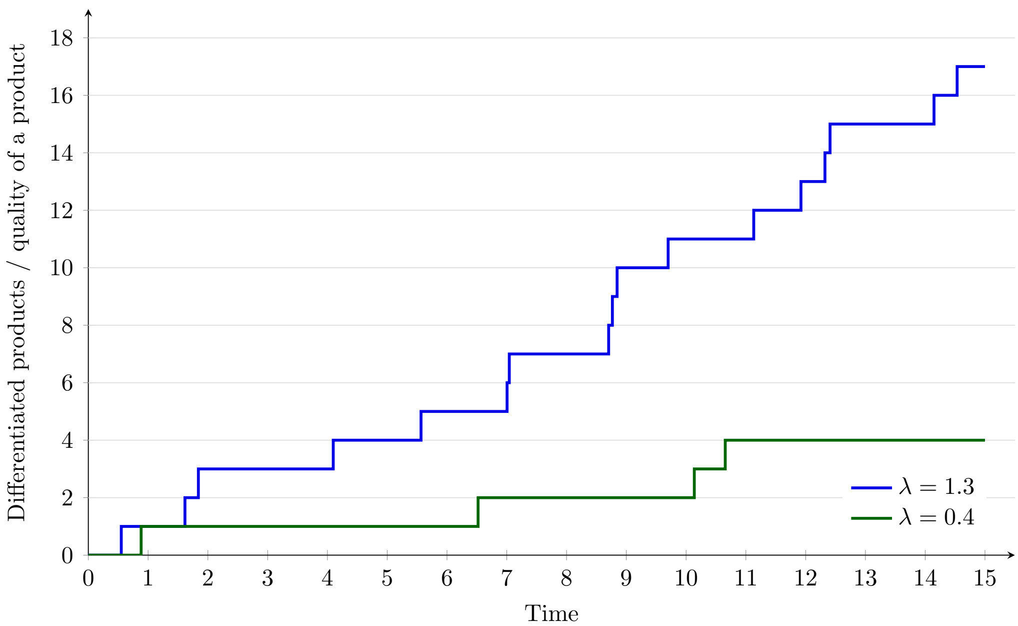


Figure C.2: Illustration of Two Poisson Processes with Intensities $\lambda = 1.3$ and 0.4, respectively.

Note: The data underlying the Poisson processes was generated in Mathematica.

Additionally, Figure C.3 illustrates that a higher value for λ is equivalent to having a larger probability mass at any value of the random variable. Therefore, the probability that an innovation occurs within a certain period of time is indeed higher for higher values of λ.

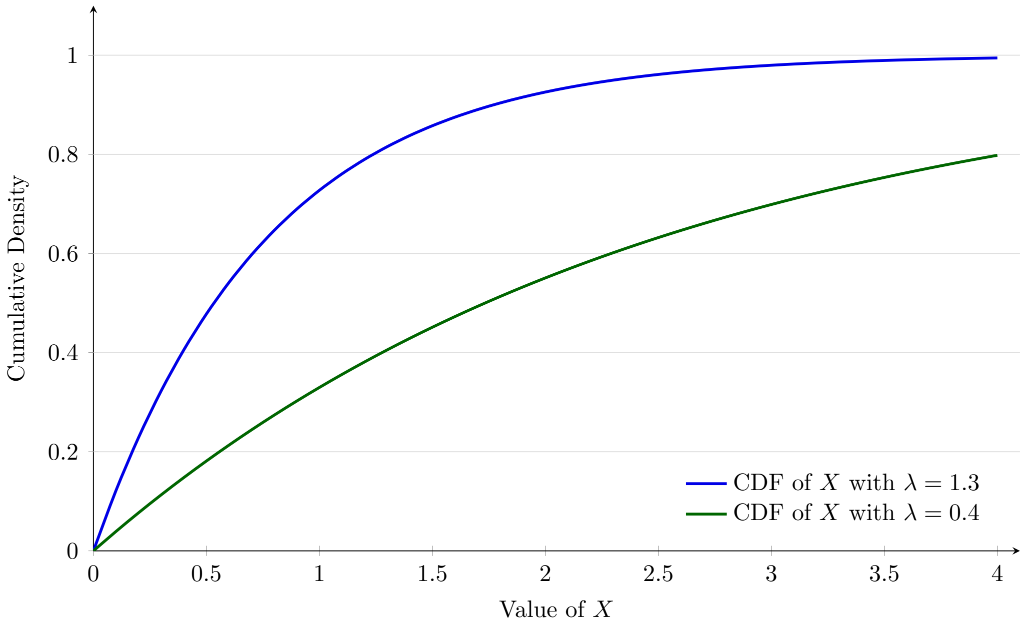


Figure C.3: Cumulative Density Functions (CDF) of an Exponentially Distributed Random Variable X with Parameter $\lambda = 1.3$ and $\lambda = 0.4$, respectively.

Note: The underlying data was generated in Mathematica.

C.2 Additional Derivations – Model

This appendix gathers a variety of derivations of (intermediate) results that are merely stated in the presentation of the model in Sections 3.2 and 3.3 in the main text.

C.2.1 Derivation of the Inverse Demand Schedule for Intermediate Goods

In the following, the inverse demand function for an intermediate good will be derived in detail.

The necessary condition for the maximization problem in Equation (3.2) is given by $\frac{d\Pi_i(v,t)}{dx_i(v,t)} = 0$. Calculating the derivative in this equation, requires an application of the following result for differentiating under the integral sign (see, for instance, Sydsæter et al., 2008, 159)

$$F(x) = \int_c^d f(x,t)\,dt \qquad \Longrightarrow \qquad F'(x) = \int_c^d \frac{\partial f(x,t)}{\partial x}\,dt.$$

Applying this general result to the problem in (3.2), leads to the necessary condition

$$Q_i^{\alpha-1} \int_0^{Q_i(t)} \alpha A_i(v,t) L_i(t)^{1-\alpha} x_i(v,t)^{-(1-\alpha)}\,dv = \int_0^{Q_i} p_i(t)\,dv.$$

Note that this expression is not identical to the solution given in Equation (3.3) in the main text. The reason is that in the equation above the derivation has been taken with respect to $x_i(v,t)$ in general (i.e. the whole continuum of varieties) and not with respect to a specific intermediate good, like, for example, good j. For a specific good j, the correct derivative to take is $\frac{\partial \Pi_i(v,t)}{\partial x_i(j,t)}$. This derivative can be stated, by rewriting the integrals in the optimization problem (maybe slightly informally interpreting the integral as a sum of discrete varieties), as

$$\begin{aligned}
\frac{\partial \Pi_i(v,t)}{\partial x_i(j,t)} = Q_i(t)^{\alpha-1} & \left[\int_{\substack{v \neq j \\ v=0}}^{Q_i(t)} \frac{\partial}{\partial x_i(j,v)} A_i(v,t) x_i(v,t)^{\alpha} L_i(t)^{1-\alpha} dv \right. \\
& \left. + \frac{\partial}{\partial x_i(j,v)} A_i(j,t) x_i(j,t)^{\alpha} L_i(t)^{1-\alpha} \right] \\
& - \left[\int_{\substack{v \neq j \\ v=0}}^{Q_i(t)} p_i(v,t) x_i(v,t) dv + \frac{\partial}{\partial x_i(j,v)} p_i(j,t) x_i(j,t) \right] \\
& - \frac{\partial}{\partial x_i(j,v)} w_i(t) L_i(t).
\end{aligned}$$

Calculating the respective derivatives in this expression and setting the result equal to zero, leads to (note that the terms with the integrals no longer depend on $x_i(j,t)$ and thus their derivative with respect to this variable is equal to zero)

$$Q_i(t)^{\alpha-1} \alpha A_i(j,t) L_i(t)^{1-\alpha} x_i(j,t)^{-(1-\alpha)} = p_i(j,t).$$

As good j is just one specific good out of the continuum $v \in [0, Q_i(t)]$, the (inverse) demand from the producers of final goods for intermediate goods in Equation (3.3) in the main text follows.

C.2.2 Deriving the Production Function in Intensive Form

In equilibrium, capital supply, $K_i(t)$, equals capital demand, $\int_0^{Q_i(t)} K_i(v,t)dv$, and hence, substituting $K_i(v,t) = A_i(v,t)x_i(v,t)$ from the production function for intermediate goods (see Equation 3.5) into this equality, leads to

$$K_i(t) = \int_0^{Q_i(t)} A_i(v,t) x_i(v,t)\, dv \qquad \Longleftrightarrow \qquad K_i(t) = x_i(t) \int_0^{Q_i(t)} A_i(v,t)\, dv \qquad \text{(C.3)}$$

where the second equation has used the property that the equilibrium in the intermediate goods sector is symmetric (see Equation (3.9)). Defining the average productivity parameter in the intermediate goods sector as

$$A_i(t) \equiv \frac{1}{Q_i(t)} \int_0^{Q_i(t)} A_i(v,t)\, dv \qquad \Longleftrightarrow \qquad A_i(t) Q_i(t) = \int_0^{Q_i(t)} A_i(v,t)\, dv \qquad \text{(C.4)}$$

and substituting from the second equation in the expression above into Equation (C.3) results in $x_i(t) = \hat{k}_i(t)\frac{L_i(t)}{Q_i(t)}$, where $\hat{k}_i(t)$ is the capital stock per effective worker, i.e. $\hat{k}_i(t) \equiv \frac{K_i(t)}{A_i(t)L_i(t)}$. With the property of the symmetric equilibrium, the expression for $x_i(t)$ just derived, and the second expression in Equation (C.4), the production function in intensive form can be written as

$$\hat{y}_i(t) = \hat{k}_i(t)^\alpha$$

where $\hat{y}_i(t) \equiv \frac{Y_i(t)}{A_i(t)L_i(t)}$ is the production per effective worker, and which is identical to Equation (3.4).

In the main text in Section 3.2.1, it was stated that as the production function is multiplied by the factor $Q_i(t)^{\alpha-1}$, technological progress in this model is due to increases in productivity and not increases in the number of varieties as in the model by Romer (1990). This result will now be demonstrated mathematically. The production function in intensive form above can be expressed in aggregate terms as $Y_i(t) = A_i(t)L_i(t)^{1-\alpha}K_i(t)^\alpha A_i(t)^{-\alpha}$. From the two expressions after the equivalence arrows in Equations (C.3) and (C.4), it follows that $\frac{K_i(t)}{A_i(t)} = x_i(t)Q_i(t)$ so that

$$Y_i(t) = A_i(t)L_i(t)^{1-\alpha}\big(x_i(t)Q_i(t)\big)^\alpha \tag{C.5}$$

which has constant returns to scale in the two input factors labor and aggregate amount of intermediate inputs. As can be seen, technological progress in this specification is only due to increases in productivity. On the other hand, without the factor $Q_i(t)^{\alpha-1}$ in the original specification of the production function in Equation (3.1), the right-hand side in Equation (C.5) would need to be multiplied by $Q_i(t)^{1-\alpha}$ to obtain a corresponding result, and increases in the number of varieties would lead to increases in productivity in this case.

C.2.3 Convergence of the Number of Workers per Product to a Constant

The result that the number of workers per intermediate good, $L_i(t)/Q_i(t) = l_i(t)$ monotonically converges to the constant n_i/ξ_i in Equation (3.13) can be derived as follows: Taking the natural logarithm of $l_i(t)$ and deriving the result with respect to time yields

$$\frac{\dot{l}_i(t)}{l_i(t)} = \frac{\dot{L}_i(t)}{L_i(t)} - \frac{\dot{Q}_i(t)}{Q_i(t)}.$$

Inserting n_i for the population growth rate and substituting for $\dot{Q}_i(t)$ from Equation (3.12) leads to the differential equation

$$\frac{\dot{l}_i(t)}{l_i(t)} = n_i - \xi l_i(t) \qquad \Longleftrightarrow \qquad \dot{l}_i(t) - n_i l_i(t) = -\xi l_i(t)^2. \tag{C.6}$$

This equation has one steady state at $l_i^* = \frac{n_i}{\xi}$, which results from setting $\dot{l}_i(t) = 0$ and solving for l_i. Asymptotic convergence to the steady state follows as $\dot{l}_i(t) < 0$ for all $l_i(t) > l_i^*$ and $\dot{l}_i(t) > 0$ for all $l_i(t) < l_i^*$ (see Part 3 of Corollary 2.2 in Acemoglu (2009) for this approach to determine global asymptotic stability).[5]

C.2.4 Derivation of the Value of an Innovation

This section provides a derivation of the expression for the value of an innovation to a monopolist stated in Equation (3.16). In particular, it will be shown how this value depends on the Poisson arrival rate of new (quality) innovations.

A firm will reap monopoly profits from the time the innovation is brought to market (e.g. $t = 0$) until it is replaced at some time T, with $T \in (0, \infty)$, by a new monopolist producing a variety of a higher quality, and profits will fall to zero.[6] Therefore, the value for a firm at time 0 is given by[7]

$$V_d(0) = \int_0^T e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau$$

where $r(s)$ is the interest rate at time s, and the exponential expression is the discount factor applied to the monopolist's profits. That the replacement will happen is certain, but the point in time T in the future when it will happen can only be determined with

[5] Note that Equation (C.6) is a Bernoulli equation (Sydsæter et al., 2008, 208), which can be transformed into a standard linear differential equation by using the transformation $z(t) = \frac{1}{l_i(t)}$ and then be solved for the general solution

$$z(t) = \left(z(0) - \frac{\xi}{n_i}\right) e^{-n_i t} + \frac{\xi}{n_i}.$$

Reversing the transformation, the general solution for $l_i(t)$ is thus given by

$$l_i(t) = \frac{1}{\left(l_i(0)^{-1} - \frac{\xi}{n_i}\right) e^{-n_i t} + \frac{\xi}{n_i}},$$

which confirms that $l_i^* = \frac{n_i}{\xi}$ is indeed a steady-state value for the differential equation.

[6] This follows from the Arrow replacement effect and the fact that the previous monopolist will be driven out of the market via Bertrand competition, as the new innovator produces a higher quality good at identical costs (Aghion et al., 2014, 518).

[7] The subscript d denotes "deterministic" in this instance.

some probability. Hence, the expected value of an innovation is a random variable and can be expressed as follows[8]

$$\begin{aligned} V(0) = E[V_d(0)] &= \int_0^\infty f(T) \left[\int_0^T e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau \right] dT \\ &= \int_0^\infty \int_0^T f(T) e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau\, dT, \end{aligned} \tag{C.7}$$

where $f(T)$ is a general probability density function with $f(T) \geq 0\, \forall\, T$ and $\int_0^\infty f(T)\, dT = 1$. The equality in Equation (C.7) follows as $f(T)$ does not depend on τ and can thus be moved into the integral with respect to τ. However, this expression is still quite different from Equation (3.16).

The next step is to change the order of integration, which requires adjusting the limits of integration (this step is explained and illustrated in more detail at the end of this section). This procedure yields

$$V(0) = \int_0^\infty \left[\int_\tau^\infty f(T)\, dT \right] e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau. \tag{C.8}$$

Referring back to the discussion on Poisson processes in Appendix C.1, and making a specific distributional assumption on the function $f(T)$ (compare Equation (C.1)), the integral in brackets is just the probability that an innovation occurs after time τ, or, equivalently, that the firm can still earn monopoly profits at time τ. Calculating the complementary probability to the one stated (in general terms) in Equation (C.2), this probability is $e^{-\phi\tau}$ so that the value of an innovation is given by[9]

$$\begin{aligned} V(0) &= \int_0^\infty e^{-\phi\tau} \cdot e^{-\int_0^\tau r(s)ds} \pi(\tau)\, d\tau \\ &= \int_0^\infty e^{-\int_0^\tau (r(s)+\phi)ds} \pi(\tau)\, d\tau. \end{aligned}$$

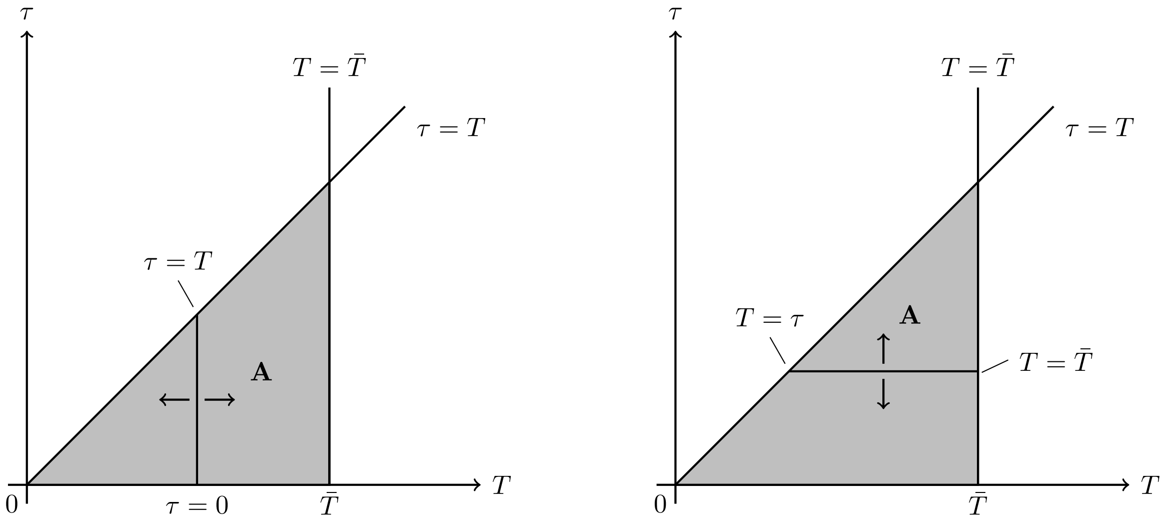
To be precise, this expression differs slightly from the more general one in the main text, as it is assumed here that ϕ is constant, which only holds in steady state (also, an identifier i for individual regions was dropped here).

Proving the validity of the change in the order of integration above, requires demonstrating that the expressions in Equation (C.7) and (C.8) are equivalent. This basically works by showing that the area of integration is identical in both cases. The following method can, for instance, be found in Sydsæter et al. (2008, 166pp) or Thomas Jr. (2005, 1074-75). In

[8] Note that this approach is basically the same as the one adopted by Yaari (1965, 142) in his model of uncertain lifetime.

[9] The second equality follows, as $e^{-\int_0^\tau \phi\, ds} = e^{-\left.\right|_0^\tau \phi s} = e^{-\phi\tau}$.

general, double integrals are evaluated by first working out the inner integral and then the outer one. In illustrating the method graphically in Figure C.4, the upper limit of integration in Equation (C.7) is changed from ∞ to an upper bound of $\bar{T}$ to simplify the graphical exposition. For the double integral in Equation (C.7), the relevant area of integration is depicted in Panel (a) of Figure C.4 and the one for Equation (C.8) in Panel (b).



(a) Illustration of Evaluating the Integral in Equation (C.7).

(b) Illustration of Evaluating the Integral in Equation (C.8).

Figure C.4: Graphical Illustration of Changing the Order of Integration and Preserving the Area of Integration.

In order to evaluate the integral in the first equation, the inner integral is evaluated along the line $\tau = 0$ to $\tau = T$, and then the outer integral is evaluated by integrating along all vertical lines from $T = 0$ to $T = \bar{T}$ (indicated by the horizontal arrows) to obtain the grey-shaded area of integration **A** (see Panel (a)). The same area is obtained by changing the order of integration, then first integrating along the horizontal line from $T = \tau$ to $T = \bar{T}$, and then the outer integral covers all horizontal lines from $\tau = 0$ to $\tau = \bar{T}$ (indicated by the vertical arrows) so that also in this case the grey-shaded area of integration **A** is obtained (see Panel (b)).

C.2.5 Derivation of the Research-Arbitrage Equation

The research-arbitrage equation results from deriving Equation (3.17) which is repeated here for convenience

$$v_i(t) = \int_t^{\infty} e^{-\int_t^{\tau}(r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)\,d\tau$$

with respect to t. Taking this derivative requires applying Leibniz's Formula (see, for example Sydsæter et al., 2008, 160)

$$F(t) = \int_{u(t)}^{v(t)} f(t,\tau)\, d\tau$$
$$\implies \quad F'(t) = \underbrace{f\big(t, v(t)\big)}_{1}\underbrace{v'(t)}_{2} - \underbrace{f\big(t, u(t)\big)}_{3}\underbrace{u'(t)}_{4} + \int_{u(t)}^{v(t)} \underbrace{\frac{\partial f(t,\tau)}{\partial t}}_{5}\, d\tau.$$

In the case at hand, the function $f(t,\tau)$ in the formula above is therefore given by $f(t,\tau) = e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)$, and the individually numbered terms above are given by the following expressions, respectively

1: $f\big(t, v(t)\big) = e^{-\int_t^\infty (r_i(s)+\phi_i(s))\,ds} \lim_{\tau\to\infty} \tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)$

2: $v'(t) = \frac{d}{dt}\text{“}\infty\text{”} = 0$

3: $f\big(t, u(t)\big) = e^{-\int_t^t (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t) = e^0\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t) = \tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t)$

4: $u'(t) = \frac{d}{dt}t = 1$

5: $\frac{\partial}{\partial t}e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau) = -[-r_i(t) + \phi_i(t)]e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)$

Using these intermediate results with the product 1 · 2 = 0 already inserted, it follows that

$$\begin{aligned}
\frac{\partial}{\partial t}v_i(t) = \dot{v}_i(t) &= -\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t) + \int_t^\infty [r_i(t) + \phi_i(t)]e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)\,d\tau \\
&= -\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t) + [r_i(t) + \phi_i(t)]\underbrace{\int_t^\infty e^{-\int_t^\tau (r_i(s)+\phi_i(s))\,ds}\tilde{\pi}_i\big(\hat{k}_i(\tau)\big)l_i(\tau)\,d\tau}_{v_i(t)} \\
&= -\tilde{\pi}_i\big(\hat{k}_i(t)\big)l_i(t) + [r_i(t) + \phi_i(t)]v_i(t)
\end{aligned}$$

and from the last expression, the research-arbitrage stated in the main text in Equation (3.17), is readily obtained.

C.2.6 Convergence of Relative Productivities

In the following, it will be demonstrated that the relative productivity parameters $a_i(v,t) = \frac{A_i(v,t)}{A_i(t)^{max}}$ converge to an invariant distribution. More specifically, it will be shown that the distribution of the fraction of sectors for which $A_i(v,t) \leq A_i(t)^{max}$ is time independent and given by $a_i^{\frac{1}{\sigma}}$. This result is based on the assumption that new and existing products have identical distributions for the productivity parameters at any time t. The proof follows along the lines of Aghion and Howitt (1998, 115).

For an arbitrary point in time t, denote the cumulative distribution of the absolute productivity parameters by $F(\cdot, t)$. At some point in time, $t_0 \geq 0$, one particular sector $v \in [0, Q_i(t)]$ with productivity parameter $A_i(v,t)$ necessarily was the leading-edge sector. Defining then the cumulative distribution function as $\Phi_i(t) = F\big(A_i(v,t), t\big)$, it needs to hold that

$$\Phi_i(t_0) = 1, \tag{C.9}$$

i.e. the probability that the particular sector that was picked out has the highest productivity across all sectors under consideration equals 1. At time t_0 "many" sectors are behind the one with the highest productivity. These sectors individually will innovate with the Poisson arrival rate for vertical innovations and, hence, in aggregate, since there are $\Phi_i(t)$ sectors, with the rate $\Phi_i(t)\lambda_i\kappa_i(t)^\phi$. This rate therefore equals the one with which the mass of sectors behind the leading one will decrease. In formal terms,

$$\dot{\Phi}_i(t) = -\Phi_i(t)\lambda_i\kappa_i(t)^\phi \qquad \forall\, t \geq t_0. \tag{C.10}$$

Equations (C.9) and (C.10) pose then an initial-value problem with solution

$$\Phi_i(t) = e^{-\int_{t_0}^{t} \lambda_i\kappa_i(s)^\phi \, ds} \qquad \forall\, t \geq 0. \tag{C.11}$$

Equation (3.20) implies the differential equation $\dot{A}_i(t)^{max} = \sigma\lambda_i\kappa_i(t)^\phi A_i(t)^{max}$, and at the start of this section it was assumed that $A_i(v,t) = A_i(t_0)^{max}$ (compare also the definition in Equation (3.15)). The solution to the differential equation for the leading-edge productivity parameter is therefore

$$A_i(t)^{max} = A_i(v,t)e^{\sigma \int_{t_0}^{t} \lambda_i\kappa_i(s) ds} \qquad \forall\, t \geq t_0. \tag{C.12}$$

From combining Equations (C.11) and (C.12), it thus follows that the distribution of the relative productivities in the long run converges to

$$\Phi_i(t) = \left(\frac{A_i(v,t)}{A_i(t)^{max}}\right)^{\frac{1}{\sigma}} = a_i(t)^{\frac{1}{\sigma}}.$$

As Aghion and Howitt (1992, 116) point out, in the long run almost all values for a_i in the interval $[0,1]$ will exist.[10]

[10] Additional remarks on the cross-section distribution, including a graphical analysis can be found in Howitt (2000, 834). See also Howitt (1999, 721).

C.2.7 Growth Rate of the Average and Leading-Edge Productivity Parameters

In the main text at the end of Section 3.2.3, an equation for the growth rate of the average productivity, $A_i(t)$, was given, which is repeated here for convenience

$$\dot{A}_i(t) = \lambda_i \kappa_i(t)^{\phi} \big(A_i(t)^{max} - A_i(t) \big).$$

It will now be demonstrated that the leading-edge and average productivity parameters will grow at identical rates. Defining the ratio between these two parameters as $\Gamma_i \equiv \frac{A_i(t)^{max}}{A_i(t)}$ and rewriting it in growth rates leads to

$$\frac{\dot{\Gamma}_i(t)}{\Gamma_i(t)} = \frac{\dot{A}_i(t)^{max}}{A_i(t)^{max}} - \frac{\dot{A}_i(t)}{A_i(t)}. \tag{C.13}$$

Substituting for the growth rate of the leading-edge parameter from Equation (3.20) and noting that the growth rate of the average productivity parameter is given by

$$\frac{\dot{A}_i(t)}{A_i(t)} = \lambda_i \kappa_i(t)^{\phi} \left[\frac{A_i(t)^{max}}{A_i(t)} - \frac{A_i(t)}{A_i(t)} \right] = \lambda_i \kappa_i(t)^{\phi} \big(\Gamma_i(t) - 1 \big),$$

it follows that the growth rate of the ratio of the productivity parameters is

$$\frac{\dot{\Gamma}_i(t)}{\Gamma_i(t)} = \sigma \lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi} \left[\Gamma_i(t) - 1 \right].$$

This expression can be rewritten as[11]

$$\dot{\Gamma}_i(t) = \big[(1+\sigma) \lambda_i \kappa_i(t)^{\phi} - \lambda_i \kappa_i(t)^{\phi} \Gamma_i(t) \big] \, \Gamma_i(t) \tag{C.14}$$

which has a trivial steady state at zero and a second one at $\Gamma_i^* = 1 + \sigma$. As long as $\lambda_i \kappa_i(t)^{\phi} > 0$, convergence to this value follows via applying the same approach as in Appendix C.2.3. From the definition of Γ_i, it holds that $A_i(t)^{max} = (1+\sigma) A_i(t)$, and both productivity parameters will therefore grow at the rate $g_i(t) = \sigma \lambda_i \kappa(t)^{\phi}$.

[11] As in the case for the differential equation for the number of workers, $l_i(t)$, Equation (C.14) is also a Bernoulli equation.

C.2.8 Derivation of the Global Technology Growth Rate

This section derives the productivity growth rate given in Equation (3.26) in the main text. Starting with inserting the expression for κ_i in Equation (3.25) into the one for g^w in Equation (3.24), yields

$$g^w = \frac{\sigma\lambda}{[(1+\sigma)\xi]^{\phi}} s_{A,i}^{\phi} y_i^{\phi} n_i^{\phi} A_i^{-\phi} \prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}}. \tag{C.15}$$

With the help of the properties of the product operator, the last factor can now be rewritten in the following way

$$\begin{aligned}
\prod_{j=1}^{N} \left(\frac{A_j}{A_i}\right)^{\gamma_i v_{ij}} &= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i v_{ij}} = \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} \prod_{j\neq i}^{N} A_i^{-\gamma_i v_{ij}} \\
&= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i \sum_{j\neq i}^{N} v_{ij}} \\
&= \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} A_i^{-\gamma_i (1-v_{ii})} \\
&= A_i^{-\gamma_i \frac{1}{\gamma_i}} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}} = A_i^{-1} \prod_{j\neq i}^{N} A_j^{\gamma_i v_{ij}}.
\end{aligned}$$

The step from the first to the second line uses the result $\sum_{j=1}^{N} v_{ij} = 1$ and the one from the second-to-last to the last line takes advantage of the definition $v_{ii} \equiv \frac{\gamma_i - 1}{\gamma_i} < 1$ (see Ertur and Koch (2011, 226-27) on these assumptions). Substituting the final result in the derivation above into Equation (C.15), leads to Equation (3.26) in Section 3.3.1.

C.2.9 Existence of $(\boldsymbol{I} - \gamma \boldsymbol{W})^{-1}$

In contrast to the case of a row-standardized interaction matrix, $\boldsymbol{I} - \gamma\boldsymbol{W}$ might be singular also for values in the interval $\gamma \in (-1, 1)$. The general condition for this matrix to be singular is $|\boldsymbol{I} - \gamma\boldsymbol{W}| = 0$, i.e. if $\frac{1}{\gamma}$ is an eigenvalue of the interaction matrix. Consider now, for instance, the matrix

$$\boldsymbol{W}_1 = \begin{pmatrix} 0 & 16 \\ 4 & 0 \end{pmatrix},$$

which is not row-standardized. Its characteristic equation is given by $\lambda^2 = 64$ so that the eigenvalues are $\lambda_1 = -8$ and $\lambda_2 = 8$. Then, for $\gamma = \frac{1}{8}$ the matrix $\boldsymbol{I} - \gamma\boldsymbol{W}_1$ will be singular. However, by restricting the parameter space for γ to $\gamma \in \left(-\frac{1}{\lambda_1}, -\frac{1}{\lambda_2}\right)$, the inverse above will be non-singular. An equivalent representation of the model under consideration can thus be obtained if the interaction matrix is normalized by this factor, i.e. $\boldsymbol{W}_1^* = \frac{\boldsymbol{W}_1}{\lambda_2}$ and by denoting $\gamma^* = \gamma\lambda_2$ with parameter space $\gamma^* \in (-1, 1)$. A similar procedure works in more general cases (Kelejian and Prucha, 2010, 56), when the eigenvalues cannot be as easily determined as in the matrix above. With the help of Gerschgorin's Circle Theorem (Gerschgorin, 1931), regions in the complex plane can be determined that contain the eigenvalues of the matrix.[12] With this information, it is possible to identify an interval for the parameter space, in which the inverse exists (see also Ertur and Koch, 2011, 231).

C.3 Additional Derivations – Econometric Theory

This appendix derives results in detail that are important in the econometric estimation of the model and for drawing inference. More specifically, a part of the score vector will be derived, before the steps in the derivation of the variance-covariance matrix, which is merely stated in Ertur and Koch (2011, 233), will be demonstrated.

C.3.1 Derivation of the Maximum Likelihood Estimator $\hat{\boldsymbol{\delta}}$

Before taking the derivative of the likelihood function with respect to $\boldsymbol{\delta}$, it will first be written in an expanded form. From Equation (3.43), it follows that

$$\begin{aligned}
\ln L(\boldsymbol{y}; \boldsymbol{\delta}, \gamma, \sigma^2) &= \underbrace{-\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|\boldsymbol{I} - \gamma\boldsymbol{W}|}_{\equiv C} \\
&\quad - \frac{1}{2\sigma^2}\Bigg[\boldsymbol{y}'(\boldsymbol{I} - \gamma\boldsymbol{W})'(\boldsymbol{I} - \gamma\boldsymbol{W})\boldsymbol{y} - \underbrace{\boldsymbol{y}'(\boldsymbol{I} - \gamma\boldsymbol{W})'\tilde{\boldsymbol{X}}\boldsymbol{\delta}}_{1\times 1} \\
&\qquad - \underbrace{\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'(\boldsymbol{I} - \gamma\boldsymbol{W})\boldsymbol{y}}_{1\times 1} + \boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\Bigg] \\
&= C - \frac{1}{2\sigma^2}\left[\boldsymbol{y}'(\boldsymbol{I} - \gamma\boldsymbol{W})'(\boldsymbol{I} - \gamma\boldsymbol{W})\boldsymbol{y} - 2\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'(\boldsymbol{I} - \gamma\boldsymbol{W})\boldsymbol{y} + \boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right]
\end{aligned}$$

where the last line has used the fact that the terms with the underbraces are identical scalars and that the matrix $(\boldsymbol{I} - \gamma\boldsymbol{W})$ is symmetric. Taking now the derivative[13] with respect to $\boldsymbol{\delta}$ leads to

$$\frac{\partial \ln L(\cdot)}{\partial \boldsymbol{\delta}} = -\frac{1}{2\sigma^2}\left[-2\tilde{\boldsymbol{X}}'(\boldsymbol{I} - \gamma\boldsymbol{W})\boldsymbol{y} + 2\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right].$$

Setting this expression equal to zero and solving for $\hat{\boldsymbol{\delta}}$, yields the expression in Equation (3.47).

[12] A more recent formal statement of this theorem can, for example, be found in Cheney and Kincaid (2008).

[13] For the rules on matrix derivation see, for example, Verbeek (2004, 394-95).

C.3.2 Derivation of the Variance-Covariance Matrix

The asymptotic variance-covariance matrix is given by the inverse of the information matrix $\boldsymbol{I}(\boldsymbol{\delta}, \gamma, \sigma^2)$, and this matrix is equal to the negative expected Hessian matrix, $\boldsymbol{H}$, for the log-likelihood function in Equation (3.43). In general terms, the information matrix thus reads

$$\boldsymbol{I}(\boldsymbol{\delta}, \gamma, \sigma^2) = -E[\boldsymbol{H}] = -E \begin{bmatrix} \frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \gamma} & \frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \sigma^2} \\ \frac{\partial^2 \ln L(\cdot)}{\partial \gamma \partial \boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial \gamma^2} & \frac{\partial^2 \ln L(\cdot)}{\partial \gamma \partial \sigma^2} \\ \frac{\partial^2 \ln L(\cdot)}{\partial \sigma^2 \partial \boldsymbol{\delta}'} & \frac{\partial^2 \ln L(\cdot)}{\partial \sigma^2 \partial \gamma} & \frac{\partial^2 \ln L(\cdot)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix}. \tag{C.16}$$

The individual entries for the first row in the Hessian matrix are calculated by taking the respective partial derivatives of Equation (C.3.1):

$$\frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} = -\frac{1}{\sigma^2} \tilde{\boldsymbol{X}}' \tilde{\boldsymbol{X}} \tag{C.17}$$

$$\frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \gamma} = -\frac{1}{\sigma^2} \tilde{\boldsymbol{X}}' \boldsymbol{W} \boldsymbol{y} \tag{C.18}$$

$$\frac{\partial^2 \ln L(\cdot)}{\partial \boldsymbol{\delta} \partial \sigma^2} = \frac{1}{\sigma^4} \left\{ -\tilde{\boldsymbol{X}}' \left[(I - \gamma \boldsymbol{W}) \, \boldsymbol{y} - \tilde{\boldsymbol{X}} \boldsymbol{\delta} \right] \right\} = -\frac{1}{\sigma^4} \tilde{\boldsymbol{X}}' \boldsymbol{\varepsilon} \tag{C.19}$$

where the last equality has used the expression for $\boldsymbol{\varepsilon}$ in Equation (3.44).

In order to calculate the entries in the second row of the Hessian, the first derivative of the log-likelihood function with respect to γ is needed. Note that the last term in the log-likelihood function can be equivalently written as $-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}$, and the derivative of this term with respect to γ is given by (compare, for example, Anselin, 1988b, 75)

$$\frac{\partial \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\partial \gamma} = \boldsymbol{\varepsilon}' \frac{\partial \boldsymbol{\varepsilon}}{\partial \gamma} + \frac{\partial \boldsymbol{\varepsilon}'}{\partial \gamma} \boldsymbol{\varepsilon} = 2 \boldsymbol{\varepsilon}' \frac{\partial \boldsymbol{\varepsilon}}{\partial \gamma}. \tag{C.20}$$

Deriving the log determinant with respect to γ makes use of Jacobi's formula (compare, for instance, Absil et al., 2008, 196). This states that the derivative of the determinant of a matrix $\boldsymbol{X}$ with respect to a can be expressed in the following way: $\frac{\partial |\boldsymbol{X}|}{\partial a} = \operatorname{tr} \left[\operatorname{adj} (\boldsymbol{X}) \frac{\partial \boldsymbol{X}}{\partial a} \right]$.[14] Alternatively, provided that $\boldsymbol{X}$ is invertible, the expression for the adjugate matrix, $\operatorname{adj}(\boldsymbol{X}) = |\boldsymbol{X}| (\boldsymbol{X})^{-1}$, can be inserted, implying that the derivative of the determinant is given by $|\boldsymbol{X}| \operatorname{tr} \left[(\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{X}}{\partial a} \right]$. In the following, a derivative of a log determinant will be taken so that taking into account the rules for differentiating logarithmic functions and the ones for determinants, Jacobi's formula reads $\frac{\partial \ln |\boldsymbol{X}|}{\partial a} = \operatorname{tr} \left[(\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{X}}{\partial a} \right]$ in this case.

[14] A proof of this result can be found, for instance, in Magnus and Neudecker (1999, 150).

Applying these rules to the case at hand and noting that $\frac{\partial \boldsymbol{\varepsilon}}{\partial \gamma} = -\boldsymbol{Wy}$, the partial derivative of the log-likelihood function with respect to γ is

$$\frac{\partial \ln L(\cdot)}{\partial \gamma} = -\mathrm{tr}\left(\boldsymbol{I} - \gamma \boldsymbol{W}\right)^{-1} \boldsymbol{W} + \frac{1}{\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{Wy}.$$

Before calculating the first entry in the second row of the Hessian matrix, note that the expressions $\boldsymbol{\varepsilon}'\boldsymbol{Wy}$ and $(\boldsymbol{Wy})'\boldsymbol{\varepsilon}$ denote an identical scalar. As $\boldsymbol{\varepsilon}'\boldsymbol{Wy}$ is a scalar, it is possible to introduce the trace operator (see Anselin, 1988b, 77) so that

$$\boldsymbol{\varepsilon}'\boldsymbol{Wy} = \mathrm{tr}\left[\boldsymbol{\varepsilon}'\left(\boldsymbol{Wy}\right)\right] = \mathrm{tr}\left\{\left[\boldsymbol{\varepsilon}'\left(\boldsymbol{Wy}\right)\right]'\right\} = \mathrm{tr}\left[\left(\boldsymbol{Wy}\right)'\boldsymbol{\varepsilon}\right] = \left(\boldsymbol{Wy}\right)'\boldsymbol{\varepsilon}'. \tag{C.21}$$

The second equality holds, since a matrix and its transpose have the same trace, and the third equality follows from the properties of transposed matrices. Substituting this expression into the first derivative above, inserting for $\boldsymbol{\varepsilon}$, and taking the partial derivative with respect to $\boldsymbol{\delta}'$ yields the result

$$\frac{\partial^2 \ln L(\cdot)}{\partial \gamma \partial \boldsymbol{\delta}'} = -\frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{Wy}\right)' \tag{C.22}$$

which is just the transpose of Equation (C.18). Anselin (1988b, 75) provides a helpful rule for taking the derivative of an inverse matrix, i.e. $\frac{\partial (\boldsymbol{X})^{-1}}{\partial a} = -\left(\boldsymbol{X}\right)^{-1}\frac{\partial \boldsymbol{X}}{\partial a}\left(\boldsymbol{X}\right)^{-1}$ and notes that the trace operator can be applied after differentiation as it is a linear operator. Hence,

$$\begin{aligned}
\frac{\partial^2 \ln L(\cdot)}{\partial \gamma^2} &= -\,\mathrm{tr}\left[-\left(\boldsymbol{I} - \gamma\boldsymbol{W}\right)^{-1}\left(-\boldsymbol{W}\right)\left(\boldsymbol{I} - \gamma\boldsymbol{W}\right)^{-1}\boldsymbol{W}\right] - \frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy} \\
&= -\,\mathrm{tr}\left[\boldsymbol{W}\left(\boldsymbol{I} - \gamma\boldsymbol{W}\right)^{-1}\boldsymbol{W}\left(\boldsymbol{I} - \gamma\boldsymbol{W}\right)^{-1}\right] - \frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy} \\
&= -\,\mathrm{tr}\left(\boldsymbol{W_A W_A}\right) - \frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy}
\end{aligned} \tag{C.23}$$

where the second equality has taken advantage of the property that the trace of a matrix is invariant to cyclical permutations (see, e.g. Meyer, 2000, 110). Additionally, in the expression in the last equality, the following definition from Ertur and Koch (2011, 233) is employed $\boldsymbol{W_A} \equiv \boldsymbol{W}\left(\boldsymbol{I} - \gamma\boldsymbol{W}\right)^{-1}$. Turning to the last entry in the second row of the Hessian, this is given by

$$\frac{\partial^2 \ln L(\cdot)}{\partial \gamma \partial \sigma^2} = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\boldsymbol{Wy}. \tag{C.24}$$

The derivatives in the third row of the Hessian matrix are the partial derivatives of the expression in Equation (3.45). The first entry in this row is

$$\begin{aligned}\frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\boldsymbol{\delta}'} &= \frac{1}{2\sigma^4}\left[-2\boldsymbol{y}'\left(\boldsymbol{I}-\gamma\boldsymbol{W}\right)'\tilde{\boldsymbol{X}}+2\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right]\\ &= -\frac{1}{\sigma^4}\left[\boldsymbol{y}'\left(\boldsymbol{I}-\gamma\boldsymbol{W}\right)'-\boldsymbol{\delta}'\tilde{\boldsymbol{X}}'\right]\tilde{\boldsymbol{X}} = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\tilde{\boldsymbol{X}} = -\frac{1}{\sigma^4}\left(\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon}\right)' \qquad \text{(C.25)}\end{aligned}$$

where again Equation (3.44) has been used. Applying one more time the rule in Equation (C.20) for differentiating the expression for the sum of squared errors, facilitates calculating the partial derivative with respect to γ so that

$$\frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\gamma} = \frac{1}{2\sigma^4}2\boldsymbol{\varepsilon}'\left(-\boldsymbol{W}\boldsymbol{y}\right) = -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y}. \qquad \text{(C.26)}$$

For the last entry in the third row, the partial derivative reads

$$\frac{\partial^2 \ln L(\cdot)}{\partial\sigma^2\partial\sigma^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}. \qquad \text{(C.27)}$$

Gathering the results in Equations (C.17) – (C.19) and Equations (C.22) – (C.27) yields the following Hessian matrix of dimension 7×7 (the first column has dimension 7×5 and the remaining two columns each have dimension 7×1)

$$\boldsymbol{H} = \begin{pmatrix} -\frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} & -\frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y} & -\frac{1}{\sigma^4}\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon} \\ -\frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y}\right)' & -\mathrm{tr}\left(\boldsymbol{W_A}\boldsymbol{W_A}\right) - \frac{1}{\sigma^2}\left(\boldsymbol{W}\boldsymbol{y}\right)'\boldsymbol{W}\boldsymbol{y} & -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y} \\ -\frac{1}{\sigma^4}\left(\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon}\right)' & -\frac{1}{\sigma^4}\boldsymbol{\varepsilon}'\boldsymbol{W}\boldsymbol{y} & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \end{pmatrix}.$$

The next step in deriving the information matrix is taking the (negative) expected value of the Hessian matrix above. Starting with the first column, its first entry contains no random variables and hence $-E\left[-\frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right] = \frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$. Moving on, the second entry equals

$$\begin{aligned}-E\left[-\frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W}\boldsymbol{y}\right)'\right] &= \frac{1}{\sigma^2}\left\{E\left[\tilde{\boldsymbol{X}}'\boldsymbol{W}(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\tilde{\boldsymbol{X}}\boldsymbol{\delta}+(\boldsymbol{I}-\gamma\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}\right]\right\}'\\ &= \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)' \qquad \text{(C.28)}\end{aligned}$$

where the last line follows as the expectation is a linear operator, the errors are assumed to be independent of all explanatory variables, and since $E[\boldsymbol{\varepsilon}] = 0$ due to the distributional assumption from Section 3.4.1. These latter two results can also be applied to calculate the last entry in the first column, implying that $-E\left[-\frac{1}{\sigma^4}\left(\tilde{\boldsymbol{X}}'\boldsymbol{\varepsilon}\right)'\right] = \boldsymbol{0}$.

In the second column of the information matrix, the first entry is simply the transpose of the 1×5 vector in Equation (C.28). However, the second entry on the diagonal requires more computations. The negative of the expected value of the first term in this entry is not a random variable and thus equals $-E\left[-\mathrm{tr}\left(\boldsymbol{W_A W_A}\right)\right] = \mathrm{tr}\left(\boldsymbol{W_A W_A}\right)$, while the following holds for the second term

$$\begin{aligned}
-E\left[-\frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy}\right] &= \frac{1}{\sigma^2}E\left\{\left[\boldsymbol{W_A\tilde{X}\delta} + \boldsymbol{W_A\varepsilon}\right]'\left[\boldsymbol{W_A\tilde{X}\delta} + \boldsymbol{W_A\varepsilon}\right]\right\}\\
&= \frac{1}{\sigma^2}\left\{E\left[\left(\boldsymbol{W_A\tilde{X}\delta}\right)'\boldsymbol{W_A\tilde{X}\delta}\right] + E\left[\left(\boldsymbol{W_A\tilde{X}\delta}\right)'\boldsymbol{W_A\varepsilon}\right]\right.\\
&\quad \left. + E\left[\left(\boldsymbol{W_A\varepsilon}\right)'\boldsymbol{W_A\tilde{X}\delta}\right] + E\left[\left(\boldsymbol{W_A\varepsilon}\right)'\boldsymbol{W_A\varepsilon}\right]\right\}.
\end{aligned}$$

Following the same arguments as above, the first term in the previous equation is completely deterministic and the cross products have an expected value of $\mathbf{0}$. For the last term, the rules from Equation (C.21) and the fact that cyclical permutations leave the trace of a matrix unchanged can be applied to demonstrate that[15]

$$\frac{1}{\sigma^2}E\left\{\mathrm{tr}\left[\boldsymbol{W_A'W_A\varepsilon\varepsilon'}\right]\right\} = \frac{1}{\sigma^2}\mathrm{tr}\left[\boldsymbol{W_A'W_A}\right]\underbrace{E\left[\boldsymbol{\varepsilon\varepsilon'}\right]}_{\sigma^2\boldsymbol{I}} = \mathrm{tr}\left[\boldsymbol{W_A'W_A}\right]. \tag{C.29}$$

Combing these partial results leads to the corresponding entry in the information matrix

$$\begin{aligned}
-E\left[-\mathrm{tr}\left(\boldsymbol{W_A W_A}\right) - \frac{1}{\sigma^2}\left(\boldsymbol{Wy}\right)'\boldsymbol{Wy}\right] &= \mathrm{tr}\left[\left(\boldsymbol{W_A} + \boldsymbol{W_A}'\right)\boldsymbol{W_A}\right]\\
&\quad + \frac{1}{\sigma^2}\left(\boldsymbol{W_A\tilde{X}\delta}\right)'\boldsymbol{W_A\tilde{X}\delta}.
\end{aligned}$$

Substituting for $\boldsymbol{y}$ in the in the last entry in the second column and transforming the resulting expression in a similar manner as in Equation (C.29) leads to

$$-E\left[-\frac{1}{\sigma^4}\boldsymbol{\varepsilon'Wy}\right] = \frac{1}{\sigma 4}E\left[\mathrm{tr}\left(\boldsymbol{W_A\varepsilon\varepsilon'}\right)\right] = \frac{1}{\sigma^2}\mathrm{tr}\boldsymbol{W_A}.$$

This entry is also identical to the second one in the third column in the information matrix, and the first entry in this column equals $\mathbf{0}$.[16] Noting that $E\left[\varepsilon'\varepsilon\right] = N\sigma^2$, the remaining entry on the diagonal reads

$$-E\left[\frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\boldsymbol{\varepsilon'\varepsilon}\right] = -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6}N\sigma^2 = \frac{N}{2\sigma^4}.$$

[15] See also Anselin (1988b, 77) for the second equality in this derivation.

[16] This follows as this value is simply the transpose of the third entry in the first column.

Collecting the results for the individual entries derived above, leads to the following information matrix

$$\boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} & \frac{1}{\sigma^2}\left(\tilde{\boldsymbol{X}}'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)' & \mathbf{0} \\ \frac{1}{\sigma^2}\tilde{\boldsymbol{X}}'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta} & \mathrm{tr}\left[\left(\boldsymbol{W_A}+\boldsymbol{W_A}'\right)\boldsymbol{W_A}\right] + \frac{1}{\sigma^2}\left(\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta}\right)'\boldsymbol{W_A}\tilde{\boldsymbol{X}}\boldsymbol{\delta} & \frac{1}{\sigma^2}\mathrm{tr}\boldsymbol{W_A} \\ \mathbf{0} & \frac{1}{\sigma^2}\mathrm{tr}\boldsymbol{W_A} & \frac{N}{2\sigma^4} \end{pmatrix}.$$

Finally, the asymptotic variance-covariance matrix, $\boldsymbol{V}(\boldsymbol{\delta},\gamma,\sigma^2)$, on which the hypotheses tests will be based, is then given by the inverse of the information matrix, i.e. $\boldsymbol{V}(\boldsymbol{\delta},\gamma,\sigma^2) = \boldsymbol{I}(\boldsymbol{\delta},\gamma,\sigma^2)^{-1}$.

C.4 Additional Estimation Results

This appendix provides estimation results from two additional analyses. The estimation results in the first section demonstrate that the omission of the state of Delaware is crucial for the results regarding the significance of the estimate of the investment rate in physical capital divided by the effective depreciation rate. Next, in Section C.4.2, the time horizon of the analysis is extended to cover the period 1990-2007, thereby ignoring the warning by the Bureau of Economic Analysis mentioned in Footnote 146 of Section 3.5.1 about appending the data series for the dependent variable.

C.4.1 Results – Benchmark Sample not Omitting Delaware

Table C.1: Estimation Results for Three Different Models for the Baseline Sample plus the State of Delaware and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.337	11.755	10.931	10.969	10.949	10.933	11.041	11.011
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.179	0.231	0.169	0.148	0.117	0.155	0.150	0.119
	(0.233)	(0.119)	(0.188)	(0.288)	(0.406)	(0.242)	(0.301)	(0.419)
$\ln s_{A,i}$	—	0.072	—	—	—	−0.013	0.001	0.001
		(0.033)				(0.745)	(0.979)	(0.984)
$\ln n_i$	—	0.280	—	—	—	0.018	0.013	0.011
		(0.406)				(0.547)	(0.680)	(0.717)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	−1.589	−0.232	0.061	−1.716	−0.258	0.034
			(0.101)	(0.608)	(0.920)	(0.084)	(0.572)	(0.956)
γ	—	—	0.149	0.129	0.135	0.159	0.126	0.133
			(0.000)	(0.001)	(0.001)	(0.003)	(0.021)	(0.014)
AIC	−3.661	−3.691	−3.869	−3.814	−3.807	−3.796	−3.736	−3.727
BIC	−3.583	−3.535	−3.498	−3.659	−3.651	−3.562	−3.502	−3.493
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Table C.2: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample plus the State of Delaware and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1997-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.155	0.150	0.119
	(0.248)	(0.306)	(0.422)
$\ln s_{A,i}$	−0.013	0.001	0.001
	(0.745)	(0.980)	(0.983)
$\ln n_i$	0.018	0.013	0.011
	(0.551)	(0.682)	(0.717)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.027	0.019	0.015
	(0.321)	(0.421)	(0.537)
$\boldsymbol{W} \ln s_{A,j}$	−0.005	−0.002	−0.002
	(0.628)	(0.797)	(0.803)
$\boldsymbol{W} \ln n_j$	0.003	0.001	0.001
	(0.616)	(0.782)	(0.811)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W} \frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.183	0.169	0.134
	(0.248)	(0.308)	(0.428)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	−0.017	−0.001	−0.001
	(0.718)	(0.987)	(0.984)
$\ln n_i + \boldsymbol{W} \ln n_j$	0.021	0.014	0.012
	(0.556)	(0.691)	(0.729)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

C.4.2 Data and Estimation Results – Baseline Sample (Extended Time Horizon: 1990-2007)

This appendix provides a brief description on how the variables have been constructed for the case when the sample period is extended to include the years 1990-1996 as well. After providing summary statistics in Table C.3, the results from the estimation of the nested models are shown in Table C.4. Estimates of the impacts are given in Table C.5.

Table C.3: Summary Statistics – Baseline Sample (Extended Time Horizon: 1990-2007).

Variable	Mean	Median	Standard deviation	Minimum	Maximum
y_i	85,012.07	80,896.36	13,921.60	66,616.49	123,281.63
$s_{K,i}$	0.079	0.075	0.014	0.061	0.127
n_i	0.016	0.014	0.009	0.002	0.043
δ_i	0.048	0.047	0.001	0.048	0.051
$n_i + g^w + \delta_i$	0.084	0.082	0.009	0.069	0.109
$s_{A,i}$	0.021	0.019	0.015	0.004	0.075
H_i	0.237	0.228	0.043	0.144	0.328
$\frac{s_{K,i}}{n_i+g^w+\delta_i}$	0.947	0.917	0.126	0.776	1.398
$\boldsymbol{W_1 s_K}$	0.219	0.214	0.044	0.130	0.332
$\boldsymbol{W_2 s_K}$	0.226	0.220	0.054	0.130	0.390
$\boldsymbol{W_3 s_K}$	0.224	0.219	0.050	0.128	0.345
$\boldsymbol{W_1 y}$	19,617.38	18,376.66	4,434.07	11,868.55	30,909.47
$\boldsymbol{W_2 y}$	19,825.09	18,874.88	5,152.72	11,411.44	40,053.49
$\boldsymbol{W_3 y}$	20,114.12	18,874.88	5,144.09	12,313.56	35,559.93

Note: The given values are the original values (i.e. not in logs) for the benchmark sample of 47 states and the period 1990-2007 with y_i the income per worker in 2007.

Even though the dependent variable is still real per worker income in 2007, values for the earlier years are needed to calculate the average real investment rate in physical capital, as Yamarik (2013) only provides values for gross real investment in physical capital. The data series on nominal gross state product for the years 1990-1996 from the Bureau of Economic Analysis' regional accounts data (BEA, 2015b) based on SIC, has been transformed as described in Section 3.5.1 into real 2000 dollars and then appended to the series for the years 1997-2007 based on NAICS.

Another complication arose in the construction of this data set, as the OECD only provides annual values for the R&D investment rate from 1997 onwards and additionally for the years 1991, 1993, and 1995 (OECD, 2015). Hence, the values for the years 1992 and 1994 have been interpolated by taking the average of the previous and subsequent years' value before calculating the average over the values from 1992 to 2007 to obtain the variable $s_{A,i}$.

As can be seen from Table C.3, no negative values for the employment growth rate occurred in this sample so that the values for all observations can be transformed into logs without any problems.

Finally, note that even though the dependent variable has not changed and the neighborhood relations and geographic distances between states are identical to the ones for the sample in the main text, this is not the case for the spatial lags, as these include a measure for the human capital stock.

Table C.4: Estimation Results for Three Different Models for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1990-2007.

Model	Solow (1956)	Howitt (2000)	Ertur and Koch (2007)			Ertur and Koch (2011)		
Interaction matrix			$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Constant	11.379	11.374	11.119	11.050	11.052	11.093	10.863	10.869
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.643	0.637	0.544	0.543	0.509	0.566	0.527	0.504
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
$\ln s_{A,i}$	—	0.044	—	—	—	0.020	−0.008	−0.007
		(0.134)				(0.505)	(0.809)	(0.826)
$\ln n_i$	—	−0.042	—	—	—	−0.042	−0.039	−0.040
		(0.180)				(0.153)	(0.182)	(0.161)
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	—	—	0.097	−0.489	−0.019	−0.041	−0.295	0.063
			(0.937)	(0.443)	(0.982)	(0.973)	(0.648)	(0.940)
γ	—	—	0.098	0.118	0.118	0.070	0.114	0.112
			(0.014)	(0.001)	(0.001)	(0.116)	(0.017)	(0.017)
AIC	−4.003	−4.056	−4.035	−4.141	−4.130	−4.019	−4.086	−4.082
BIC	−3.925	−3.898	−3.877	−3.983	−3.972	−3.783	−3.850	−3.846
Number of observations	47	47	47	47	47	47	47	47

Note: p-values are given in parentheses.

Table C.5: Estimation Results for the Direct, Indirect and Total Impacts in the Multi-Region Schumpeterian Model for the Baseline Sample of 47 States and Interaction Matrices $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, and $\boldsymbol{W_3}$ for the Period 1990-2007.

Interaction matrix	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$	$\boldsymbol{W_3}$
Direct impacts:			
$\ln s_{K,i} - \ln(n_i + 0.02 + \delta_i)$	0.565	0.527	0.504
	(0.000)	(0.001)	(0.002)
$\ln s_{A,i}$	0.020	−0.008	−0.007
	(0.508)	(0.812)	(0.824)
$\ln n_i$	−0.042	−0.039	−0.040
	(0.160)	(0.189)	(0.168)
Indirect impacts:			
$\boldsymbol{W}[\ln s_{K,j} - \ln(n_j + 0.02 + \delta_j)]$	0.042	0.066	0.062
	(0.155)	(0.049)	(0.053)
$\boldsymbol{W} \ln s_{A,j}$	0.001	−0.002	−0.002
	(0.765)	(0.667)	(0.675)
$\boldsymbol{W} \ln n_j$	−0.003	−0.005	0.006
	(0.358)	(0.294)	(0.957)
Total impacts:			
$\frac{\ln s_{K,i}}{\ln(n_i+0.02+\delta_i)} + \boldsymbol{W}\frac{\ln s_{K,j}}{\ln(n_j+0.02+\delta_j)}$	0.607	0.593	0.566
	(0.000)	(0.001)	(0.002)
$\ln s_{A,i} + \boldsymbol{W} \ln s_{A,j}$	0.021	−0.010	−0.001
	(0.518)	(0.788)	(0.801)
$\ln n_i + \boldsymbol{W} \ln n_j$	−0.045	−0.044	−0.045
	(0.161)	(0.189)	(0.169)

Note: p-values are given in parentheses. These were constructed using a set of 500,000 random draws from the estimation.

C.5 List of States Included in the Empirical Analyses

This appendix lists the states that are included in the different empirical analyses and also provides a correspondence with the state abbreviations used in Figure 3.1.

Table C.6: Alphabetical List of the 48 US States plus the District of Columbia.

State	Code	State	Code
Alabama	AL	Nebraska	NE
Arizona	AZ	Nevada	NV
Arkansas	AR	New Hampshire	NH
California	CA	New Jersey	NJ
Colorado	CO	New Mexico	NM
Connecticut	CT	New York	NY
Delaware	DE	North Carolina	NC
District of Columbia	DC	North Dakota	ND
Florida	FL	Ohio	OH
Georgia	GA	Oklahoma	OK
Idaho	ID	Oregon	OR
Illinois	IL	Pennsylvania	PA
Indiana	IN	Rhode Island	RI
Iowa	IA	South Carolina	SC
Kansas	KS	South Dakota	SD
Kentucky	KY	Tennessee	TN
Louisiana	LA	Texas	TX
Maine	ME	Utah	UT
Maryland	MD	Vermont	VT
Massachusetts	MA	Virginia	VA
Michigan	MI	Washington	WA
Minnesota	MN	West Virginia	WV
Mississippi	MS	Wisconsin	WI
Missouri	MO	Wyoming	WY
Montana	MT		

Bibliography

ABEL, G. J. AND N. SANDER (2014). "Quantifying Global International Migration Flows," *Science*, 343(6178), 1520–1522.

ABSIL, P.-A., R. MAHONY, AND R. SEPULCHRE (2008). *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton.

ACEMOGLU, D. (2009). *Introduction to Modern Economic Growth*, Princeton University Press, Princeton.

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

AGHION, P., U. AKCIGIT, AND P. HOWITT (2014). "What Do We Learn From Schumpeterian Growth Theory?" in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 2B, Chapter 1, Elsevier B.V., Oxford, 515–563.

AGHION, P. AND P. HOWITT (1992). "A Model of Growth through Creative Destruction," *Econometrica*, 60, 323–351.

——— (1998). *Endogenous Growth Theory*, MIT Press, Cambridge.

AHRENS, T. J. (editor) (1995). *Global Earth Physics: A Handbook of Physical Constants*, Volume 1 of *AGU Reference Shelf Series*, American Geophysical Union, Washington.

AKAIKE, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle," in B. N. Petrov and F. Cszaki (editors), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267–281.
Reprinted in: S. Kotz and N. L. Johnson (editors) (1992), *Breakthroughs in Statistics Volume 1. Foundations and Basic Theory*, Springer Series in Statistics, Perspectives in Statistics, Springer, Berlin, 610-624.

ALESINA, A. AND E. SPOLAORE (2003). *The Size of Nations*, MIT Press, Cambridge.

ANSELIN, L. (1988a). "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity," *Geographical Analysis*, 20(1), 1–17.

——— (1988b). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

——— (2003). "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics," *International Regional Science Review*, 26(2), 153–166.

ANSELIN, L., A. K. BERA, R. FLORAX, AND M. J. YOON (1996). "Simple diagnostic tests for spatial dependence," *Regional Science and Urban Economics*, 26, 77–104.

Arrow, K. J. (1962). "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29, 155–173.

Ashraf, Q. and O. Galor (2013). "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103(1), 1–46.

Avent, R. (2011). *The Gated City*.

Backus, D. K., P. J. Kehoe, and T. J. Kehoe (1992). "In Search of Scale Effects in Trade and Growth," *Journal of Economic Theory*, 58, 377–409.

Baier, S. L. and J. H. Bergstrand (2001). "The growth of world trade: tariffs, transport costs, and income similarity," *Journal of International Economics*, 53, 1–27.

Bairoch, P. (1988). *Cities and Economic Development - From the Dawn of History to the Present*, Mansell Publishing Limited, Chicago.

Baldwin, R. E. and R. Forslid (2000). "The Core-Periphery Model and Endogenous Growth: Stabilizing and Destabilizing Integration," *Economica*, 67, 307–324.

Baldwin, R. E., P. Martin, and I. P. Ottaviano, Gianmarco (2001). "Global Income Divergence, Trade, and Industrialization: The Geography of Growth Take-Offs," *Journal of Economic Growth*, 6, 5–37.

Barro, R. J. (1990). "Government Spending in a Simple Model of Endogenous Growth," *Journal of Political Economy*, 98, 103–125.

Barro, R. J. and X. Sala-i-Martin (1992). "Convergence," *Journal of Political Economy*, 100(2), 223–251.

——— (2004). *Economic Growth*, 2nd Edition, MIT Press, Cambridge.

Basile, R. (2014). "Regional productivity growth in Europe: a Schumpeterian perspective," Geocomplexity Discussion Paper No. 1.

Behrens, K. and J.-F. Thisse (2007). "Regional economics: A new economic geography perspective," *Regional Science and Urban Economics*, 37, 457–465.

Bode, E., P. Nunnenkamp, and A. Waldkirch (2012). "Spatial effects of foreign direct investment in US states," *Canadian Journal of Economics*, 45(1), 16–40.

Bolt, J. and J. L. van Zanden (2013). "The First Update of the Maddison Project; Re-Estimating Growth before 1820," Maddison Project Working Paper 4.

Bosker, M. and H. Garretsen (2009). "Economic Development and the Geography of Institutions," *Journal of Economic Geography*, 9, 295–328.

Brakman, S., H. Garretsen, and C. van Marrewijk (2009). *The New Introduction to Geographical Economics*, 2nd Edition, Cambridge University Press, Cambridge.

Brakman, S. and C. van Marrewijk (2008). "It's a big world after all: on the economic impact of location and distance," *Cambridge Journal of Regions, Economy and Society*, 1, 411–437.

Breton, T. R. (2012). "Penn World Table 7.0: Are the data flawed?" *Economics Letters*, 117, 208–210.

Brülhart, M. (2009). "An Account of Global Intra-industry Trade, 1962-2006," *The World Economy*, 32(3), 401–459.

Bureau of Economic Analysis (2015a). *National Economic Accounts.*
URL: *http://www.bea.gov/national/index.htm, accessed: 29 May, 2015.*

——— (2015b). *Regional Economic Accounts.*
URL: *http://www.bea.gov/regional/index.htm, accessed: 22 June, 2015.*

Bureau of Labor Statistics (2015). *Current Employment Statistics.*
URL: *http://www.bls.gov/sae/, accessed: 22 June, 2015.*

Caballero, R. J. and A. B. Jaffe (1993). "How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in O. Blanchard and S. Fischer (editors), *NBER Macroeconomis Annual*, Volume 8, Chapter 1, MIT Press, Massachusetts, 15–86.

Campbell, D. L. and J. H. Pyun (2015). "The Diffusion of Devlopment: Along Genetic or Geographic Lines?" *Journal of International Development.*

Caselli, F. (2005). "Accounting for Cross-Country Income Differences," in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 1A, Chapter 9, Elsevier B.V., Oxford, 679–741.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza (1994). *The History and Geography of Human Genes*, Princeton University Press, Princeton.

Central Intelligence Agency (2013). *The World Factbook, 2013-14*, Central Intelligence Agency, Washington, DC.
URL: *https://www.cia.gov/library/publications/the-world-factbook/index.html, accessed 22 Januar, 2015.*

Cheney, W. and D. Kincaid (2008). *Numerical Mathematics and Computing*, 6th Edition, Thomson Higher Education, Belmont.

Cliff, A. and K. Ord (1972). "Testing for Spatial Autocorrelation Among Regression Residuals," *Geographical Analysis*, 4(3), 267–284.

Cooke, P., M. G. Uranga, and G. Etxebarria (1997). "Regional Innovation Systems: Institutional and Organisational Dimensions," *Research Policy*, 26, 475–491.

Crespo Cuaresma, J., M. Feldkircher, B. Grün, P. Hofmarcher, and S. Humer (2015). "Measuring Income with Penn World Table Data: Reaching Consensus?" *mimeo.*
URL: *http://www.alde.es/encuentros/trabajos/c/pdf/209.pdf, accessed: 11 August, 2015.*

Dalgaard, C.-J. and C. T. Kreiner (2003). “Endogenous Growth: A Knife Edge or the Razor’s Edge?” *Scandinavian Journal of Economics*, 105(1), 73–85.

Dalgaard, C.-J. and H. Strulik (2013). “The history augmented Solow model,” *European Economic Review*, 63, 134–149.

Davidson, R. and J. G. MacKinnon (2004). *Econometric Theory and Methods*, Oxford University Press, Oxford.

Davis, D. R. (1995). “Intra-industry trade: A Heckscher-Ohlin-Ricardo approach,” *Journal of International Economics*, 39, 201–226.

Deeken, T. (2015a). “Knowledge Spillovers: On the Impact of Genetic Distance and Data Revisions,” Working Paper Series in Economics 74, KIT.
URL: *http://econpapers.wiwi.kit.edu/, accessed: 9 November, 2015.*

——— (2015b). “Schumpeterian Growth with Technological Interdependence: An Application to US States,” Working Paper Series in Economics 75, KIT.
URL: *http://econpapers.wiwi.kit.edu/, accessed: 9 November, 2015.*

Deeken, T. and I. Ott (2014a). “Integration as a Spatial Institution: Implications for Agglomeration and Growth,” in R. Funck and W. Rothengatter (editors), *Man, Environment, Space and Time - Economic Interactions in Four Dimensions*, Volume 34 of *Karlsruhe Papers in Economic Policy Research*, Nomos, Baden-Baden, 413–449.

——— (2014b). “Integration as a Spatial Institution: Implications for Agglomeration and Growth,” Working Paper Series in Economics 59, KIT.
URL: *http://econpapers.wiwi.kit.edu/, accessed: 18 August, 2015.*

Dell, M., B. F. Jones, and B. A. Olken (2012). “Temperature Shocks and Economic Growth: Evidence from the Last Half Century,” *American Economic Journal: Macroeconomics*, 4(3), 66–95.

Derbyshire, J., B. Gardiner, and S. Waights (2011). “Estimating the capital stock for the NUTS 2 regions of the EU-27,” *European Union Regional Policy Working Papers*, (1).

Desmet, K., M. Le Breton, I. Ortuño-Ortín, and S. Weber (2011). “The stability and breakup of nations: a quantitative analysis,” *Journal of Economic Growth*, 16, 183–213.

Dixit, A. K. and J. Stiglitz (1977). “Monopolistic Competition and Optimum Product Diversity,” *American Economic Review*, 67, 297–308.

Dowdy, S., S. Wearden, and D. Chilko (2004). *Statistics for Research*, 3rd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken.

Duranton, G. and D. Puga (2000). “Diversity and Specialisation in Cities: Why, Where and When Does it Matter?” *Urban Studies*, 37(3), 533–555.

Durlauf, S. N. and D. T. Quah (1999). "The New Empirics of Economic Growth," in J. Taylor and M. Woodford (editors), *Handbook of Macroeconomics*, Volume 1A, Chapter 4, Elsevier Science B.V., Oxford, 235–308.

Easterly, W., L. Freschi, and S. Pennings (2015). "A Long History of a Short Block: Four Centuries of Development Surprises on a Single Stretch of a New York City Street," .
URL: *http://www.greenestreet.nyc/paper, accessed: 16 August, 2015.*

Eaton, J. and S. Kortum (1996). "Trade in ideas Patenting and productivity in the OECD," *Journal of International Economics*, 40, 251–278.

——— (1999). "International Technology Diffusion. Theory and Measurement," *International Economic Review*, 40(3), 537–570.

Eicher, T. and S. Turnovsky (2000). "Scale, Congestion and Growth," *Economica*, 67, 325–346.

Eicher, T. S. and S. J. Turnovsky (1999). "Non-Scale Models of Economic Growth," *The Economic Journal*, 109, 394–415.

Elhorst, P. J. (2010). "Applied Spatial Econometrics: Raising the Bar," *Spatial Economic Analysis*, 5(1), 9–28.

Ertur, C. and W. Koch (2006). "Convergence, Human Capital and International Spillovers," Working Paper, LEG 2006-03, Université de Bourgogne, France.

——— (2007). "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence," *Journal of Applied Econometrics*, 22(6), 1033–1062.

——— (2011). "A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions," *Journal of Economic Growth*, 16, 215–255.

Feenstra, R. C., R. Inklaar, and T. Marcel (2013a). "PWT 8.0 - a user guide," *mimeo*.
URL: *http://www.rug.nl/research/ggdc/data/pwt/pwt-research-papers, accessed: 18 August, 2015.*

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2013b). "The Next Generation of the Penn World Table," .
URL: *http://www.rug.nl/research/ggdc/data/pwt/pwt-8.1, accessed: 18 August, 2015.*

——— (2015). "The Next Generation of the Penn World Table," *American Economic Review*, (forthcoming).

Fischer, M. M. (2011). "A spatial Mankiw-Romer-Weil model: theory and evidence," *The Annals of Regional Science*, 47, 419–436.

Fischer, M. M., M. Bartkowska, A. Riedl, S. Sardadvar, and A. Kunnert (2010). "The Impact of Human Capital on Regional Labor Productivity in Europe," in M. M. Fischer and A. Getis (editors), *Handbook of Applied Spatial Analysis – Software Tools, Methods and Applications*, Chapter E.1, Springer, Berlin, 585–597.

Fischer, M. M. and J. Wang (2011). *Spatial Data Analysis: Models, Methods and Techniques*, SpringerBriefs in Regional Science, Springer, Heidelberg.

Florax, R. J., H. Folmer, and S. J. Rey (2003). "Specification searches in spatial econometrics - the relevance of Hendry's methodology," *Regional Science and Urban Economics*, 33, 557–579.

Friedman, T. L. (2005). *The World is Flat - A Brief History of the Twenty-First Century*, Farrar, Straus and Giroux, New York.

Fujita, M. and J.-F. Thisse (2002). *Economics of Agglomeration: Cities, Regions, and International Trade*, Cambridge University Press, Cambridge.

Gabaix, X. and Y. M. Ioannides (2004). "The Evolution of City Size Distributions," in J. V. Henderson and J.-F. Thisse (editors), *Handbook of Regional and Urban Economics*, Volume 4, Chapter 53, Elsevier B.V., Oxford, 2341–2378.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999). "Geography and Economic Development," *International Regional Science Review*, 22(2), 179–232.

Galor, O. (2005). "From Stagnation to Growth: Unified Growth Theory," in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 1A, Chapter 4, Elsevier B.V., Oxford, 171–293.

——— (2011). *Unified Growth Theory*, Princeton University Press, Princeton.

Garofalo, G. A. and S. Yamarik (2002). "Regional Convergence: Evidence from a New State-by-State Capital Stock Series," *The Review of Economics and Statistics*, 84(2), 316–323.

Gerschenkron, A. (1962). *Economic Backwardness in Historical Perspective: A Book of Essays*, The Belknap Press of Harvard University Press, Cambridge.

Gerschgorin, S. (1931). "Über die Abgrenzung der Eigenwerte einer Matrix," *Bulletin de L'Académie des Sciences de L'URSS*, 749–754.

Gilfillan, S. C. (1935a). *Inventing the Ship*, Follett, Chicago.

——— (1935b). *The Sociology of Invention*, Follett, Chicago.

Giuliano, P., A. Spilimbergo, and G. Tonon (2014). "Genetic distance, transportation costs, and trade," *Journal of Economic Geography*, 14, 179–198.

Gollin, D. (2002). "Getting Income Shares Right," *Journal of Political Economy*, 110(2), 458–474.

Greene, W. H. (2003). *Econometric Analysis*, 5th Edition, Pearson Education Ltd., Upper Saddle River.

Halberstam, D. and R. M. Hills, Jr. (2001). "State Autonomy in Germany and the United States," *Annals of the American Academy of Political and Social Science*, 574, 173–184.

HANUSHEK, E. A., J. RUHOSE, AND L. WOESSMANN (2015). "Human Capital Quality and Aggregate Income Differences: Development Accounting for U.S. States," Working Paper 21295, National Bureau of Economic Research.

HENDERSON, V. J., A. STOREYGARD, AND D. E. WEIL (2012). "Measuring Economic Growth from Outer Space," *American Economic Review*, 102(2), 994–1028.

HERNDON, T., M. ASH, AND R. POLLIN (2014). "Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff," *Cambridge Journal of Economics*, 38, 257–279.

HESTON, A., R. SUMMERS, AND B. ATEN (2002). "Penn World Table Version 6.1," Center for International Comparison of Production, Income and Prices at the University of Pennsylvania.

——— (2006). "Penn World Table Version 6.2," Center for International Comparison of Production, Income and Prices at the University of Pennsylvania.

——— (2012). "Penn World Table Version 7.1," Center for International Comparison of Production, Income and Prices at the University of Pennsylvania.

HEWINGS, G. J. D. AND J. OOSTERHAVEN (2014). "Interregional Trade Models," in M. M. Fischer and P. Nijkamp (editors), *Handbook of Regional Science*, Chapter 46, Springer-Verlag, Berlin, 903–925.

HOLSINGER, K. E. AND B. S. WEIR (2009). "Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}," *Nature Reviews Genetics*, 10, 639–650.

HOLTZ-EAKIN, D. (1993). "Solow and the States: Capital Accumulation, Productivity, and Economic Growth," *National Tax Journal*, 46(4), 425–439.

HOWITT, P. (1999). "Steady Endogenous Growth with Population and R & D Inputs Growing," *Journal of Political Economy*, 107(4), 715–730.

——— (2000). "Endogenous Growth and Cross-country differences," *American Economic Review*, 90, 829–846.

HUMMELS, D. (2007). "Transportation Costs and International Trade in the Second Era of Globalization," *Journal of Economic Perspectives*, 21(3), 131–154.

INKLAAR, R. AND M. P. TIMMER (2013). "Capital, labor and TFP in PWT8.0," *mimeo*. **URL:** *http://www.rug.nl/research/ggdc/data/pwt/pwt-research-papers*, *accessed: 18 August, 2015.*

ISLAM, N. (1995). "Growth Empirics: A Panel Data Approach," *The Quarterly Journal of Economics*, 110(4), 1127–1170.

ISPHORDING, I. E. AND S. OTTEN (2013). "The Costs of Babylon - Linguistic Distance in Applied Economics," *Review of International Economics*, 21(2), 354–369.

JACOBS, J. (1969). *The Economy of Cities*, Random House, New York.

JOHNSON, S., W. LARSON, C. PAPAGEORGIOU, AND A. SUBRAMANIAN (2013). "Is newer better? Penn World Table Revisions and their impact on growth estimates," *Journal of Monetary Economics*, 60, 255–274.

JONES, C. I. (1995a). "R & D-Based Models of Economic Growth," *The Journal of Political Economy*, 103(4), 759–784.

——— (1995b). "Time Series Tests of Endogenous Growth Models," *The Quarterly Journal of Economics*, 110(2), 495–525.

JONES, C. I. AND P. M. ROMER (2010). "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, 2(1), 224–245.

KALDOR, N. (1957). "A Model of Eocnomic Growth," *Economic Journal*, 67(268), 591–624.

——— (1961). "Capital Accumulation and Economic Growth," in F. A. Lutz and D. C. Hague (editors), *The Theory of Capital*, St. Martin's Press, New York, 177–222.

KEILBACH, M. (2000). *Spatial Knowledge Spillovers and the Dynamics of Agglomeration and Regional Growth*, Physica-Verlag, Heidelberg.

KELEJIAN, H. H. AND I. R. PRUCHA (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, 157, 53–67.

KELLER, K. R. I. AND P. POUTVAARA (2005). "Growth in OECD countries and elsewhere: how much do education and R&D explain?" *Economics Bulletin*, 15(16), 1–11.

KELLER, W. (2002). "Geographic localization of international technology diffusion," *American Economic Review*, 92(1), 120–142.

KIMURA, M. (1968). "Evolutionary Rate at the Molecular Level," *Nature*, 217, 624–626.

KLENOW, P. J. AND A. RODRÍGUEZ-CLARE (2005). "Externalities and Growth," in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 1A, Chapter 11, Elsevier B.V., Oxford, 817–861.

KNOWLES, S. AND P. D. OWEN (1995). "Health capital and cross-country variation in income per capita in the Mankiw-Romer-Weil model," *Economics Letters*, 48, 99–106.

KREMER, M. (1993). "Population Growth and Technological Change: One Million B.C. to 1990," *The Quarterly Journal of Economics*, 108(3), 681–716.

KRUGMAN, P. (1991). "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

——— (1998). "What's New About the New Economic Geogrpahy?" *Oxford Review of Economic Policy*, 14(2), 7–17.

Le Gallo, J. (2014). "Cross-Section Spatial Regression Models," in M. M. Fischer and P. Nijkamp (editors), *Handbook of Regional Science*, Chapter 76, Springer-Verlag, Berlin, 1511–1533.

Lee, L. (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72(6), 1899–1925.

LeSage, J. and R. K. Pace (2009). *Introduction to Spatial Econometrics*, Chapman and Hall/CRC, Boca Raton.

LeSage, J. P. and K. R. Pace (2014). "The Biggest Myth in Spatial Econometrics," *Econometrics*, 2, 217–249.

Lindner, I. and H. Strulik (2014). "The great divergence: a network approach," Discussion Papers, Center for European Governance and Economic Development Research, No. 193.

López-Bazo, E., E. Vayó, and M. Artís (2004). "Regional Externalities and Growth: Evidence from European Regions," *Journal of Regional Science*, 44(1), 43–73.

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107, 407–437.

Marshall, A. (1890). *Principles of Economics*, Macmillan, London.

Masel, J. (2011). "Genetic drift," *Current Biology*, 21(20), R837–R838.

Meissner, C. M. (2014). "Growth from Globalization? A View from the Very Long Run," in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 2B, Chapter 8, Elsevier B.V., Oxford, 1033–1069.

Melitz, J. and F. Toubal (2014). "Native language, spoken language, translation and trade," *Journal of International Economics*, 93, 351–363.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia.

Michalopoulos, S. and E. Papaioannou (2013). "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81(1), 113–152.

National Science Board (2014). *Science and Engineering Indicators 2014*, National Science Foundation, Arlington.

Nei, M. (1972). "Genetic Distance between Populations," *The American Naturalist*, 106(949), 283–292.

——— (1973). "Analysis of Gene Diversity in Subdivided Populations," *Proceedings of the National Academy of Sciences*, 70(12), 3321–3323.

Nelson, R. R. and E. S. Phelps (1966). "Investment in Humans, Technological Diffusion and Economic Growth," *American Economic Review*, 61(1/2), 69–75.

Nonneman, W. and P. Vanhoudt (1996). "A Further Augmentation of the Solow Model and the Empirics of Economic Growth for OECD Countries," *The Quarterly Journal of Economics*, 111(3), 943–953.

North, D. (1990). *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge.

Oeppen, J. and J. W. Vaupel (2002). "Broken Limits to Life Expectancy," *Science*, 296, 1029–1031.

Ord, K. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70(349), 120–126.

Organisation for Economic Co-operation and Development (2010). *OECD Science, Technology and Industry Outlook*, OECD Publishing.

——— (2014). *OECD Science, Technology and Industry Outlook*, OECD Publishing.

——— (2015). *OECD Regional Database*.
URL: *https://www.oecd.org/gov/regional-policy/regionalstatisticsandindicators.htm, accessed: 12 June, 2015.*

Ottaviano, G. and J.-F. Thisse (2004). "Agglomeration and Economic Geography," in J. Henderson and J.-F. Thisse (editors), *Handbook of Regional and Urban Economics*, Volume 4, Chapter 58, Elsevier B.V., Oxford, 2563–2608.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011). "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *The World Bank Economic Review*, 25(2), 12–56.

Pace, R. K. and R. Barry (1997). "Quick Computation of Spatial Autoregressive Estimators," *Geographical Analysis*, 29(3), 232–247.

Peri, G. (2012). "The Effect of Immigration on Productivity: Evidence From U.S. States," *The Review of Economis and Statistics*, 94(1), 348–358.

Peterson, J. C. and R. D. Smith (2012). *Introductory Technical Mathematics*, 6th Edition, Cengage Learning, Clifton Park.

Ponomareva, N. and H. Katayama (2010). "Does the version of the Penn World Tables matter? An analysis of the relationship between growth and volatility," *Canadian Journal of Economics*, 43, 152–179.

Porter, M. E. (1990). *The Competitive Advantage of Nations*, Macmillan, New York.

Pritchett, L. and L. H. Summers (1996). "Wealthier is Healthier," *The Journal of Human Resources*, 31(4), 841–868.

Puga, D. (2010). "The Magnitude and Causes of Agglomeration Economies," *Journal of Regional Science*, 50(10), 203–219.

Ram, R. (2007). "IQ and economic growth: Further augmentation of Mankiw-Romer-Weil model," *Economis Letters*, 94, 7–11.

Ram, R. and S. Ural (2014). "Comparison of GDP Per Capita Data in Penn World Table and World Development Indicators," *Social Indicators Research*, 116, 639–646.

Reinhart, C. M. and K. S. Rogoff (2010). "Growth in a Time of Debt," *American Economic Review: Papers & Proceedings*, 100, 573–578.

Rey, S. J. and B. D. Montouri (1999). "US Regional Income Convergence: A Spatial Econometric Perspective," *Regional Studies*, 33(2), 143–156.

Rodrigues Júnior, W., P. H. Melo Albuquerque, C. Rey Laureto, G. P. Marina, R. Dantas Guimarães, and G. Rezende de Almeida Júnior (2010). "Augmenting the Solow-Romer Model with Spatial Externalities: An Application to the Brazilian Case," in *32º Encontro Brasileiro de Econometria*.

Rodrik, D., A. Subramanian, and F. Trebbi (2004). "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development," *Journal of Economic Growth*, 9, 131–165.

Romer, P. M. (1986). "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94, 1002–1037.

——— (1990). "Endogenous Technological Change," *The Journal of Political Economy*, 98(5), S71–S102.

Ross, S. M. (2010). *Introduction to Probability Models*, 10th Edition, Elsevier, Amsterdam.

Samuelson, P. A. (1954). "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *The Economic Journal*, 64(254), 264–289.

Sarel, M. (1996). "Nonlinear Effects of Inflation on Economic Growth," *IMF Staff Papers*, 43(1), 199–215.

Schwarz, G. (1978). "Estimating the Dimension of a Model," *The Annals of Statistics*, 6(2), 461–464.

Shekhar, S. and H. Xiong (editors) (2008). *Encyclopedia of GIS*, Springer, New York.

Smith, A. (1904). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Methuen & Co, London.
URL: `http://www.econlib.org/library/Smith/smWN.html` *(Library of Economics and Liberty), accessed: 13 August, 2015.*

Solow, R. M. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70(1), 65–94.

——— (1994). "Perspectives on Growth Theory," *Journal of Economic Perspectives*, 8(1), 45–54.

Spolaore, E. and R. Wacziarg (2009). "The Diffusion of Development," *The Quarterly Journal of Economics*, 124(2), 469–529.

——— (2015). "Ancestry, Language and Culture," Working Paper 21242, National Bureau of Economic Research.

Starrett, D. (1978). "Market Allocations and Location Choice in a Model with Free Mobility," *Journal of Economic Theory*, 17, 21–37.

Swan, T. W. (1956). "Economic Growth and Capital Accumulation," *The Economic Record*, 32(2), 334–361.

Sydsæter, K., P. Hammond, A. Seierstad, and A. Strøm (2008). *Further Mathematics for Economic Analysis*, 2nd Edition, Prentice Hall, Harlow.

The Economist (2013). "The Missing $20 Trillion - How to stop companies and people dodging tax, in Delaware as well as Grand Cayman," February 16.
URL: *http://www.economist.com/news/leaders/21571873-how-stop-companies-and-people-dodging-tax-delaware-well-grand-cayman-missing-20, accessed: 11 August, 2015.*

Thom, R. and M. McDowell (1999). "Measuring Marginal Intra-Industry Trade," *Weltwirtschaftliches Archiv*, 135(1), 48–61.

Thomas Jr., G. B. (2005). *Thomas' Calculus*, 11th Edition, Addison-Wesley, Boston.

Tobler, W. R. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 46, 234–240.

Turnovsky, S. J. (2000). *Methods of Macroeconomic Dynamics*, 2nd Edition, MIT Press, Cambridge.

Uchida, H. and A. Nelson (2010). "Agglomeration Index: Towards a New Measure of Urban Concentration," Working Paper 2010/29, World Institute for Development Economic Research.

United Nations, Department of Economic and Social Affairs, Population Division (2012). *World Urbanization Prospects, the 2011 Revision*, Final Report with Annex Tables, New York.

United States Census Bureau (2015). *Current Population Survey.*
URL: *http://www.census.gov/hhes/socdemo/education/data/cps/index.html, accessed: 22 June, 2015.*

Varian, H. R. (1992). *Microeconomic Analysis*, 3rd Edition, W.W. Norton & Company, New York.

Verbeek, M. (2004). *A Guide to Modern Econometrics*, 2nd Edition, John Wiley & Sons, Ltd, Chichester.

Vincenty, T. (1975). "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations," *Survey Review*, XXII, 88–93.

von Thünen, J. H. (1826). *Der isolirte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, oder Untersuchungen über den Einfluß, den die Getreidepreise, der Reichthum des Bodens und die Abgaben auf den Ackerbau ausüben*, Friedrich Perthes, Hamburg.
URL: *http://www.deutschestextarchiv.de/book/show/thuenen_staat_1826, accessed: 29 July, 2015.*

Wälde, K. (2011). *Applied Intertemporal Optimization, Edition 1.1*, Mainz University Gutenberg Press.

Weil, D. N. (2014). "Health and Economic Growth," in P. Aghion and S. N. Durlauf (editors), *Handbook of Economic Growth*, Volume 2B, Chapter 3, Elsevier B.V., Oxford, 623–682.

Wetterstrand, K. (2015). "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)," .
URL: *http://www.genome.gov/sequencingcosts/, accessed: 17 August, 2015.*

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*, 5th Edition, South-Western, Mason.

World Bank (2009). *World Develoment Report 2009: Reshaping Economic Geography*, World Bank, Washington.

——— (2013). *World Development Indicators*, World Bank.

World Trade Organization (2011). *World Trade Report 2011 - The WTO and preferential trade agreements: From co-existence to coherence*, World Trade Organization, Geneva.

Wright, S. (1951). "The Genetical Structure of Populations," *Annals of Eugenics*, 15, 323–354.

Yaari, M. E. (1965). "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *The Review of Economic Studies*, 32(2), 137–150.

Yamarik, S. (2006). "Solow and the States: New Evidence," *Regional Studies*, 40(6), 571–582.

——— (2011). "Human capital and state-level economic growth: what is the contribution of schooling?" *The Annals of Regional Science*, 47, 195–211.

——— (2013). "State-Level Capital and Investment: Updates and Implications," *Contemporary Economic Policy*, 31(1), 62–72.

Young, A. (1998). "Growth without Scale Effects," *Journal of Political Economy*, 106(1), 41–63.