Knowledge Diffusion Processes: Theoretical and Empirical Considerations

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

vorgelegt von

Diplom-Volkswirt Univ. Torben Alexander Klarl

Augsburg, im Juni 2009

Erstgutachter:	Prof. Dr. Horst Hanusch
Zweitgutachter:	Prof. Dr. Alfred Maußner
Vorsitzender der mündlichen Prüfung:	Prof. Dr. Fritz Rahmeyer

Tag der mündlichen Prüfung:	23. Dezember 2009

Für meine Eltern und Unterstützer

Contents

List of Tables

List of Figures

1 Knowledge in economics

In March 2000 the Members of the European Council set a new and very ambitious goal, which states to become by 2010 "the most competitive and dynamic knowledge-based economy in the world, capable of sustainable growth with more and better jobs and greater social cohesion". Throughout the 20^{th} century, the question, what knowledge is or what it could be, and disputes about a concise definition of the term knowledge and its economic importance has been engaged several economists. Nowadays, there is consensus, that knowledge plays an important role not only in economics but also in politic debates, such as propagated by OECD and EU. The so called "knowledge society" is one crucial topic on the research agenda of the "socio-economic sciences and humanities (SSH)" programme which is embedded into the "Seventh Framework Programme (FP7)". The aim of SSH is to advance the understanding of the socio-economic challenges facing Europe in the near future. Besides, growth, employment, social cohesion, education, migration and sustainability, the major aim of FP7 is to maintain a leadership in the global knowledge economy. Hence, knowledge can be recognised as one of the competitive advantages in a globalised economy.

Although the "knowledge based society" is a hot topic from a politician point of view, the idea, that knowledge should be treated as a key determinant for economic development, is not new nor it is a new introduced fact. All areas of economic development are based on knowledge. But since the industrial revolution the degree of knowledge and information has become so great that knowledge itself, as an input factor of production for instance, exhibits a strong influence on economic development. Hence, the increasing knowledge intensity in the globalised economy needs to focus on the determinants of the "knowledge based society". Two major determinants on which the "knowledge based society" relies are the creation and the diffusion of knowledge, besides the use of knowledge. The development of new information and communication technologies has favoured particularly the knowledge diffusion but also enforces indirectly knowledge creation. As laid out in the FP 6 the "knowledge based society" is not only focused on national but also on regional level. Regions as "knowledge laboratories" should ensure long run economic development and networks between regions should promote knowledge diffusion.

From an economic point of view, the research field of the economic analogon of the "knowledge based society", the so called "knowledge based economy", is rather new and gives several revenues for research. In particular, the role of network effects for knowledge diffusion, the role of knowledge in the production process and the question which effect exhibit territorial structures such as spatial proximity of regional knowledge laboratories on regional growth are predominant topics on the research agenda and define the motivation for this work.

This chapter provides not only a short overview of specific theoretical and empirical knowledge creation and diffusion topics mentioned in early economic related literature but also focus on more recent strands of the relevant literature. The aim of this chapter is not only to recapitulate research results but also to highlight economic intuition of the models which have been used in this thesis. Finally, the last section of this chapter deals with concretion of the research fields based on the topics laid out in this chapter and the derivation of research questions which will be discussed in the following chapters.

1.1 Basic considerations

From an economic growth point of view, knowledge is not only treated as a pure input factor in production process, but also as the result of a production process itself. Hence, sources of new knowledge are commonly associated with learning-by-doing, accumulation of human capital, R&D or patent activity and via spillovers generated by universities[1]. From empirical analysis, two (Kaldor and Mirrlees, 1962) stylized facts are of importance with respect to knowledge: first, the average product of labour is decreasing over time and second, labour productivity varies over regions. The last stylized fact can be explained with (Arrow, 1962). He proposes, that knowledge generation itself is often assumed to be path depended: a historically given knowledge stock determines the creation of new knowledge. In this way, historic endowment of knowledge can be used to predict future's knowledge stock. In consequence of that and with respect to regions, lurching effects are excluded and everlasting regional divergence can be observed. Therefore, knowledge has an explicit time dimension. But from (Arrow, 1962) an another important character of knowledge can be derived: knowledge is space depended, or has a spatial dimension as already mentioned by (Hayek, 1945).

Another important feature of knowledge is its context dependence. (Polany, 1967) has pointed out that knowledge has an implicit and explicit context. Implicit knowledge cannot be captured instantaneously, over time and space and it must be transmitted by personal contact, e. g. via face-to-face communication, because it is often embodied.

[1]Refer to (Audretsch and Feldman, 1996) and (Audretsch, 1998).

Explicit knowledge, on contrast can be easily articulated, transferred and saved and is often disembodied.

Closely linked with knowledge creation is knowledge diffusion which was implicitly addressed before. With respect to the dimension of knowledge we have to distinguish between context, time and spatial knowledge diffusion. Practically and theoretically, knowledge diffusion is hardly to measure, because "[k]nowledge flows are invisible; they leave no paper trail by which they may be measured and tracked[...]", as stated by (Krugman, 1991). The diffusion process itself can be imagined as an epidemic or as a hierarchical phenomenon. The first assumes that from a given source knowledge diffusion spreads uniformly over space, the latter instead interprets knowledge diffusion as depended from agglomeration phenomena: Knowledge first flows from the source to agglomerated areas and then with a certain delay to peripheral economic areas.

To sum up, knowledge creation and knowledge diffusion are not only context or problem based dependent, but further have a time and a spatial dimension.

1.2 Knowledge diffusion, knowledge transfer and network effects

Economists and sociologists both seek better understanding, of why some knowledge disperses widely whereas other knowledge does not exhibit this kind of pattern. As mentioned above, one reason could be that knowledge itself should be treated as heterogeneous. For instance, if absorptive capacity is required to understand tacit knowledge and further a certain group of people does not have the qualification in terms of skills, then the implication is, that knowledge diffusion tends to be very slow, et vice versa.[2] As a result, this group cannot benefit from new technologies or other applications which contain a significant knowledge component, as highlighted by (Henderson and Cockburn, 1996), (Teece, 1998) or (McEvily and Chakravarthy, 2002).

Some of the early models, that study diffusion of innovation and knowledge flows, are the so called epidemic diffusion models. This kind of models assume, that a given number of potential adopters exists, which adopt a new innovation, or in more medical terms are inflected by a new innovation due to external and internal influence. The communality of these models is that the cumulative adoption follows a sigmoid pattern. This corresponds to the idea, that adoption at the beginning is slow, then it is sharply rising in the middle, followed by a slabbing adoption tendency at the end. Hence, this

[2]Refer to (Cohen and Levinthal, 1990) for this topic.

model inherently acknowledges a certain kind of interpersonal communication within the group of adopters and potential adopters, as mentioned right before. Thus, it is necessary, that at least one member of the population has adopted the innovation. This can be justified with the assumption that some members of the group exhibit innovators behaviour and show search activity or other idiosyncratic features (Griliches, 1957), which are closely related to innovators actions. This indemnifies that innovators have adopted the new technology right from the beginning of the diffusion process.

Diffusion stops automatically, after an exogenous market saturation potential has been equalized by the number of adopters and hence the number of potential adopters tends to zero. In the beginning of the adoption process, diffusion is relatively slow, because potential adopters wait until adopters have communicated them some characteristics of the new knowledge. Through knowledge transfer, which is not a sufficient condition of knowledge diffusion, diffusion can be accelerated. It is worth mentioning that knowledge transfer means the technical transfer mechanism of knowledge via face to face communication for example. Knowledge is diffused, if one can benefit from using new knowledge, because she is able to understand it. Of course, the acceleration depends positively on knowledge transfer possibilities. Right after the inflection point of adoption has passed, acceleration of knowledge diffusion stagnates until market saturation potential has reached, which implies that acceleration speed of diffusion becomes zero. This scenario can be described as throwing a stone into water, and waiting until the ripples have steadily spread over the entire surface.

Albeit that immagination of knowledge diffusion seems pretty easy. Researches found, that adoption of innovation and knowledge over time can be accurately described by an S-shaped pattern, as mentioned by (Hargadon, 1996). Over the years, starting with (Bass, 1969), the S-shaped diffusion models have found wide acceptance in economics, especially in economics of innovation and empirical marketing research. (Rogers, 1983) provides a formidable review of the advancement of this type of diffusion models. Some of the models explicitly abandon the assumption that adopters are homogenous and introduce an innovators and imitators relations[3]. Others include Bayesian learning such as (Oren and Schwartz, 1988) and (Chatterjee and Eliashberg, 1990).

Although, these types of diffusion models are heavily used in knowledge diffusion application, they exclude relevant aspects of knowledge diffusion: the role of networks for knowledge diffusion. As shown by (Hansen, 1999) strong network relations are necessary to transfer complex knowledge from sender to receiver. The implication is that networking and the ability of successful knowledge transfer are closely related.

[3]Refer to the models of (Tanny and Derzko, 1988) and (Van den Bulte and Joshi, 2007).

Another fact, which is of importance for knowledge diffusion and also related to knowledge transfer, is the consideration of feedback loops which are predominant in (knowledge) networks. But in this context it is assumed that knowledge transfer, if it happens, happens without errors. Of course, this assumption neglects uncertainty of knowledge transfer. From this point of view, knowledge transfer itself should be treated as a trial-and-error process as noted by (Sorenson et al., 2005).

1.3 Knowledge diffusion and learning, firm size and market structure

Diffusion of knowledge generally depends, as mentioned before, on communication channels and social networks, but also on the personal ability to understand new knowledge and thus on absorptive capacity of the adopters and potential adopters. As mentioned by (Cohen and Levinthal, 1990), absorptive capacity is the ability to identify, accumulate value, assimilate, transform, and exploit knowledge resources to enable learning. (Arrow, 1962) noticed on page 156 that (technical) learning can be seen as the result of experience engaging in a certain activity itself. He also stressed the time aspect of learning, because undertaking learning activity leads to "favourable responses [which] are selected over time"[4]. In the relevant literature, the sigmoid curve is the most used for modelling learning-curves[5]. The idea to model learning curves particularly with sigmoid functions originally stems from psychology. There is a close relationship between absorptive capacity, learning and knowledge diffusion. If an individual has reached the saturation level of learning curve, then individual absorptive capacity should take the maximum level.

If learning effects are in place on firm level, this should yield a reduction of production costs, using a new technology, which is based on a process innovation as explained by (D'Aspremont and Jacquemin, 1988) and (Kamien et al., 1992). Closely related to these before mentioned articles is the work of (Yildizoglu, 2002), who found that learning effects in firms lead to more efficiency and finally to social welfare gains. In consequence, the more learning curve effects can be exploited, the faster production costs can be reduced.

Of course, learning in this sense has a strong normative attitude. A large bulk of literature deals with the question, how knowledge can be retained in firms and how

[4](Arrow, 1962), p. 156.

[5]Refer for example to (Young and Ord, 1989), (Gamerman and Migón, 1991), (Meade, 1988) and (Meade and Islam, 1995).

learning affects firm's specific strategy and structure. The link between learning and how it can affect firm structure is sketchy. This is owed to the fact, that learning new knowledge cannot be observed directly, as mentioned above. From an empirical point of view, R&D effort is often used as a proxy for knowledge accumulation as a result of learning endeavour.[6]

The interaction between innovation of new technologies and competitiveness is essential in evolutionary economics for discussing market structure development. The link is straightforward: as mentioned above and further highlighted by (D'Aspremont and Jacquemin, 1988) and (Kamien et al., 1992) cost reduction potential is generally linked with innovation activity. Going ahead, different cost potential in a given economy should also affect firm structure and also market structure. Indeed, this presupposes that firms are heterogeneous in terms of cost potentials. The question now occurs, how firm heterogeneity, learning efforts, and market structure can be acknowledged into a model frame. Replicator dynamics do a formidable job in this purpose. The drawback of replicator dynamics models, which idea is based on natural selection and mutation phenomena, is that they often cannot be solved analytically, but only numerically on the basis of simulation studies.[7] This is especially valid, if one assumes that innovative activity is modeled explicitly as a trial- and -error-process.

Market structure can be described by the following different "structural factors", as noted by (Malerba et al., 1997): concentration and asymmetries between innovative firms, firm size, evolution of the ranking of innovative firms, and the importance of new innovations with respect to existing innovations. The first factors should answer the question, if innovative activity is either concentrated on fewer firms or equally distributed over the entire population of firms and whether small or large firms are innovative drivers. These curled questions are closely related to the so called *Schumpeter-Mark-I* and *Schumpeter-Mark-II* hypothesis, labeled by (Malerba et al., 1997). As highlighted by (Schumpeter, 1912) creative destruction is particularly caused by small and young firms which cause market instability. In the relevant literature this scenario is labelled *Schumpeter-Mark-I*. On contrary to (Schumpeter, 1912), (Schumpeter, 1942) singled out the importance of established large firms which dispose of an own R&D division, for market stability. This scenario is labelled *Schumpeter-Mark-II*. Therefore, the question arises, under which conditions market structure coincides with *Schumpeter-Mark-I* or *Schumpeter-Mark-II*. (Arthur, 1989), (David, 1985) and (Klepper, 1996) argue that dynamic increasing returns to scale create lock-in-dynamics which allow firms to grow persistently more than for other firms in the same market.

[6]See (Geroski and Mazzucato, 2002) for this topic.
[7]See (Kwasnicki, 1996), (Kwasnicki and Kwasnicka, 1992) and (Saviotti and Mani, 1995).

This success-breeds-success-scenario is closely related to *Schumpeter-Mark-II* scenario, whereas *Schumpeter-Mark-I* is associated to decreasing returns to scale.

Empirically, there have been a large number of papers, which investigate the link between firm size, market structure and innovation activity. (Comanor, 1967) and (Dosi, 1984) found, that markets should tend to be much more concentrated in industries with a low rate of innovative activity, or on a macro perspective and given the industry, that countries are more innovative if they are innovative leaders. On contrary, (Audretsch, 1995), (Abernathy and Wayne, 1974), (Klein, 1977), (Tushman and Nadler, 1986), (Markides, 1998) and (Christensen, 1997) found, that market share instabilities are much more likely in markets in which small firms are more innovative than large firms. (Dosi and Orsenigo, 1985) come to the conclusion that firm homogeneity goes in line with less concentrated markets.

To sum up, the link between firm size, innovation and market structure is investigated in several studies. Increasing returns to scale, caused by learning-by-doing for example, are related to *Schumpeter-Mark-II* scenario, whereas decreasing returns to scale are more appropriate for characterizing market structure in which small firms are dominant. But although network effects and learning exhibit an impact on market structure[8], these elements are ignored so far in the relevant literature.

1.4 Knowledge diffusion, scale effects and spatial proximity

Since Alfred Marshall, several economists are engaged to define the question what diffusion of knowledge is. Marshall gave a new insight into this question. He noted that "[...]if one man starts a new idea, it is taken up by others and combined with suggestion of their own; and thus it becomes the source of new ideas."[9] From this cognition, Marshall concluded that firms should profit from densely populated areas. Especially, regions which are endowed with a high qualified labour stock, which is often called human capital, and a widely differentiated supply of highly specialized suppliers in the region specific industries should provide a fruitful framework for knowledge diffusion.

For this reason, Marshall can be named as the door opener of a new discipline which is a valuable source of new insights into the topic knowledge diffusion: the new economic geography, which development was mainly driven by (Krugman, 1991) and (Fujita

[8] Refer to (Campagni, 1991), (Best, 2001) and (Porter, 2000) for this topic.
[9] (Marshall, 1920), p. 225.

and Thisse, 2002) and (Brakman et al., 2001), who found that uneven distribution of economic development is mainly caused by agglomeration effects of mobile production factors. Aboriginally, the new economic geography was mainly designed to explain trade patterns within a country, e. g. between regions, or between countries, with the focus on intra- and interindustrial trade flows. From this point of view, the new economic geography logically expands the theoretical new trade perspective with an empirical view. Particularly, this kind of econometric application should present a comprehensive approach to identify spatial trade patterns within data. (Helpman and Krugman, 1985) were the first, who introduced a growth model with a new economic geography context.

Closely related to those models of the (Helpman and Krugman, 1985) type mentioned before, in which trade costs play a crucial role, another focus on new growth theory is, that production of knowledge generates positive external effects by assumption. Of course, external effects can also be negative, such as pollution, but it is common to focus on the positive side of externalities as mentioned before. Again, in general the generator of new knowledge cannot entirely appropriate the new knowledge completely, but has to worry about the fact that a third party can participate from this new knowledge without costs. In an extreme case, where knowledge cannot be appropriated entirely, knowledge has to be characterized as a public good. Generally, it can be assumed that knowledge contains a specific public good item, but can be appropriated by the knowledge generator[10]. The more knowledge has been accumulated in the past in this way, the more current production is influenced. Thus, we have to label these effects as dynamic positive externalities.

It depends on the space and on the kind of knowledge, whether such effects have only a productivity increasing effect in the contiguous neighbourhood, which we can label as a small cluster, or if such effects have sweeping effects in space. In the later case we should expect wide cluster effects.

From this point of view, economic geography is not only relevant for explaining growth effects of knowledge diffusion on macro level, but also on regional level. Despite the blatantly relevance of the spatial effect of knowledge diffusion especially on regional level, such effects have been neglected quite often in recent literature covering the topic of knowledge diffusion.

One of the reasons why those spatial effects have played a minor role in economic context could be, that different conceptions of knowledge diffusion are employed in the relevant literature. On the one hand, so called "cumulative-causation-models" assume that technological know-how, and thus knowledge, which is often embodied through

[10]Refer to (Tirole, 1995).

new technologies, is entirely immobile. This could be due to the fact, that knowledge, such as experience is inherently tacit, which means that knowledge is linked to a human being[11]. The implication of those models is, that knowledge spillovers are more or less excluded. Tacit knowledge could be an explanation why the catching-up process of regions is left out, by virtue of missing knowledge spillover processes.

On contrary, early approaches of neoclassical growth theory assume, that knowledge transfer is a natural given phenomenon, which means that knowledge can be transferred immediately, without costs and any difficulties from the sender to receiver. In this context, knowledge is disembodied, and spatial proximity to sender is not of importance. In this context, knowledge is, as mentioned before, more or less a public good. As a consequence, those early neoclassical approaches fail to explain regional growth and income disparities, which could be caused by regional knowledge and technology differences.

(Romer, 1990), (Grossman and Helpman, 1991a), and (Aghion and Howitt, 1992) for instance take explicitly R&D as profit-maximizing activities and technological progress as the result of these activities into account. Particularly, (Romer, 1990) assumes that researchers of a firm create a new kind of knowledge for the production of a new homogeneous good, driven by monopoly profits from the final product sales of the good. The key is, that although the production of these goods is monopolized, the stock of knowledge created, can be accessed by the entire population in the economy and researchers use them for free to generate new knowledge. The production of knowledge is the key parameter of growth in this model context, by which the speed of innovation is treated proportionally to the number of researchers in R&D. Hence, population size produces positive effects on the the GDP per capita, because if population increases, the speed of innovation becomes also faster and also does the growth rate of GDP per capita.

If we take a look at post war data, there is minor support for the theoretical findings of the (Romer, 1990) model. (Backus et al., 1992) have found for 57 countries during 1970-1985 that pure number of researches in R&D sector have not the expected effects on the average growth rate of GDP per capita. (Jones, 1995b) has shown for US data that the number of engineers and scientists engaged in R&D increased since 1954 from 237,000 to a million in 1995. Comparable observations have been conduced for other industrialized countries, such as Germany, France and Japan for example. Given the (Romer, 1990) model, this should lead to an increased GDP growth rate per capita in the same time span. But the GDP growth rate for the United States of America has been roughly constant at about 2% over the same time interval. As a

[11] Refer to (Myrdal, 1959) for this topic.

consequence and on the basis of the finding of (Jones, 1995b), subsequent approaches have tried to eliminate those scale effects. For instance (Jones, 1995b), (Kortum, 1997) and (Segerstrom, 1998) have assumed diminishing returns in the production of new technological knowledge which results in a felicitously elimination of scale effects. Thus, the development of new innovations becomes more difficult as the underlying technology improves and more researchers are required to let the speed of innovation unaffected. The implication is, first that scale effects appear in the level of per capita income instead of its per capita growth rate. Second, research subsidies schemes only affect level of income but not its long run growth rate.

(Aghion and Howitt, 1998) in chapter 12, (Dinopoulos and Thompson, 1998), (Peretto, 1998) and (Young, 1998) instead follow a new strategy to eliminate the scale effect by considering two different types of R&D: innovation of new goods and improvement of existing or quality goods. Hence a new dimension has been added into the models of the (Romer, 1990), (Grossman and Helpman, 1991a) and (Aghion and Howitt, 1998) type. The argumentation in this framework is that a growing population leads to an increase in the number of goods, but does not affect the number of researches for each specific good, because quality improvement instead of new good innovation is of importance. But within these models, there remains still an inconsistency: although population size does not affect GDP growth, the growth rate of population does. The consequence is, that scale effects on levels occur within these models as pointed out by (Jones, 1998). However, if we look at the data, there is no distinct relation between the growth rate of population size and GDP per capita. On the one hand it is found that scale effects on GDP level are not supported by post war data.[12] But on the other side, there is evidence, that in the long run history scale effects on economic growth can be found (Kremer, 1993). This seems to be puzzling. One attempt to resolve this puzzle is to accept the concept of international knowledge diffusion, which requires open-economy models. (Todo, 2001) opens the (Romer, 1990) model in two aspects: First he assumes that international knowledge diffusion is costly. Second, intranational knowledge diffusion is treated as in the (Romer, 1990) model as a costless phenomenon. In this context (Jaffe and Trajtenberg, 1996), (Jaffe and Trajtenberg, 1998), (Jaffe and Fogarty, 2000), (Jaffe and Henderson, 1993) and (Brendstetter, 2001) have performed studies regarding knowledge diffusion based on patent citations. The idea behind those works is, that home country patents are cited with a higher probability by patents from the home country, which implies that new knowledge is more or less country specific and thus knowledge spillovers are spatial limited. There is some evidence that international knowledge diffusion can reconcile the post war inconsistency of scale effects in

[12]Refer to (Todo, 2001).

growth implied by the (Romer, 1990) model.

We conclude, that the early neoclassic assumption that knowledge diffusion can be characterized by a process which is costless, spatially unlimited and possible in every period of time is inconsistent with the data. From the discussion above we have seen, that post war inconsistency can be at least partly resolved by the assumption of international knowledge diffusion, which is done in several contributions in the new growth literature and new economic geography, as mentioned above. Such scale effects should be taken into account, as well as the economic role of space within the knowledge diffusion process.

To sum up, we have to notice that neither the neoclassical approach nor the cumulative-causation approach seems to be appropriate to explain knowledge diffusion. The reality is stacked somewhere in the middle between these two approaches: Knowledge can diffuse, but one has to keep in mind that knowledge diffusion is a function of the kind of knowledge and of space. Although the last assumption follows from the scale effects discussion within the new growth theory, the first is not well established in the strand of research.

1.5 Knowledge diffusion and spatial econometrics

The transformation from an industrial to a more or less knowledge-based economy is one of the key challenges for political institutions. "To become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion".[13]

From this excerpt one can easily conclude that, particularly the concept of learning regions has been preached as a strategy for successful future development. Further, as another implication of the above mentioned excerpt and as laid out before, learning and knowledge diffusion are closely interlinked or more precisely, knowledge diffusion and learning are reciprocally presupposed.

As a consequence, knowledge diffusion should be primarily seen not as a country level, but as a supra-national or subnational entity phenomenon, which makes sense mainly because of the following reason: knowledge diffusion is evolving not evenly in space, by virtue of heterogeneous endowment of production factors. The implication is, that policy should foster network effects, to create an environment for learning regions. Especially, the European Union (EU) is championing the concept of learning

[13]Refer to (EU, 2004).

regions, which aims to reduce interregional disparities. Within the "learning regions" or Regional networks of Life-Long-Learning (R3L) initiative launched in April 2003 the EU[14] is highlighting the importance of learning for regional economic growth.

(Florida, 1995) gave a very precise definition of what makes a region a learning region: "The new age of capitalism requires a new kind of region. In effect, regions are increasingly defined by the same criteria and elements which comprise a knowledge-intensive firm-continuous improvement, new ideas, knowledge creation and organizational learning. Regions must adopt the principles of knowledge creation and continuous learning; they must in effect become learning regions. Learning regions provide a series of related infrastructures which can facilitate the flow of knowledge, ideas and learning."[15]

From this point of view, the EU policy aims to create knowledge networks, from which regions can benefit via knowledge spillovers which should in the long run lead to cohesion of regions as highlighted by EU[16]. Hence, the concept of a learning region, which implicitly fosters knowledge spillovers potential between regions, should incorporate not only innovation orientated approaches, which should foster the diffusion of knowledge, but also a policy approach, which is focused on the sustainable creation of networks of regions and of course a human capital element which is a precondition for creating a knowledge base. Prima facie this seems to be reasonable.

But if we look again at the goals of the Lisbon Agenda, we can find that it bears an inherent conflict of aims: It is not possible to foster economic growth, which goes hand in hand with agglomeration tendencies, as mentioned above due to spillovers, and cohesion on the other side. Thus, it is worth to ask the question, whether spatial knowledge spillover exist, and if yes to which extent they can contribute to explain growth effects. Are knowledge spillovers more or less local, or global regarding their grasp? As a consequence, if knowledge spillovers are more local, then policy as the Lisbon Agenda should set their key aspects of activities on local level. Therefore, the question which one should primarily be focused is, how important are knowledge spillovers in a spatial environment.

These questions, especially the last, can of course only be answered empirically. In the past several studies have contributed to these questions, not only in the macroeconomic sphere, but also on the microeconomic level. The macroeconomic method is to measure knowledge spillovers via patent citations or R&D efforts, where the distance of patentee and locations of patent's citations is from relevance. On the microeconomic sphere, firm level data could reveal spatial phenomena. For example location decisions

[14] Refer to (EU, 2003) for an overview.
[15] (Florida, 1995), p. 532.
[16] Refer to (EU, 2003).

could depend on infrastructure and life quality of a region. With this approach, cluster phenomena, for instance the greater Munich area, could be explained.

An attractive method to deal with spatial phenomena empirically is the employment of spatial econometrics methods, which can be described as a sub discipline of new economic geography. Because of the fact that spatial econometric tools primarily focus on regional science application, this methods seem to be appropriate to deal with spatial knowledge spillovers.

Spatial econometrics can be distinguished from traditional econometrics in two directions: the first view could be to argue, that all econometric models in regional science fall automatically within the spatial econometric toolbox. This distinction seems to be rather flippantly and therefore not appropriate. Instead, a second view focus on the spatial characteristics of the data itself. If these data precludes the application of traditional established methods in econometrics, due to spatial effects such as spatial dependence and spatial heterogeneity, then those non standards methods can be subsumed under the term spatial econometrics.[17]

Spatial dependence can be due to several reasons. The first reason why spatial dependence could occur is misspecification. This is already known from traditional econometric applications, in the context of time series analysis. Although on the first sight, spatial dependence seems to be directly comparable to the phenomenon of autocorrelation from time series context. This is only partly the case. Spatial dependence occurs primarily in cross-section or panel applications, whereas autocorrelation is a time series problem. Spatial spillovers do not have a clear directional development towards time, in contrast to time series, but should be, instead of time series argumentation, characterized with feedback effects. Although (Kmenta, 1971) has worked out this problem as precisely as possible, until today, neither standard econometricians academic book, nor any standard econometricians toolboxes, such as EViews do provide space for spatial applications. Hence, if the researcher tries to cope with this problem he has to program its own spatial estimation routines employing `GAUSS`, `Mathematica`, `Matlab` or `R`. Today, we find sporadically some application toolboxes for spatial model estimation, the toolbox of LeSage designed for `Matlab`, the open source `R`-based programme `Geoda`, the `R`-package `spdep` or the `Stata` toolbox `spatreg`. (Kmenta, 1971) argues that "In many circumstances the most questionable assumption [...] is that the cross sectional units are mutually independent. For instance, when the cross-sectional units are geographical regions with arbitrarily drawn boundaries-such as the states of the United States - we would not expect this assumption to be well satisfied."[18]

[17]Refer to (Anselin, 1988) for an introduction.
[18](Kmenta, 1971), p. 512.

On contrast to spatial dependence, the idea behind spatial heterogeneity is, that spatial entities such as regions are not homogeneous. This phenomenon likely occurs in cross-sectional application, when using dataset including both poor regions from the South and rich regions from the North. More technically spoken, the assumption of constant variance over space is not justified within such a setup. It is easy to conclude that neglecting spatial heterogeneity although it matters leads to estimations of parameters which violate the *Gauss-Markov* assumption. Sometimes it is difficult to differentiate between spatial heterogeneity on the one hand, and spatial dependence on the other hand, due to the fact, that often a combination of both effects occurs.

Independently, if spatial dependence or spatial heterogeneity is relevant, the model itself can be estimated with conventional maximum likelihood methods. Additionally to the before mentioned so called frequentest methods, so called Bayesian methods have been prevailed and proved itself in spatial econometric application. The difference between both approaches is, that Bayesian methods treat the coefficient vector of estimators itself as random, whereas frequentest say that the resulting estimates of the coefficient vector is random. Bayesian methods hold a great deal for several reasons: first it is possible to model hierarchy of places or regions, second a more or less systematic change of variance over space, and thus spatial heterogeneity and third a hierarchy of regions or places. Bayesian methods can incorporate these ideas because of its underlying concept to employ prior information additionally to existent sample data information, whereas frequentest methods can solely rely on the latter mentioned. As mentioned before, although Bayesian methods seem to be very attractive, their usage in application is very limited. On the other side as mentioned above, frequentest methods lead to insufficient parameter estimates, if spatial heterogeneity is neglected and only for spatial dependence it is consistently controlled.

Closely related to the question of existence of spatial knowledge spillovers is the question, to what distance does spatial proximity matter with respect to knowledge spillovers. And if yes, is the relevance of influence of spillovers a constant or not a non constant function of space, which implies that spatial strength of spatial influence depends on contiguousness. Most of existing studies do not contribute to the question how far knowledge spillovers reach. (Anselin et al., 1997), (Varga, 1998) and (Anselin et al., 2000) are one of the few studies that have mentioned concrete numbers of knowledge spillover scope. (Anselin et al., 1997), (Varga, 1998) and (Anselin et al., 2000) found by investigating the influence of university related research and private R&D effort on knowledge transfer that a significant positive effect can be detected within a 50 mile radius of Metropolitan Statistical Areas (MSAs) only for the university research. For private R&D such a significant effect could not be detected. (Varga, 1998), with a similar setup as (Anselin et al., 1997) show, that not only spillovers within MSA but also between MSA can be found. But also without exact geographical distance measures, limited spatial influence can be measured. (Audretsch and Mahmood, 1994) show on patent basis for 59 US metropolises, that knowledge spillovers are limited towards the metropolis boarders. They come to this conclusion because they found that only for research institutes which are settled within a metropolis, significant knowledge spillovers can be detected, whereas for research institutes, settled in each metropolis related country, no such effects could be found.

For Germany (Funke and Niebuhr, 2000) is one of the view studies, which directly investigates the scope of knowledge spillovers, but only for West German data. They use data of 75 so called "Raumordnungsregionen", and find that the half-life of knowledge spillovers lies in a range between 23 km and 30 km, where agglomeration areas are the source of knowledge diffusion. Similar results have been obtained by (Greif, 1998) and (Frauenhofer, 2000). (Badinger and Tondl, 2002) found similar evidence for 159 EU regions. They have noted that capital, human capital and knowledge transfer play decisive roles for regional growth. More general, (Bottazzi and Peri, 2003) concluded that knowledge diffusion is significant within 300 km distance for European regions. Common for all studies is that they neglect possible spatial heterogeneity, which should not be excluded ex ante.

So far we know something about the spatial scope of knowledge spillovers. Closely related to the scope of knowledge spillovers is, what type of knowledge is relevant for explaining (regional) growth differentials. In related literature, there is made a distinction between so called urbanisation externalities and location externalities. So called MAR externalities, which follow the idea of (Marshall, 1920), (Arrow, 1962) and (Romer, 1986), assume that knowledge transfer takes place between firms within a branch. Hence, firms itself are assumed to be similar. In contrast to MAR externalities, so called (Jacobs, 1970) externalities describe spillovers between different industries, which lead to the exploitation of so called economics of scope potentials. Although the empirical differentiation of knowledge spillovers seems to be plausible to some extent, from empirical evidence we have to conclude that spillover effects in combination, hence a combination of Jacobs and MAR externalities, are relevant for explaining economic development, as concluded by (Forni and Paba, 2001).

In summary, spatial econometrics mostly deals with the treatment of spatial dependence and spatial heterogeneity in both cross-sectional and panel data model contexts. The focus on space, as an important economic dimension has not only gained attention in theoretical growth model context, but also more or less recently in applied econometrics context, albeit the focus of spatial econometricians application mainly lays on spatial dependence. The focus on spatial heterogeneity and the closely linked question to what extent, regarding space, knowledge spillovers matter, are less acknowledged in the relevant literature, for German regions.

1.6 Motivation for further research

From the discussion above, we can subsume, that knowledge diffusion plays a crucial role, not only on a macro level growth context, but also on a more traditional micro level. Knowledge and knowledge diffusion are important cornerstones and elements in modern economics, as highlighted before. Although knowledge and with some cutbacks also knowledge diffusion seem to be well established in economic theory and empirics, there are more implications of knowledge diffusion which are not acknowledged in the cited literature. Of course, the economic field of knowledge diffusion is broad, and hence only some topics of the above mentioned fields can be deeper investigated in this thesis. Therefore, three main topics have been distilled, from which two of them are more theoretically based. These topics are motivated in this paragraph.

The second chapter of this thesis deals with the question, how knowledge transfer affects knowledge diffusion. As mentioned above, diffusion process of knowledge on a microeconomic level is often modeled as a sigmoid process of time. Although such diffusion models have a long history and mushroomed over time, not only in economics, but also in psychologies and medicine research, the seminal paper of (Bass, 1969) has made this epidemic models quiet popular. This is particularly true for applied diffusion research, as marketing or product related innovation of economics, because that type of models is easy to translate and to embed into in an economic framework. However, there are several drawbacks of this easy to implement and easy to use diffusion models, especially when employing these type of models in a knowledge diffusion context. First of all, a large number of contributions have relaxed the strict assumption of homogeneous adopters, because for knowledge diffusion processes interactions between innovators and imitators are from importance, as highlighted by (Tanny and Derzko, 1988) and (Van den Bulte and Joshi, 2007).

The model of (Tanny and Derzko, 1988), which is one of the first extension of the (Bass, 1969) diffusion model, indeed incorporates innovators and imitators behaviour. This type of model as many others, for instance (Easingwood et al., 1983), (Mahajan and Peterson, 1985) and (Mahajan and Wind, 1986), p. xiii, is not eligible to replicate bimodal adoption and patterns, although these seem relevant, as noted by (Moore, 2002). The exceptions are the recent contributions of (Goldenberg et al., 2006) and (Van den Bulte and Joshi, 2007) which are able to replicate bimodal patterns.

As laid out above, in relevant literature knowledge diffusion and knowledge transfer are often treated as synonyms. Although, knowledge diffusion and knowledge transfer are close related, in fact knowledge diffusion and knowledge transfer are two sides of a coin, as mentioned above. Hence, the implication is, that knowledge transfer, if possible can have a direct influence on knowledge diffusion via amplification and acceleration effects. As shown by (Hansen, 1999) strong network relations are necessary to transfer complex knowledge from sender to receiver. Thus, the implication is that knowledge networks and the ability of successful knowledge transfer are closely related. If we refer to the relevant literature, no recent work has examined the effects of knowledge networks on knowledge diffusion.

Before mentioned works have in common that they are discussed in a deterministic model frame. As noted by (Boswijk and Franses, 2005) and (Boswijk et al., 2006), diffusion process of a technology should be better treated as a stochastic phenomenon, because adoption is not certain. Further they assume that uncertainty of adoption is not constant over time: at the beginning and at the end of the diffusion process uncertainty regarding the adoption should tend to zero, while in the middle of the diffusion process uncertainty of adoption is high. The drawback of the (Boswijk and Franses, 2005) framework is, that it only replicates the (Bass, 1969) assumptions and neglects network effects as well as it assumes homogeneous adopters. (Boswijk et al., 2006) have extended (Boswijk and Franses, 2005) in a multivariate way, but in their model context, network effects are not of direct importance.

Thus, the aim of the second chapter is to set up a model based on (Goldenberg et al., 2006), (Van den Bulte and Joshi, 2007), (Boswijk and Franses, 2005) and (Boswijk et al., 2006) which first, assumes heterogeneous adopters, second includes knowledge networks and third is modeled in a stochastic framework. From this background the question, how knowledge networks do influence knowledge diffusion via knowledge transfer between innovators and imitators is answered. Another appealing feature of this model is, that it can be estimated directly.

It is shown that the shape of the adoption pattern depends on the fact, whether knowledge diffusion occurs or not. If knowledge transfer occurs, the stronger network effects, so called unimodal patterns are more probable, because right before innovators have realized the inflection point, imitators have nearly reached themselves their inflection point. In contrast, the longer the discrepancy between the realization of the inflection point of innovators and the beginning of imitators adoption is, the less important are network effects, the more probable are so called bimodal adoption phenomena. Thus "chasm" pattern of adoption curves occur if network effects are of less importance. Further it has been shown, that uncertainty is largest around the inflection point of the adoption curve. Finally, some econometric annotations regarding an appropriate estimation scheme are given.

The third chapter of this work is devoted to the question how learning and knowledge diffusion between heterogeneous firms, if small or large, affects market structure development. To obtain an insight into these questions, an industrial dynamics model based on the works of (Mazzucato, 1998) or (Cantner and Hanusch, 1998) has been developed. The idea of (Mazzucato, 1998) or (Cantner and Hanusch, 1998) rest upon the phenomenon of natural selection based on (Fisher, 1930). Within this approach the survival of the fittest principle of Charles Darwin can be dispatched.

The basis to answer this question is a model which is set up on the model of (Mazzucato, 1998) or (Cantner and Hanusch, 1998) with some elements borrowed from (Noailly et al., 2003). The model assumes that a given number of firms, which are heterogeneous exists in a market. Heterogeneity is modeled via cost differentials in production. In terms of the replicator dynamics approach, only those firms with low costs compared to the others will survive. Production costs can be reduced over time because of cost reduction innovation. It is further assumed that there is a negative relationship between market share and cost level. The fitness of a firm is replicated via its market share. If we combine the fitness and heterogeneity of firms, then it is possible to derive a replicator differential equation, which is responsible for the dynamic and thus the selection competition in the system. Firm size in this context is introduced via the so called *Schumpeter-Mark-I* and *Schumpeter-Mark-II* hypothesis. In literature this scenario is labeled *Schumpeter-Mark-I*. On contrary to (Schumpeter, 1912), (Schumpeter, 1942) singled out the importance of established large firms which disposed of an own R&D division, for market stability. This scenario is labeled *Schumpeter-Mark-II*. The question now arises, under which conditions market structure coincides with *Schumpeter-Mark-I* or *Schumpeter-Mark-II*. (Arthur, 1989), (David, 1985) and (Klepper, 1996) argue that dynamic increasing returns to scale create lock-in-dynamics which allow growing some firms persistently more than other firms in the same market. This success-breeds-success-scenario is closely related to the *Schumpeter-Mark-II* scenario, whereas *Schumpeter Mark I* is closely related to decreasing returns to scale. Because of its inherent complexity these models cannot be solved analytically. The key finding in a duopolistic simulation study is, that under constant and increasing returns to scale only the more efficient and larger firm will survive. The large technological progress is, which coincides with a fast rate of cost reduction, the more probable a monopolistic market structure is. In contrast under the assumption of decreasing returns to scale, and under a suitable parameter regime, it can be shown, that despite of turbulences at the beginning, a coexistence of both firms will be the result, whereas the small firm becomes market leader. Although simulation results are more or less in line with empirical observations and partly confirms the stylized fact regarding the observation of early stage market turbulences[19], as mentioned above the model itself has limitations.

(Mazzucato, 1998) and (Cantner and Hanusch, 1998) assume, that the ability of firms to understand knowledge is treated as an exogenously given and constant over time. The implication is, that learning effects are fixed which contradicts the learning curve literature as mentioned above. Hence, the aim of this chapter is, to introduce a model which is based on (Mazzucato, 1998), (Cantner and Hanusch, 1998) and (Noailly et al., 2003) which incorporates learning curve effects. Learning curve effects are first treated as deterministic but then a stochastic version of a sigmoid learning curve is introduced. It is shown that introducing learning effects has an influence on market structure. If learning curve effects exist, first endogenous learning has a positive effect on inferior strategies for low values of the technological progress. For all given scenarios of returns to scale it can be shown that even laggard firms remain in the market. This is particularly true for the increasing returns to scale case, where small firms are at a disadvantage to large firms. This observation coincides with the works of (Campagni, 1991), (Best, 2001), (Porter, 2000) and (Krugman, 1991) in a more spatial context, which highlight that inter-firm cooperation based on knowledge sharing can explain the predominance of small firms in the market. Also the new model confirms the stylized fact of early market turbulence, but it is shown that learning effects exhibit a smoothing effect on market turbulences.

The fourth chapter of this work is dedicated to the spatial dimension of knowledge diffusion. Given we know the source of knowledge creation, how can we describe concisely the way of how knowledge is transferred from sender to receiver? Is it always the case, that knowledge finds a receiver or does it depend on where the receiver is located? The question we have to ask is therefore, is it always true that knowledge creation is an unlimited process regarding space? To give an answer, we first have to think about the kind of knowledge we have in mind. For example, if knowledge is tacit than face-to-face communication or spatial proximity is a necessary condition for knowledge diffusion. On the other hand, if knowledge is codified, modern communication facilities can be used to transfer knowledge from sender to receiver. Codified knowledge is less space depended than tacit knowledge as highlighted by (Anselin et al., 1997). Therefore, we should expect that tacit knowledge dissemination is different from explicit knowledge dissemination with respect to time and space.

From this point of view, it is plausible to focus not only on time as done in the preceding chapters, when integration knowledge diffusion in a growth model context,

[19]Please refer to (Mazzucato, 2000), p. 49 for an overview.

but also to consider a possible space limitation of knowledge transfer. Empirically, we find evidence for space limitation regarding knowledge diffusion, as mentioned above.

From a macroeconomic perspective the economic role of space should be taken under consideration. The modeling of explicit space dependence is rather complex, because, if we try to integrate the space in a growth model context, we have to acknowledge heterogeneous regions, for which we have to consider both a vector of control variables and a vector of state variables for each region. Due to its inherent complexity, such a model can only be solved numerically. The aim of this chapter is, first to give a new insight into the role of spatial dependence in a theoretical semi-endogenous growth model, which core is based on the model of (Uzawa, 1965) and (Lucas, 1976). It is assumed that, spatial spillovers are local and not constant over space which implies that the tacit element in knowledge overwhelms the explicit element. With this model it can be shown as noted by (Fujita and Thisse, 2002), that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[20], which is also known as the "'folk theorem of spatial economics"'. One key result of the theoretic model is, that the disparity of income distribution is largest if the "'folk theorem of spatial economics"' matters. Further, higher order spatial influence has a positive effect on income distribution in that sense that these effects reduce inequality, because more regions benefits from knowledge spillovers.

In a further section, an econometric study has been performed to answer the question, if knowledge spillovers are more global or local end and tries to find support for the before developed spatial model. The study is based on a spatial cross section production function approach, proposed by (Griliches, 1979) which should measure the effects of innovativeness, represented by knowledge capital, such as human capital, patents or $R\&D$ and spatial spillovers on output for German NUTS-2 regions. Because spatial econometric model selection still is a highly disputed topic, a new model selection mechanism is proposed, which combines frequentest and Bayesian model selection criteria. One key result is, that the assumption of spatial heterogeneity is appropriate for explaining economic performance of German NUTS-2 regions. This last finding is additionally supported by a conducted spatial filtering procedure. Another key result is that, based on spatial weighting matrices, knowledge spillovers seems to be rather local than global.

Chapter five summarizes the findings of the thesis and gives major conclusions. Additionally, the role of knowledge diffusion for economics is stressed, particularly in the light of findings which will be discussed in section four. Comments on further research possibilities both on theoretical and empirical side are made with respect to

[20](Fujita and Thisse, 2002), p. 342.

the topics discussed in chapter two, three and four.

2 Knowledge diffusion and the role of knowledge transfer: a stochastic approach

2.1 Motivation

To know the way in which knowledge is technically produced and to understand its diffusion path is from fundamental importance in the innovative process. But what is knowledge? Knowledge itself can be embodied in new products, or can be approximated by citation of scientific publications[1], but loosely spoken there is no clear cut definition what knowledge is. What we know first is, that technological knowledge is often not transferred as itself, but within new technologies, via licensing, FDI or products for instance. Thus if we talk about knowledge diffusion, it is either a direct transfer or indirectly linked with the diffusion of new technologies, intermediate and capital goods as (Rivera-Batiz and Romer, 1991) have argued. In this model direct knowledge diffusion is assumed but with the annotation that with an empirical view, proxies of knowledge diffusion as the above mentioned are required.

Second, we know, that knowledge adoption is no homogeneous process over potential adopters. With diffusion one could associate the picture of dropping colour in a glass of water and waiting until the colour has more or less uniformly distributed over time and space within the glass. Such an imagination is of course too simple. It is, if any appropriate for the "homo economicus world", in which everybody knows everything right from the beginning or with a less strict assumption, everybody can learn everything with probability one. In such a world, the question what kind of knowledge can diffuse easily and what kind of knowledge can diffuse less easily is obviously obsolete. Assuming that the world is not perfect with respect to learning abilities and information potentials for instance, the question what kind of knowledge can easy diffuse and what kind of knowledge is diffusible is from importance. (Polany, 1967) takes this question seriously and separates implicit knowledge from explicit knowledge. The first is labelled tacit the latter not, which means, explicit knowledge can be transferred without any limits, tacit knowledge can not. For instance, assume that knowledge is

[1] Refer to (Fok and Franses, 2007) for instance.

partly or completely tacit. Then knowledge diffusion is embedded in a community and diffusion depends on the specific characteristics of that community. Thus, some people are more in touch with new developments than others. This is especially the case for two important groups of adopters, which play an important role within the diffusion process: the innovators and imitators of new knowledge.

(Bass, 1969) in his seminal work and others such as (Easingwood et al., 1983), (Mahajan and Peterson, 1985) and (Mahajan and Wind, 1986), p. xiii, for instance mentioned, that innovators and imitators behaviour is different in diffusion process. This assumption is reasonable, because it is comprehensible that each subgroup of adopters, the innovators and imitators have different intentions to adopt. Common for both groups instead is the assumption of S-shaped pattern diffusion process. Following (Kalish, 1985) one can differentiate between so called "search attributes" and "experience attributes". As a consequence, the innovators need only "search" information to adopt the new knowledge, while the latter imitators require "experience" type information before they adopt. As noted by (Gatignon and Robertson, 1985) and (Rogers, 1983) the speed of diffusion of knowledge depends on several characteristics, such as complexity, relative advantage, status value and observability etc.. These characteristics influence innovators and imitators in different ways. However most of previous studies more or less failed to highlight the different behaviour of these two specific groups.

Although (Schmalen, 1982) has mentioned that innovator's and imitator's behaviour regarding their adoption decision differs, he does not capture this facts in a notational form. The famous so called "two compartment model", proposed by (Tanny and Derzko, 1988) goes in line with the model of (Schmalen, 1982) but their definition of "innovators" and "imitators" seems not to be clear cut: "innovators" adopt because of learning effects driven by external information, whereas "imitators" adopt because of external knowledge by prior adopters. In this model it is hypothesized that "innovators" adopt because of "search" information while imitators adopt due to "experience" but also due to "knowledge transfer" which can be justified with the existence of networks.

It is therefore assumed that the adoption decision is also influenced by networks which are a necessary condition for knowledge transfer between both groups. But it is worth to mention, that "knowledge transfer" is not a sufficient condition for "knowledge diffusion". If a dense network structure is available, "knowledge transfer" is easier and thus the imitator should adopt faster. On contrary, if networks do not exist, knowledge transfer is excluded and thus adoption takes place later. The latter scenario often leads to the so called "chasm" pattern between early and late adoptions, which is extensively discussed in (Moore, 2002) and mentioned in diffusion related literature[2].

Therefore, network effects should have also an influence regarding the shape of the adoptions curve, which is in the latter case not necessarily unimodal but bimodal for the entire market. The point is, that the introduced model treats the "chasm" pattern as endogenous, not as a given exogenous number. The literature is still silent about this topic and only a few micro based paper take this network effects into account, for instance (Van den Bulte and Lilien, 2001), (Van den Bulte and Joshi, 2007), (Hill et al., 2006) and (Goldenberg et al., 2006).

The aim of this chapter can be layed out as follows: on the basis of (Goldenberg et al., 2006), (Van den Bulte and Joshi, 2007), (Boswijk and Franses, 2005) and (Boswijk et al., 2006) a knowledge diffusion model is set up, which includes "innovators" and "imitators" behaviour. Further this model is able to replicate both, unimodal and bimodal adoption pattern. Which pattern occurs, depends on the fact if network effects play a crucial role within the diffusion process. Additionally, the model will be extended towards a stochastic knowledge diffusion model to capture the idea that uncertainty of adoption is a function of time, which means at the beginning and at the end of the diffusion process uncertainty regarding the adoption should tend to zero, while in the middle of the diffusion process uncertainty of adoption is high. Another feature of the proposed model is, that it can be applied directly empirically.

The chapter is structured as follows: in the second section, I start off with an introduction and discussion of the (Bass, 1969) model. In the third section a deterministic knowledge diffusion model is setup. After the solution of this model the solution's stability is discussed. The fourth section deals with the deterministic knowledge diffusion model which is embedded into in a stochastic framework. In the fifth section a simulation study of both, the deterministic and stochastic model is conducted. Before giving some remarks in the seventh section, some econometric annotations regarding the estimation of the stochastic knowledge diffusion model are given in section six.

2.2 The Bass diffusion model

The (Bass, 1969) model, loosely spoken, describes how a new product or technique is adopted over time by interaction between potential and de facto adopters or users. Adoption stops, if the market saturation level m has been reached, that means that every potential adopter has become a de facto adopter. For each potential adopter the time of adoption is random, that means ex ante the potential adopter does not know

[2]Refer to (Van den Bulte and Joshi, 2007) for instance.

when he will adopt the product. In statistical terms, time of adoption is a random variable with a distribution Function $F(t)$ and the corresponding density $f(t)$. The (Bass, 1969) model assumes that the portion of potential adopters who adopt at time t, given they have not adopted yet can be written as a linear function of adopters or in mathematical terms:

$$\frac{f(t)}{[1 - F(t)]} = p + qF(t). \tag{2.1}$$

The left hand side of equation 2.1 can be also interpreted as the hazard rate of potential adopters. The parameter p is the probability that a potential adopter adopts at t influenced by external influence, such as word of mouth influence through the adopters. On contrary q can be interpreted as the probability that a potential adopter adopts at t for a given internal influence caused by the adopters. This covers the intrinsic motivation of potential adopters that the product or technique generates some utility.

Therefore, the diffusion process[3] of the (Bass, 1969) model can be also written as follows[4]

$$\frac{dF(t)}{dt} = f(t) = [p + qF(t)][1 - F(t)], \tag{2.2}$$

which can be interpreted as follows: on the left hand side of equation we can find the rate of change with respect to time t of the cumulative number of adopters. This is equal to the hazard rate $p + qF(t)$ times the adopters which have not adopted in t. Thus, $[1 - F(t)]$ are the potential adopters. If $p = 0$ we obtain a diffusion process which is completely driven by internal influence of adopters in t, whereas $q = 0$ the diffusion process depends solely on external influence. In general, a mixture influence model is assumed, that means that $\{p, q\} \in (0, 1)$.

Labeling the cumulative number of adopters at t as $N(t) = mF(t)$, the rate of change of adopters is given by

[3]A mathematical diffusion function ca be expressed as the solution $y = y(t)$ of a deterministic differential equation $\frac{dy}{dt} = f(y, t)$. $f(\cdot)$ describes the pattern of the diffusion path and y gives information about the evolution of the diffusion process over time. Thus $f(\cdot)$ is a dependent function of y and diffusion time t. This is the basic idea of modelling diffusion path.

[4]Refer to (Kalish and Sen, 1986) and (Mahajan et al., 1984) for instance.

$$n_t \equiv \frac{dN(t)}{dt} = m\frac{dF(t)}{dt} = mf(t), \tag{2.3}$$

or inserting 2.2 in 2.3 and noting that $N(t) = mF(t)$ yields

$$n_t \equiv \frac{dN(t)}{dt} = m\left\{\left[p + q\frac{N(t)}{m}\right]\left[1 - \frac{N(t)}{m}\right]\right\}, \tag{2.4}$$

or

$$n_t \equiv \frac{dN(t)}{dt} = \left[p + q\frac{N(t)}{m}\right][m - N(t)] = \chi(t)[m - N(t)], \tag{2.5}$$

with $N(t) = \int_{t_0}^{t} n_t dt$. The last derived so called *Ricatti*-differential equation with constant coefficients can be interpreted as the rate of change with respect to time t of the cumulative number of adopters which is equal to a time dependent variable $\chi(t)$, which covers the mixture influence of adoption, given $\{p, q\} \in (0, 1)$ times the cumulative number of potential adopters in t given by $[m - N(t)]$. From equation 2.5 we can easily see, that change rate of cumulative adopters is zero, given the number of potential adopters equals the number of cumulative adopters which is equal to the postulation that $[m - N(t)] = 0$.

The solution of 2.5 for the cumulative number of adopters is given by:

$$N_t = mF(t) = m\left[\frac{1 - exp\{-(p+q)t\}}{1 + \frac{q}{p}exp\{-(p+q)t\}}\right], \tag{2.6}$$

and for the adoption in t:

$$n_t = mf(t) = m\left[\frac{p(p+q)^2 exp\{-(p+q)t\}}{(p + qexp\{-(p+q)t\})^2}\right]. \tag{2.7}$$

The problem which now occurs is, how to translate this theoretical model in practical application. The number of adopters are usually discrete values, whereas the above derived diffusion equation 2.5 is written in continuous time. For this reason, (Bass, 1969) applied a simple *Euler*-discretization scheme to obtain the following discrete time difference equation of the continuous time differential equation 2.5:

$$N_t = N_{t-1} + \left[p + q\frac{N(t-1)}{m}\right][m - N(t-1)]\,. \tag{2.8}$$

Due to its parsimonious specification, the (Bass, 1969) diffusion model and its extensions are so popular in diffusion research[5]. Besides, it should be mentioned that from equation 2.8 it is quite clear that the (Bass, 1969) model is very attractive also for empirical application, especially for out-of-sample forecasts[6], because theoretically equation 2.8 can be estimated without modifications[7]. Although, the (Bass, 1969) model seems to be very intuitive and well established both in theoretical and empirical application, there are several drawbacks.

(Bass, 1969) mentioned in his publication, that innovators and imitators behaviour is different in diffusion process. This assumption is reasonable, because it is reasonable that each subgroup of adopters, the innovators and imitators have different intentions to adopt. Following (Kalish, 1985) one can differentiate between so called "search attributes" and "experience attributes". As a consequence of that the innovators need only "search" information to adopt knowledge, while the latter imitators require "experience" type information before they adopt. As noted by (Rogers, 1983) for instance, the speed of diffusion of knowledge depends on several characteristics, such as complexity, relative advantage, status value and observability etc.. These characteristics influence innovators and imitators in different ways.

Additionally, communication between those two types and thus network effects are a second channel which influence imitators adoption decision.[8] But within his mathematical representation layed out above, for instance in 2.8 this fundamental assumption is not reflected, although it is from central importance for knowledge diffusion.[9]

Another limitation of the (Bass, 1969) model is, that it also can reproduce a bell shaped single peak adoption curve. (Kluyver, 1977) has pointed out, that "one drawback of such models (diffusion type models) is that only unimodal phenomena can be fitted". If one refers to the empirical literature there is strong evidence that life cycle of innovations fits to a more bimodal pattern[10]. This is due to the fact that in the early

[5]Refer to (Parker, 1994), (Mahajan et al., 1990), (Mahajan et al., 1993), (Sultan et al., 1990) and (Mahajan et al., 2000) for an overview.

[6]For instance, refer to (Bass, 1993) and (Bass, 1995) for this topic.

[7]It should be mentioned that there is a large bulk of paper which discuss estimation strategies for the Bass model. Refer to (Boswijk and Franses, 2005) for a discussion of that topic.

[8]Refer to (Gladwell, 2000), (Moore, 1995), (Rosen, 2000) and (Slywotzky and Shaprio, 1993).

[9]Already (Jeuland, 1981) has pointed out this fact.

[10]Refer to (Rink and Swan, 1979) and (Tellis and Crawford, 1981).

stage of an innovation life cycle, a new product in which new knowledge is embodied, innovators demand leads to an often sharp rise, then to a plateau or a fall in adoption, followed by an imitators caused raise of adoptions, when network effects are in place.

A further interesting aspect, which is still not incorporated in innovation diffusion models is the phenomenon of "knowledge transfer". To be precise, one has to distinguish between pure "knowledge transfer", which is for instance practiced via face-to-face communication and "knowledge diffusion". "Knowledge transfer" must not necessarily influence the adoption decision but it can. In this context it is meant that "knowledge transfer" is only possible if the knowledge is transferable, for instance via face-to-face communication.

For this reason, in the next section a more general (Bass, 1969) type model is set up which first includes a heterogeneous potential adopter group which is split in innovators and imitators. Second, the new model formulation also includes network effects, which are not symmetric: it is assumed that only imitators can benefit from information about the adoption of knowledge from the innovators. This means, the often mentioned effects of knowledge transfer via network effects and its effect on knowledge diffusion, embodied by the adoption of a new knowledge are incorporated. In this manner, it is possible both to replicate unimodal as well as bimodal shapes of the adoption curves. The shape of the curve only depends on the fact if knowledge transfer is easy or totally excluded. The easier knowledge transfer is, the faster knowledge diffusion should be, the lesser the probability that bimodal adoption pattern occurs or so called "chasm" between early and the later parts of the adoption curve[11].

2.3 Deterministic knowledge diffusion model

As mentioned before, innovators and imitators behaviour should be acknowledged when talking about knowledge diffusion, because heterogeneous adopters could explain bimodal adoption shapes.

2.3.1 Setup

The group of adopters $N(t)$ is separated in innovators and imitators $N(t)_k$ for $k = \{1, 2\}$.[12] $k = 1$ represents the subgroup of innovators, whereas $k = 2$ symbolizes the group of imitators. Now, the key idea is, to incorporate a communication channel

[11]Refer to (Van den Bulte and Joshi, 2007).
[12]In the following time index t is only used if clarity demands it.

between these two groups. In this way, an asymmetric communication flow is created, because per definition the subgroup of innovators could not learn anything about the subgroup of imitators regarding their adoption decision. Innovators, per definition are the first, entering the market. Thus the model contains a communication parameter q_{12} which stands for the communication between the group of innovators and the group of imitators. The diffusion process for innovators $N(t)_1$ is similar to the above mentioned Bass diffusion equation 2.5 and can be written as:

$$\frac{dN_1}{dt} = \left[p_1 + \left(q_1 \frac{N_1}{m_1}\right)\right][m_1 - N_1]\,. \tag{2.9}$$

The diffusion process for the imitators $N_{t,2}$ instead should be written as:

$$\frac{dN_2}{dt} = \left[p_2 + \left(q_2 \frac{N_2}{(m_1 + m_2)}\right) + \left(q_{12} \frac{N_1}{(m_1 + m_2)}\right)\right][m_2 - N_2]\,, \tag{2.10}$$

with q_{12} representing the "knowledge transfer" probability. Therefore, the change rate of cumulative group of imitators $\frac{dN_2}{dt}$, is also affected by network effects. If q_{12}=0 then innovators and imitators adoption are independent from each other, but still not symmetric, because even if $q_{12} = 0$ the entire market saturation level $m_1 + m_2$ is from importance for the imitators.

These two model segments 2.9 and 2.10 can be stacked into a system of equations as follows:

$$\begin{bmatrix} \frac{dN_1}{dt} \\ \frac{dN_2}{dt} \end{bmatrix} = \begin{bmatrix} \left[p_1 + \left(q_1 \frac{N_1}{m_1}\right)\right] & 0 \\ 0 & p_2 + \left[\left(q_2 \frac{N_2}{(m_1+m_2)}\right)\right] \end{bmatrix} \begin{bmatrix} [m_1 - N_1] \\ [m_2 - N_2] \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 & 0 \\ 0 & \left(q_{12} \frac{N_1}{(m_1+m_2)}\right) \end{bmatrix} \times \begin{bmatrix} [m_1 - N_1] \\ [m_2 - N_2] \end{bmatrix}\,, \tag{2.11}$$

or in compact form

$$\dot{\mathbf{N}} = \mathbf{\Xi} a + \mathbf{\Pi} b. \tag{2.12}$$

From 2.11 we can see that information flow is asymmetric, because the first diagonal element of $\mathbf{\Pi}$ in 2.12 is zero.

2.3.2 Solution

Given $N(0)_1 = 0$, the solution of our differential equation system for $N(t)_1$ can be written as before as:

$$N(t)_1 = mF(t) = m\left[\frac{1 - exp\{-(p_1+q_1)t\}}{1+\frac{q_1}{p_1}exp\{-(p_1+q_1)t\}}\right]. \tag{2.13}$$

In contrast to the solution of $N(t)_1$, the derivation of solution for $N(t)_2$ is cumbersome. Given $N(0)_2 = 0$ it can be expressed as:

$$N(t)_2 = \frac{1}{\Theta}\Big\{[m_1(p_2+q_{12}) + m_2(p_2+q_2)]\times$$

$$\Bigg[-(exp(-(p_1+q_1)t)^{\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)}}(m_1+m_2)(p_1+q_1)+$$

$$+m_2(p_1+exp(-(p_1+q_1)t)^{\frac{m_1q_{12}}{(m_1+m_2)q_1}}q_2C\Bigg]+$$

$$+(exp(-(p_1+q_1)t)^{\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)}}m_2(m_1+m_2)(p_1+q_1)\left(1+\frac{exp(-(p_1+q_1)t)q_1}{p_1}\right)^{\frac{m_1q_{12}}{(m_1+m_2)q_1}}\times$$

$$q_2\tilde{H}\left[\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)}, \frac{m_1q_{12}}{(m_1+m_2)q_1}, \Phi, -\frac{exp(-(p_1+q_1)t)q_1}{p_1}]\right]\Bigg\} \tag{2.14}$$

with Θ defined as:

$$\Theta \equiv q_2\Bigg\{(p_1+exp(-(p_1+q_1)t)q_1)^{\frac{m_1q_{12}}{(m_1+m_2)q_1}}(m_1(p_2+q_{12})+m_2(p_2+q_2))C+$$

$$+(exp(-(p_1+q_1)t)^{\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)}}(m_1+m_2)(p_1+q_1)\left(1+\frac{exp(-(p_1+q_1)t)q_1}{p_1}\right)^{\frac{m_1 q_{12}}{(m_1+m_2)q_1}}\times$$

$$\times\tilde{H}\left[\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)},\frac{m_1 q_{12}}{(m_1+m_2)q_1},\Phi,-\frac{exp(-(p_1+q_1)t)q_1}{p_1}\right]\Bigg\}, \quad (2.15)$$

with

$\Phi \equiv 1+\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)}$,

and

$$C\equiv\frac{1}{m_2 q_2[m_1(p_2+q_{12})+m_2(p_2+q_2)]}\Bigg[(m_1+m_2)(p_1+q_1)^{1-\frac{m_1 q_{12}}{(m_1+m_2)q_1}}\times$$

$$\times(m_1(p_2+q_{12})+m_2(p_2+q_2)-m_2\left(\frac{p_1+q_1}{p_1}\right)q_2\times$$

$$\times\tilde{H}\left[\frac{m_1(p_2+q_{12})+m_2(p_2+q_2)}{(m_1+m_2)(p_1+q_1)},\frac{m_1 q_{12}}{(m_1+m_2)q_1},\Phi,-\frac{exp(-(p_1+q_1)t)q_1}{p_1}\right]\Bigg]. \quad (2.16)$$

Note, that $\tilde{H}(\cdot)$ is the hypergeometric function, which series expansion is given by

$$\tilde{H}\equiv_2 F_1(a,b,c,x)=\sum_{w=0}^{\infty}\frac{(a)_w(b)_w}{(c)_w}\frac{x^w}{w!}=$$

$$=1+\frac{abx}{c1!}+\frac{a(a+1)b(b+1)x^2}{c(c+1)2!}+\frac{a(a+1)(a+2)b(b+1)(b+2)x^3}{c(c+1)(c+1)3!}+..., \quad (2.17)$$

where $(i)_w$ is the Pochhammer symbol which is defined as $(i)_0=1$ and $(i)_w=i(i+1)...(i+w-1)=\frac{\Gamma(i+w)}{\Gamma(i)}$ for $i=a,b,c$ where $\Gamma(\cdot)$ is called the *Euler-Gamma* function.[13] Further note that $\tilde{H}$ has a branch cut discontinuity in the complex z plane from 1 to

[13](Abramowitz and Stegun, 1972) p. 255.

∞ and terminates if a and b are non positive integers. Of course, $n(t)_k = \frac{dN(t)_k}{dt}$ for $k = \{1, 2\}$.

2.3.3 Stability

Before we proceed the equilibrium points of model 2.11 or 2.12 are identified and additionally, the stability of them is proofed.

Proposition 1: On behalf of the assumption that the partial derivatives of $\frac{dN_1}{dt}$, and $\frac{dN_2}{dt}$ exist and that $\frac{dN_1}{dt}$ and $\frac{dN_2}{dt}$ hold simultaneously $\forall t$, the system 2.11 has a unique steady state vector $\mathbf{S}$, which contains N_1^* and N_2^* in the long run.□

Proof 1: An optimal steady state vector exists, if and only if $\frac{dN_1}{dt} = \frac{dN_2}{dt} = 0$ holds. This is realized, if

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \left[p_1 + \left(q_1 \frac{N_1}{m_1}\right)\right] & 0 \\ 0 & p_2 + \left[\left(q_2 \frac{N_2}{(m_1+m_2)}\right) + \left(q_{12} \frac{N_1}{(m_1+m_2)}\right)\right] \end{bmatrix} \begin{bmatrix} [m_1 - N_1] \\ [m_2 - N_2] \end{bmatrix} . \tag{2.18}$$

To find the elements for the steady state vector the first equation from the derived system 2.12 has been examined first. Given $\frac{dN_1}{dt} = 0$, this equation can be written as follows:

$$\left[p_1 + \left(q_1 \frac{N_1}{m_1}\right)\right][m_1 - N_1] = 0. \tag{2.19}$$

An equilibrium is found if $\frac{dN_1}{dt} = 0$ holds. Thus, if $m_1 = N_1^*$, then $\frac{dN_1}{dt} = 0$. If $m_1 = N_1^*$ then the number of innovators of new knowledge have realized their market saturation level m_1, which implies that every potential innovator has adopted new knowledge.

Second, if

$$N_1^* = \frac{-m_1 p_1}{q_1} < 0, \tag{2.20}$$

equation 2.19 is zero again and thus $\frac{dN_1}{dt} = 0$ also holds. Note that this equilibrium can be ruled out because $N_1 > 0$ per definition.

Now let us turn to the second equation of system 2.12, which can be written as

$$(m_2 - N_2)\left[p_2 + \left(q_2 \frac{N_2}{(m_1 + m_2)}\right) + \left(q_{12} \frac{N_1}{(m_1 + m_2)}\right)\right] = 0, \tag{2.21}$$

given $\frac{dN_2}{dt} = 0$.

Again, if $m_2 = N_2^*$ then $\frac{dN_2}{dt} = 0$, which implies again, that the number of imitators reached their market saturation level m_2. Additionally, if

$$N_2^* = -\frac{1}{q_2}\left[p_2 m + q_{12} N_1^*\right] < 0, \tag{2.22}$$

then $\frac{dN_2}{dt} = 0$ holds. This equilibrium can be ruled out ex ante because $N_2 > 0$ per definition.

From this discussion it is possible to derive four long run equilibria: the first equilibrium is given by

$$m_1 = N_1^* \text{ and } m_2 = N_2^*. \tag{2.23}$$

This is the case, when both, the innovators and imitators have reached their specific market saturation levels.

The second equilibrium is obtained if

$$N_1^* = \frac{-m_1 p_1}{q_1} \text{ and } \ m_2 = N_2^*. \tag{2.24}$$

The third equilibrium is characterized by

$$N_1^* = \frac{-m_1 p_1}{q_1} \text{ and } \ N_2 = -\frac{1}{q_2}\left[p_2 m + q_{12} N_1^*\right]. \tag{2.25}$$

Noting the fact, that $N_1^* = \frac{-m_1 p_1}{q_1}$ and inserting this expression in $N_2 = -\frac{1}{q_2}\left[p_2 m + q_{12} N_1^*\right]$ yields $N_2^* = \frac{m_1(q_{12}p_1 - p_2 q_1) - p_2 m_2 q_1}{q_2 q_1}$.

Obviously, the sign of N_2^* is not clearly determined. For a given value of N_1^*, N_2^* can be positive or negative. The last equilibrium is defined by

$$N_1^* = m_1 \text{ and } \ N_2^* = -\frac{1}{q_2}\left[m_1(p_2 + q_{12}) + m_2 p_2\right]. \blacksquare \tag{2.26}$$

Next the system 2.12 is linearized around the steady state values to establish the stability of obtained equilibria. After linearizing the entire system the Jacobian matrix for each equilibrium of our system 2.12 has been evaluated. The following table 2.1 provides a summary of the obtained equilibria. Further the equilibrium specific Eigenvalues with their corresponding signs are reported.

It is obvious, that the first equilibrium is a stable node. The stability of the remaining equilibria is not from importance, because from an economical point of view only the

Equilibrium	*Equilibrium conditions*	*Signs of Eigenvalues*
E_1	$N_1^* = m_1$ and $N_2^* = m_2$	$\lambda_1 < 0, \lambda_2 < 0$
E_2	$N_1^* = \frac{-m_1 p_1}{q_1}$ and $N_2^* = m_2$	$\lambda_1 > 0, \lambda_2 <> 0$
E_3	$N_1^* = \frac{-m_1 p_1}{q_1}$ and $N_2^* = \frac{m_1(q_{12}p_1 - p_2 q_1) - p_2 m_2 q_1}{q_2 q_1}$	$\lambda_1 > 0, \lambda_2 <> 0$
E_4	$N_1^* = m_1$ and $N_2^* = -\frac{1}{q_2}\left[m_1(p_2 + q_{12}) + m_2 p_2\right]$	$\lambda_1 < 0, \lambda_2 > 0$

Table 2.1: Stability analysis of obtained equilibria from system 2.11 (I)

Equilibrium	*Imaginary part*	*Stability*
E_1	no	stable node
E_2	no	saddle path or unstable node
E_3	no	saddle path or unstable node
E_4	no	saddle path

Table 2.2: Stability analysis of obtained equilibria from system 2.11 (II)

first equilibrium ensures a plausible result, which means that both N_1^* and N_2^* are positive, given the parameter definition above. These result can be fleshed out also graphically in figure 2.1 for positive values for $N_1^* = N_2^* > 0$.

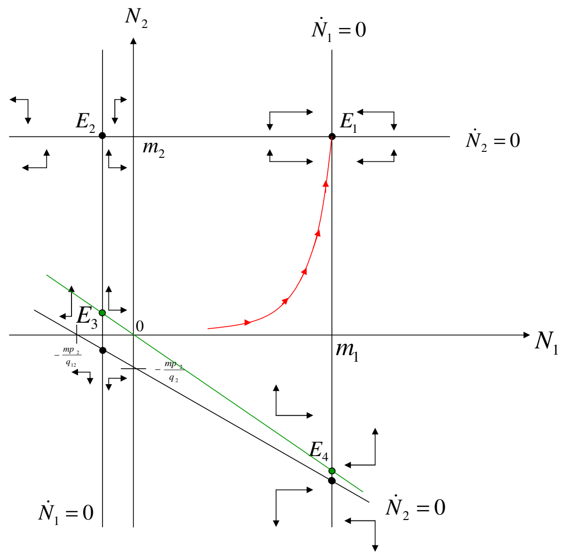

Figure 2.1: Phase plot of model 2.11 (I)

From figure 2.1 we can conclude that only for the first equilibrium an economic interpretation is possible. In the long run, the market saturation level will be reached for both groups of adopters. Moreover, this equilibrium is stable. The third equilibrium is ex ante not clearly determined, because for a given parameter constellation, positive as well as negative values for N_2 are possible. Theoretically, the $\dot{N}_2 = 0$ straight line, which slope and location is determined by E_3 and E_4, can result in another feasible solution. But if we take a closer look at our model we can rule out this possibility: If we refer again to figure three we can see, that $-\frac{p_2 m}{q_2}$ and $-\frac{p_2 m_2}{q_{12}}$ determine the location of $\dot{N}_2 = 0$ straight line $N_2 = -\frac{q_{12}}{q_2} N_1 - \frac{p_2 m}{q_2}$.

Of course, the maximum limit expression of $\dot{N}_2 = 0$ straight line is given by: $N_2 = \frac{q_{12}}{q_2} N_1$ which is graphically replicated by a green straight line through the origin, as can be seen in figure 2.1. This is the maximum limit because, $-\frac{p_2 m}{q_2}$ cannot be positive by definition, as all parameter in expression $\frac{p_2 m}{q_2}$ are positive. This is also true for $-\frac{p_2 m_2}{q_{12}}$. Note also, that the $\dot{N}_2 = 0$ straight line will not be translated parallel, because the upper limit for N_2 is given by $\tilde{N}_2 = \frac{q_{12} m_1 p_1}{q_2 q_1}$ and hence the difference between the upper limit of $\tilde{N}_2$ and N_2 is given by $\Delta N_2 = \frac{q_1 p_2 m_1}{q_2 q_1} + \frac{p_2 m_2}{q_2} = \frac{q_1 p_2 m_1}{q_2 q_1} + \Delta^+ N_2$ which is by expression on modulus greater as $\Delta^+ N_2 = \frac{p_2 m_2}{q_2}$ if we refer to equilibrium four.

From this discussion it can be concluded that E_3 cannot be a possible candidate for a relevant economic equilibrium. Again, from an economic point of view we only focus on the first equilibrium which is given by: $N_1^* = m_1$ and $N_2^* = m_2$. Thus from any given and feasible starting point within the rectangular area bounded by the parallel $\dot{N}_2 = 0$ line to the hypotenuse and the parallel $\dot{N}_1 = 0$ line to the ordinate we can always realize the equilibrium point E_1 for given starting values $N(0)_1 \geq 0$ and $N(0)_2 \geq 0$. Referring again to figure 2.1 the red line symbolizes the steady state path for given but arbitrary starting values $N(0)_1 > 0$ and $N(0)_2 > 0$.

The so far derived model 2.11 has to be criticized as it assumes a short and long run deterministic behaviour of the adoption process, which means that being on the S-shaped diffusion path, no deviations from this path are possible, even in the short run. The implication is, that uncertainty regarding the adoption process should not be treated as constant over time, as the (Bass, 1969) model does. Especially, in the middle of the diffusion process, say around the inflection point, uncertainty should be much more higher than at the beginning or at the end, which implies that fluctuations of the adoption curve should be largest around the inflection point. From this point of view, a stochastic expansion of 2.11 is required which will be derived in the next section.

2.4 Stochastic knowledge diffusion model

2.4.1 Setup

In this section a stochastic expansion of system 2.11 will be derived. I follow (Boswijk and Franses, 2005) and (Boswijk et al., 2006) who derived a stochastic "counterpart" of the (Bass, 1969) model by assuming short-run deviations from the deterministic diffusion process. To arrive at our stochastic "counterpart" of 2.11 it has to point out first, that the cumulative number of innovators and imitators are both random variables with

$$\bar{N}(t)_k = E[N(t)_k] = mF(t),\ k = \{1,2\}, \tag{2.27}$$

where $k = 1$ stands for the innovators and $k = 2$ for imitators and t is measured still in continuous time.

Defining $\frac{dN(t)_k}{dt} = n(t)$ for $k = \{1,2\}$ we can theoretically derive two different systems: the first system could assume that mean reverting takes place from the mean number $\bar{n}(t)$ or from the actual alteration rate of adoptions $\tilde{n}(t)$. The difference is, that the mean alteration rate of adoptions $\bar{n}(t)$ is treated as an exogenous variable, whereas $\tilde{n}(t)$ is endogenous. For this reason, we should prefer to work with $\tilde{n}(t)$.

Keeping this in mind the following stochastic expansion of system 2.11 is defined:

$$\begin{bmatrix} dn(t)_1 \\ dn(t)_2 \end{bmatrix} = \begin{bmatrix} \zeta[\tilde{n}(t)_1 - n(t)_1]\frac{dt}{dW(t)_1} + \sigma n(t)_1^{\gamma} & 0 \\ 0 & \zeta[\tilde{n}(t)_2 - n(t)_2]\frac{dt}{dW(t)_2} + \sigma n(t)_2^{\gamma} \end{bmatrix} \begin{bmatrix} dW(t)_1 \\ dW(t)_2 \end{bmatrix}, \tag{2.28}$$

where $W(t)_k$ is the standard Wiener process, $\zeta > 0$ is the adjustment speed. Please note, that $W(t)_1$ and $W(t)_2$ are eventually correlated. Further $\sigma > 0$ and $\gamma \geq 0.5$. Therefore, the speed of mean reversion depends on the value of ζ. This system 2.28 is a generalized stochastic version of 2.11, because it contains an error term in continuous time with a standard deviation which equals to $\sigma n(t)_k^{\gamma}$. As $n(t)_k \to 0$, the error term $\sigma n(t)_k^{\gamma} \to 0$ and thus it is guaranteed that $n(t)$ takes non negative values. It should be clear that 2.11 is obtained if $\zeta \to \infty$ and $\sigma \to 0$. Because of the fact, that for $\gamma = 1$ it can be shown that $n(t)_k$ is strictly positive. Hence, $\gamma = 1$ has been set. In this work, the examination of system´s 2.28 dynamic behaviour is done on the basis of simulation experiments. Alternatively, one can show formally, the existence and solution of 2.28. One aspect which can be easily seen from 2.28 is that, given the starting value $N(0)_k = 0$, $N(t)_k$ increases monotonically to $N(t)_k = m$ for $t \rightarrow$ for large T. Please additionally note, that the speed of adjustment ζ is assumed to be the same for both the innovators and imitators. That is also the case for σ.

Inserting model 2.11 in 2.28 yields the following system of stochastic differential equations (sde):

$$\begin{bmatrix} dn(t)_1 \\ dn(t)_2 \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} dW(t)_1 \\ dW(t)_2 \end{bmatrix}, \tag{2.29}$$

with

$$A \equiv \zeta \left\{ \Theta_1 - n(t)_1 \right\} \frac{dt}{dW(t)_1} + \sigma n(t)_1, \tag{2.30}$$

and

$$B \equiv \zeta \left\{ \Theta_2 - n(t)_2 \right\} \frac{dt}{dW(t)_2} + \sigma n(t)_2, \tag{2.31}$$

and

$$\Theta_1 \equiv \left[p_1 + \left(q_1 \frac{N(t)_1}{m_1} \right) \right], \tag{2.32}$$

and

$$\Theta_2 \equiv p_2 + \left[\left(q_2 \frac{N(t)_2}{(m_1 + m_2)} \right) + \left(q_{12} \frac{N(t)_1}{(m_1 + m_2)} \right) \right]. \tag{2.33}$$

To simulate 2.28, the continuous time model has to be transformed into a time discrete model with discrete observations $N_{i,k} = N(t_i)_k$ for $i = 1, 2, ..., T$ and $k = \{1, 2\}$. Thus, adoption of new knowledge over the interval $(t_{i-1,k}, t_{i,k}]$ is given by $\Psi \equiv N_{i,k} - N_{i-1,k}$.

2.4.2 Euler-Maruyama approximation

The discretization model 2.29 is based on the so called *Euler-Maruyama* approximation[14]. On a given interval $[t_0, T]$ and for a given discretization $t_0 < t_1 < ... < t_i < ... < t_N = T$ of $[t_0, T]$ an *Euler-Maruyama* approximation of an one dimensional Ito sde $dX_t = f(X_t, \theta) + g(X_t, \theta)dW_t$ is a so called time stochastic process which satisfies the proposed iterative scheme

$$y_{i+1} = y_i + h_i f(y_i) + g(y_n)\Delta W_i, \tag{2.34}$$

with $y_0 = x_0$ for $i = 0, 1, ..., N-1$, where $y_i = y(t_i)$, $\xi_i = [t_i - t_{i-1}]$ is the step size and $\Delta W_i = W(t_i) - W(t_{i-1}) \sim \mathcal{N}(0, \xi_i)$ with $W(t_0) = 0$.

The last follows, because of the definition of a *Wiener* process we conclude that these increments are independent *Gaussian* random variables with mean 0 and variance h_i. The increments ΔW_n can be computed as $\Delta W = \int_{t_i}^{t_{i+1}} dW_t = W(t_{i+1}) - W(t_i)$. It is straightforward that the proposed *Euler-Maruyama* approximation still holds for systems, like 2.29. Please note again, that $W(t)_1$ and $W(t)_2$ are eventually correlated.

It is known that *Euler-Maruyama* method converges with strong order $\gamma = 1$ for additive noise and for constant diffusion term g the *Euler-Maruyama* method should provide a reasonable approximation.[15] For other cases, however the method provides eventually a poor estimate of the solution, especially if the coefficients of interest have to be treated as non-linear, a fact, which is known from the deterministic *Euler*-approximation. To get a higher accuracy of approximation higher order schemes, like the *Milstein* scheme, should be consulted, because it has to be pointed out that as the order of *Euler-Maruyama* is only satisfactory regarding approximation results if a fine time span $\xi_i = \frac{H}{T}$ is used.[16]

Applying the *Euler-Maruyama* approximation for system 2.29, using $n(t_i)_k - n(t_{i-1})_k$, the following expression is obtained:

$$\begin{bmatrix} n(t_i)_1 - n(t_{i-1})_1 \\ n(t_i)_2 - n(t_{i-1})_2 \end{bmatrix} \approx \begin{bmatrix} \zeta \left\{ \left[p_1 + \left(q_1 \frac{N(t_{i-1})_1}{m_1} \right) \right] - n(t_{i-1})_1 \right\} \xi + \vartheta_1 \\ \zeta \left\{ p_2 + \left[\left(q_2 \frac{N(t_{i-1})_2}{(m_1+m_2)} \right) + \left(q_{12} \frac{N(t_{i-1})_1}{(m_1+m_2)} \right) \right] - n(t_{i-1})_2 \right\} \xi + \vartheta_2 \end{bmatrix}, \tag{2.35}$$

[14]Refer to for (Kloeden and Platen, 1992), p. 305 instance.

[15]On general, the *Euler-Maruyama* method has strong order of convergence $\gamma = 0.5$ and for weak order of convergence $\gamma = 1$.

[16]Refer to (Kloeden and Platen, 1992), p. 345.

with $\xi = [t_i - t_{i-1}]$ and $\vartheta_k = \sigma[W(t_i)_k - W(t_{i-1})_k] \sim$ i.i.d. $\mathcal{N}(0, \sigma^2\xi)$.

The approximation of adopting new knowledge $\Psi_{i,k}$ over the time interval $(t_{i-1}, t_i]$ can be written as

$$\Psi_{i,k} = N(t_i)_k - N(t_{i-1})_k = \int_{t_{i-1}}^{t_i} n(t)_k dt \approx n(t_i)_k (t_i - t_{i-1}) = n(t_i)_k \xi. \tag{2.36}$$

Thus the alteration rate of adopting new knowledge is given by $\Delta\Psi_{i,k} \equiv \Psi_{i,k} - \Psi_{i-1,k}$, or

$$\Delta\Psi_{i,k} \equiv \Psi_{i,k} - \Psi_{i-1,k} \approx \xi[n(t_i)_k - n(t_{i-1})_k]. \tag{2.37}$$

Using 2.37 together with 2.28 and 2.12 or 2.29 we obtain

$$\begin{bmatrix} \Delta\Psi_{i,1} \\ \Delta\Psi_{i,2} \end{bmatrix} \approx \begin{bmatrix} \xi\zeta\left\{\left[p_1 + \left(q_1\frac{N(t_{i-1})_1}{m_1}\right)\right] - \frac{\Psi_{i-1,1}}{\xi}\right\}\xi + \xi\frac{\Psi_{i-1,1}}{\xi}\vartheta_{i,1} \\ \xi\zeta\left\{\left[p_2 + \left(q_2\frac{N(t_{i-1})_2}{(m_1+m_2)}\right) + \left(q_{12}\frac{N(t_{i-1})_1}{(m_1+m_2)}\right)\right] - \frac{\Psi_{i-1,2}}{\xi}\right\}\xi + \xi\frac{\Psi_{i-1,2}}{\xi}\vartheta_{i,2} \end{bmatrix}, \tag{2.38}$$

or

$$\begin{bmatrix} \Delta\Psi_{i,1} \\ \Delta\Psi_{i,2} \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} \zeta\xi^2 p_1(m_1 - N_{(i-1),1}) + \xi^2\frac{q_1}{m_1}N_{(i-1),1}(m_1 - N_{(i-1),1}) - \zeta\xi\Psi_{i-1,1} + \Psi_{i-1,1}\vartheta_{i,1} \\ \zeta\xi^2 p_2(m_2 - N_{(i-1),2}) + \xi^2\frac{q_2}{(m_1+m_2)}N_{(i-1),2}(m_2 - N_{(i-1),2}) + \varsigma - \zeta\xi\Psi_{i-1,2} + \Psi_{i-1,2}\vartheta_{i,2} \end{bmatrix}, \tag{2.39}$$

with

$$\varsigma \equiv \xi^2\zeta\frac{q_{12}}{(m_1 + m_2)}N_{(i-1),2}(m_2 - N_{(i-1),2}), \tag{2.40}$$

or

$$\begin{bmatrix} \Delta\Psi_{i,1} \\ \Delta\Psi_{i,2} \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} \phi_{0,1} + \phi_{1,1}N_{(i-1),1} + \phi_{2,1}N^2_{(i-1),1} + \phi_{3,1}\Psi_{i-1,1} + \Psi_{i,1}\vartheta_{i,1} \\ \phi_{0,2} + \phi_{1,2}N_{(i-1),2} + \phi_{2,2}N^2_{(i-1),2} + \phi_{3,2}N_{(i-1),1} + \phi_{4,2}N_{(i-1),1}N_{(i-1),2} + \phi_{5,2}\Psi_{i-1,2} + \Psi_{i,2}\vartheta_{i,2} \end{bmatrix}, \quad (2.41)$$

with $\vartheta \sim$ i.i.d. $\mathcal{N}(0, \sigma^2\xi)$ and

$$\phi_{0,1} = p_1 m_1 \zeta\xi^2, \quad (2.42)$$
$$\phi_{1,1} = \zeta\xi^2(q_1 - p_1), \quad (2.43)$$
$$\phi_{2,1} = \frac{-q_1\zeta\xi^2}{m_1}, \quad (2.44)$$
$$\phi_{3,1} = -\zeta\xi, \quad (2.45)$$
$$\phi_{0,2} = p_2 m_2 \zeta\xi^2, \quad (2.46)$$
$$\phi_{1,2} = \zeta\xi^2[(\frac{m_2}{m_1 + m_2}q_1) - p_2], \quad (2.47)$$
$$\phi_{2,2} = \frac{-q_2\zeta\xi^2}{m_1 + m_2}, \quad (2.48)$$
$$\phi_{3,2} = \frac{m_2}{m_1 + m_2}\zeta\xi^2 q_{12}, \quad (2.49)$$
$$\phi_{4,2} = \frac{-q_{12}\zeta\xi^2}{m_1 + m_2}, \quad (2.50)$$
$$\phi_{5,1} = -\zeta\xi. \quad (2.51)$$

In this notational form we can interpret ς as the knowledge transfer parameter function, which depends among other values on q_{12}, which is again the probability of knowledge transfer. If $q_{12} = 0$ then $\varsigma = 0$ and thus no knowledge transfer from the innovators to the imitators takes place.

2.5 Simulation

2.5.1 Simulation of deterministic knowledge diffusion model

In this section the adoption curves of our model 2.11 are simulated. For simulation purposes we first have to assign a set of parameters. The values of the external knowledge transfer coefficients p_1 and p_2 are set to $p_1 = 0.13$ and $p_2 = 0.01$, which means that the innovators are more influenced by external knowledge transfer as the imitators. The value for the internal knowledge transfer coefficient q_1 and q_2 are determined to $q_1 = 0.75$ and $q_2 = 0.50$, which means that internal knowledge transfer matters more for the group of imitators. The knowledge transfer coefficient q_{12} is for now set to

$q = 0.07$. Later on in this chapter a sensitivity analysis regarding q_{12} for the stochastic knowledge diffusion model is performed to determine the effect on the overall adoption curve for different parameter constellations of q_{12}. Table 2.3 summarizes the calibrated values for the simulation study.

Parameter	Value
p_1	0.13
p_2	0.01
q_1	0.75
q_2	0.50
q_{12}	0.07
m_1	1.00
m_2	1.00
ξ	0.05
ζ	5.00
σ	0.50

Table 2.3: Parameter values for system 2.11

The simulation of model 2.11 is done with `Matlab 6.5.0` with cross checks conducted with `Mathematica 5.2.`. Simulation results have been graphically represented in figure 2.3. In the left upper figure the adoption curves for $N(t)_1$, $N(t)_2$ and the overall market $N(t)_{all}$ have been drawn in red, green and blue colour respectively. We can see, that the knowledge diffusion process of the innovators comes to an end after around 6 periods, because the entire population of innovators has adopted new knowledge, which means that $m_1 = N(6)_1 = 1$ and thus $\dot{N}(t)_1 = 0$. On contrary, the knowledge diffusion process of the imitator group stops after around 20 periods of time with $m_2 = N(20)_2 = 1$ and thus $\dot{N}(t)_2 = 0$. Using the results from our stability analysis, we have realized a stable equilibrium at $m_1^* = m_2^* = 1$. Figure 2.2 gives a graphical representation of the equilibrium path for the simulated model based on parameter values in table 2.3 and with arbitrary starting values $N(0)_1 = N(0)_2 = 0$. Furthermore, figure 2.3 shows, that the unique equilibrium $m_1^* = m_2^* = 1$ is stable.

The left lower graphic contains the same information as the left upper figure but in relative numbers related to the market potential m_1 and m_2 respectively. The inflection points of the innovators and imitators are realized at around 2 periods for the innovators and at around 9 periods for the imitators. In the lower right panel you can find the relative analogon of the upper right panel. It is easy to see, that for $m_1 = 1$ and $m_2 = 1$ the upper right and the lower right figure coincide.

What impression can we get from figure 2.3 regarding the overall diffusion process $n(t)_{all}$. First, the knowledge diffusion process does not exhibit a bell shaped pattern,

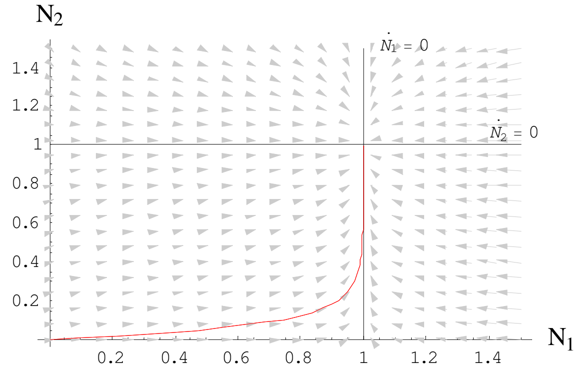

Figure 2.2: Phase plot of model 2.11 (II)

as the original (Bass, 1969) model but is unimodal with dent towards right. This is because the innovators still have adopted the entire knowledge and have realized the inflection point, whereas the imitators just start to adopt. Please note, that we do not observe the typically bimodal "chasm" pattern because cross sectional external knowledge transfer ($q_{12} > 0$) takes place. As shown later, the typically "chasm" pattern[17] of the knowledge diffusion process is only realized if $q_{12} = 0$.

As mentioned before, one of the drawbacks of this model is that the adoption curves $N(t)_1$ and $N(t)_2$ still both exhibit the typical S-shaped pattern, as one can see from the upper left and lower left pictures of figure 2.3. This assumption seems to be to strict. Thus, this strict pattern structure has been relaxed by assuming that the diffusion process is a mean reverting event and hence, short term deviation from a deterministic sigmoid adoption path, should be allowed. The simulation of model 2.29 is performed in the next section.

2.5.2 Simulation of stochastic knowledge diffusion model

In this section a simulation study of model 2.29 is conducted. Inserting the calibrated values from table 2.3 in 2.42 to 2.51 leads to

[17] Again, it is referred to (Van den Bulte and Joshi, 2007) for instance.

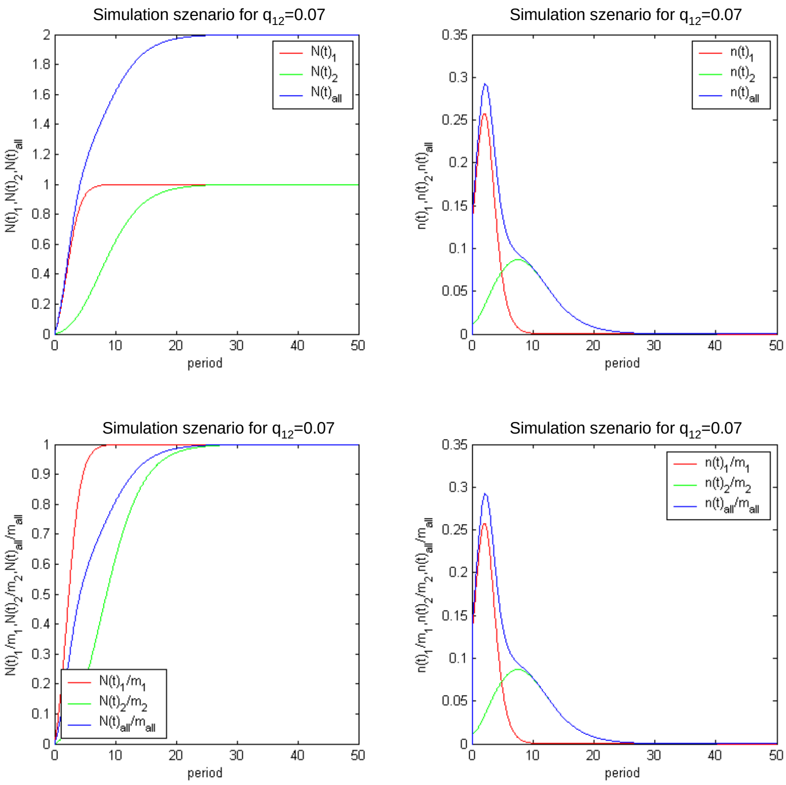

Figure 2.3: Graphical representation of simulated model 2.11 with $q_{12} = 0.07$

$$\phi_{0,1} = p_1 m_1 \zeta \xi^2 = 1.625 \times 10^{-3}, \tag{2.52}$$

$$\phi_{1,1} = \zeta \xi^2 (q_1 - p_1) = 7.75 \times 10^{-3}, \tag{2.53}$$

$$\phi_{2,1} = \frac{-q_1 \zeta \xi^2}{m_1} = -9.375 \times 10^{-3}, \tag{2.54}$$

$$\phi_{3,1} = -\zeta \xi = -0.25, \tag{2.55}$$

$$\phi_{0,2} = p_2 m_2 \zeta \xi^2 = 1.625 \times 10^{-3}, \tag{2.56}$$

$$\phi_{1,2} = \zeta \xi^2 [(\frac{m_2}{m_1 + m_2} q_1) - p_2] = 3 \times 10^{-3}, \tag{2.57}$$

$$\phi_{2,2} = \frac{-q_2 \zeta \xi^2}{m_1 + m_2} = -3.125 \times 10^{-3}, \tag{2.58}$$

$$\phi_{3,2} = \frac{m_2}{m_1 + m_2} \zeta \xi^2 q_{12} = 4.375 \times 10^{-4}, \tag{2.59}$$

$$\phi_{4,2} = \frac{-q_{12} \zeta \xi^2}{m_1 + m_2} = -4.375 \times 10^{-4}, \tag{2.60}$$

$$\phi_{5,1} = -\zeta \xi = -0.25. \tag{2.61}$$

As mentioned above, the simulation of the system of stochastic differential equation 2.29 has been performed with `Matlab 6.5.0`. If we refer to 2.5 on the upper left figure we can find the distribution functions $F(t)_1$ and $F(t)_2$ for the adoption process for innovators and imitators. Additionally, one can find in the same figure the discrete approximation of the distribution functions expressed by $\frac{N(t)_1}{m_1}$ for the innovators and $\frac{N(t)_2}{m_2}$ for the imitators. The overall distribution function $F(t)_{all}$ exhibits a dint pattern which commemorates slightly on a S-Shaped pattern. This is also the case for the discrete approximation $\frac{N(t)_{all}}{m_{all}}$ in the same subpicture 2.5. We can find also the density function $f(t)_1$ and $f(t)_2$ for the adoption process for innovators and imitators and the corresponding approximations $\frac{n(t)_1}{m_1}$ and $\frac{n(t)_2}{m_2}$. In the lower right figure additionally the relative alteration rates for adopting new knowledge $\Delta\Psi_{i,k} \equiv \Psi_{i,k} - \Psi_{i-1,k}$ for $k = \{1, 2\}$ are drawn. Further, for the entire population $N(t)_{all}$ we observe a S-shaped mean reverting behaviour with the largest deviation from the mean around the inflection point, as we should expect. Additionally, overall adoptions $\Psi_{i,all}$ exhibit a mean reverting behaviour with the largest fluctuations around the peak both for innovators and imitators.

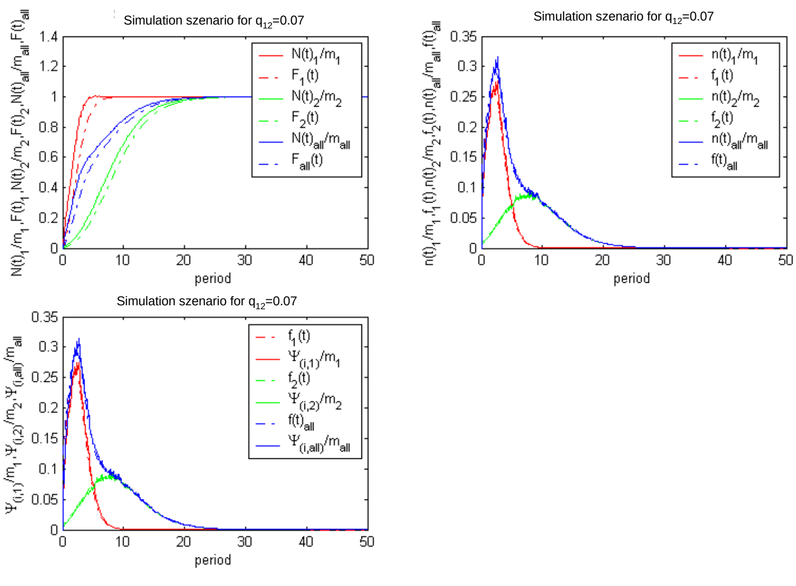

Figure 2.4: Graphical representation of simulated model 2.12 with $q_{12} = 0.07$

The interesting point is, how do changes of the knowledge transfer parameter q_{12} influence system 2.29 and how do variations of knowledge transfer affect the cumulative and adoption curves of the model 2.29 for both groups. For this purpose a sensitivity analysis for three scenarios has been performed: in the first scenario it is assumed, that knowledge transfer from the group of innovators to the group of imitators is prohibited. Please note again, that the knowledge transfer process is asymmetric, which means that knowledge transfer goes from the group of innovators to the group of imitators and not vice versa. The second scenario is characterized by a limited knowledge transfer, with $q_{12} = 0.07$ which corresponds to the already performed simulation. The last scenario assumes nearly complete knowledge transfer, which implicitly means that strong network effects are in place. For the simulation scenario, this means that $q_{12} = 0.99$.

The simulation results for the first and last simulation scenarios are depicted in figure 2.5 and 2.6.

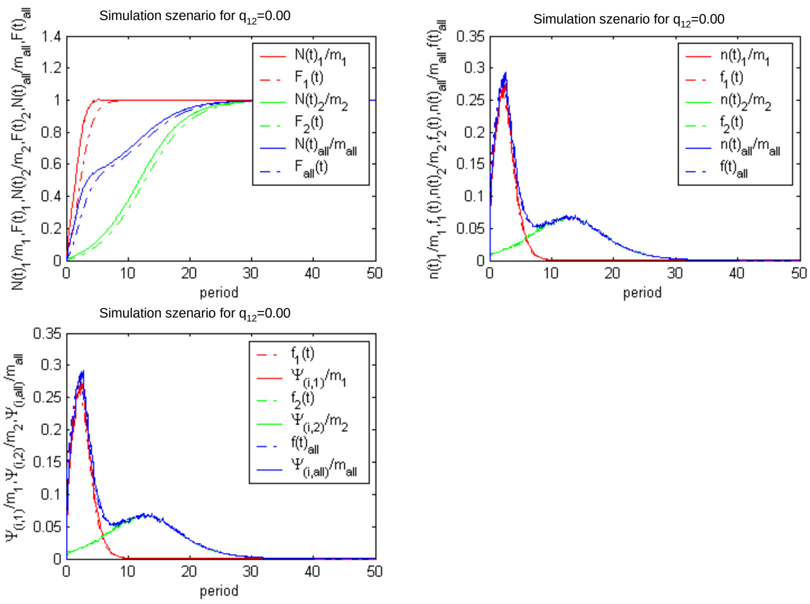

Figure 2.5: Graphical representation of simulated model 2.12 with $q_{12} = 0.00$

If we compare the figures 2.4, 2.5, 2.6 we can conclude the following: the less important network effects are, which coincides with parameter value of $q_{12} \to 0$ the more realistic is the so called "chasm" pattern. With other words: the longer the discrepancy between the realization of the inflection point of innovators and the beginning of imitator's adoption is, the more realistic is a bimodal shape of the adoption curve. On the other side, the stronger network effects are, the greater the parameter value of q_{12} is, the less realistic is the so called "chasm" pattern, because right before innovators have realized the inflection point imitators have nearly reached themselves their inflection point. In this way we can conclude that a bimodal pattern of overall knowledge diffusion is more likely if network effects are from less importance, whereas unimodal but not necessarily bell-shaped pattern in the sense of (Bass, 1969) of diffusion is more likely if strong network effects are in place.

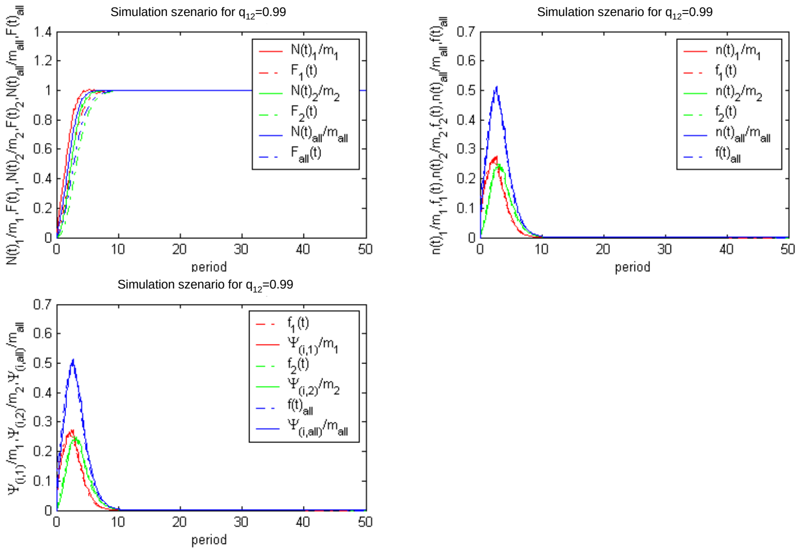

Figure 2.6: Graphical representation of simulated model 2.11 with $q_{12} = 0.99$

2.6 Econometric Annotations

The question which is unanswered is, how knowledge diffusion can be measured empirically, especially the parameter q_{12}. First, one has to find suitable proxies for knowledge diffusion. One possibility is to assume, that new knowledge is stored in scientist journals and citations of specific articles could be a proxy for diffusion of this new knowledge. Citations typically often have similarities with the diffusion of new products. At the beginning citations are low then they start growing and reach a peak before the citations tend to zero.

As one can see, system 2.29 can be estimated directly. Obviously, a seemingly unrelated regression (SUR) seems to be appropriated estimating system 2.29, because 2.29 is block recursive. Note, that this model assumes heteroscedastic errors because of the term $\sigma n(t)_k^{\gamma} \neq 0$. This again reflects the idea, that diffusion is more certain at the beginning and at the end of the diffusion process.

Before performing the SUR regression, the question to be answered is, whether the estimated coefficients are consistent or not. (Boswijk and Franses, 2005) have shown that the estimators do not exhibit the desired asymptotic normality behaviour by estimating a stochastic version of the (Bass, 1969) model. More precisely, the authors have shown that, even by increasing the sample period, the estimators $\phi \in \Phi$ cannot be estimated consistently at all. This result seems to be reasonable, because after realizing the saturation level m_1 or m_2 respectively, information no longer increases which is necessary to obtain consistent estimators of the parameter vector.

Although the before mentioned authors have shown, that inconsistent parameter estimates occur by estimating their stochastic version of the (Bass, 1969) model, they also concluded on the basis of Monte Carlo simulations for different time spans $H = \xi T$, that standard normal distribution can be consulted to approximate t-statistics of the estimated parameter vector, provided the inflection point lies within the sample period. But even if the estimators are asymptotically consistent, are they unbiased in small samples, which are standard for econometric applications in this research field? This topic is until today not addressed in literature.

2.7 Conclusion

In this chapter the link between knowledge transfer, knowledge diffusion and implicitly network effects has been investigated. For this reason a new diffusion model was put forward which focuses on those before mentioned aspects. The relevant literature has payed less attention investigating the link between knowledge transfer and knowledge diffusion. Particullarly, the question in which way knowledge transfer has an influence on the behavoiur of innovators and imitators within the adoption process is from interest.

The basis for this stochastic differential equation (sde) model is the well known (Bass, 1969) model. Although (Bass, 1969) mentioned that communication between innovators and imitators is relevant for adoption decision, this fact is not reflected in his mathematical derivations. Following (Kalish, 1985) and assuming that innovators need only "search" information to adopt new knowledge, while the latter imitators require "experience" type information before they adopt, a model which includes both the adoption decision of innovators and imitators is set up. In this way, the group of adopters has to be treated as heterogeneous. Further it was assumed, that information flows only in one direction, from innovators to imitators. Thus, the information flow is asymmetric.

After an appropriate discretization, in a simulation study it was shown, that the shape of the adoption pattern depends on the fact, if knowledge diffusion occurs or not. If knowledge transfer occurs, the stronger network effects are, so called unimodal patterns are more probable, because right before innovators have realized the inflection point, imitators have nearly reached themselves their inflection point. On contrary, the longer the discrepancy between the realization of the inflection point of innovators and the beginning of imitators adoption is, the less important network effects are, the more probable the called bimodal adoption phenomena are. Thus "chasm" patterns of adoption curves occur if network effects are from less importance.

The advantage of this new model is twofold: from a theoretical point of view, not only so called unimodal diffusion phenomena can be modeled, but also bimodal diffusion phenomena can occur. From an empirical point of view, the model which incorporates heteroscedastic errors and mean reverting can be theoretically estimated directly with a SUR approach.

So far this study suggests some avenues for further research. First of all and for now, the assumption that the market saturation level is exogenous and constant over time is very strict. Second, from a technical point of view, mean reverting is assumed to be the same over the entire population. Thus another source of heterogeneity can be introduced in the model by assuming different values for ζ. Third, after examining the large and small sample properties of the derived model the forecasting ability should be of interest.

3 The impact of learning and knowledge diffusion on industrial dynamics

3.1 Introduction

These days, firms R&D activities are mainly focused on the development of new processes, products and services. In a globalised world it is more than ever from outstanding importance to enhance the own firm competitiveness not only for present market positioning but also for long term market survival. From this background, it is not astonishing, as highlighted by (Clark and Fujimoto, 1991) and (Tushman and Nadler, 1986) for instance, that engagement into innovation activities cannot be considered as a piece of work beyond the call of duty but rather than the crucial duty to ensure firms future existence. Thus, innovation activities exert a direct influence on market activity, and thus on market share development. Further it can be assumed that innovation activities are directly linked to firm size and thus firm size also influences market activity.

From this background, the exploring of so called feedback processes between innovation, market share and firm size has gained much attention during the last years. For many years, the effects of innovation and firm size and the relationship between market share evolution and innovation have been discussed in isolation.

As stated by (Cohen and Levinthal, 1989) on p. 1070, "[a] methodological problem common to almost all the studies of the relationship between size and innovation is that they overlook the effect of innovation on firm growth (and hence, ultimately firm size). It is curious that the endogeneity of firm size, central to Schumpeter's notion of creative destruction, has been neglected, while the simultaneity associated with creative destruction has been recognized in some studies of the relationship between innovation and market concentration. This lacuna probably reflects the profession's primitive understanding of the determination of the size and growth of firms, and area of research that has just recently been revived."

As mentioned before, there subsists a large body of literature covering the relationship between firm size and innovation, which are primarily focused on manufacturing industries. These studies are heavily empirical based and are ambiguous with respect to the effects of firm size on innovation. For instance, (Mansfield, 1968) and (Schmookler, 1972) pointed out, that small firms tend to be more innovative than larger firms, whereas (Fisher and Temin, 1973) and (Vernon, 1974) found the contrary. (Kumar and Saqib, 1996) have found that the probability of engaging in R&D activities is positive correlated with firm size up to a certain threshold. Beyond this threshold R&D activity is declining. On contrary, (Wakasugi and Koyata, 1997) found, that firm size and innovation activity are not direct linked. They highlighted, that hence larger firms are more aggressive to pursue their innovation efforts but the efficiency of innovation is not necessarily enhanced by a growing firm size. (Cohen and Klepper, 1996) differentiates between process innovation and product innovation and found that process innovation increases with firm size.

If we now turn the focus on the effect of market structure on innovation, in principle two different scenarios are cogitable: the first is, that a positive relationship between monopoly power and innovative activity can be assumed, the second is, that innovative activity suffers from monopoly power. The first as well as the second relationship is from an empirical view documented in a voluminous literature.[1] (Scherer, 1967) and (Levin et al., 1985) for instance found an inverted U-pattern between market structure and innovation. This reflects the fact, that insufficient market power hinders firms to reduce so called up-front R&D effort, whereas an increasing market power reduces the incentive to engage in further R&D effort.

The problem of the before mentioned empirical orientated branches of literature are, that endogeneity problems occur, that means, ex ante it is not clear whether first the innovative activity determines firm size or firm size determines activity and second, the innovative activity determines market structure or market structure determines innovative activity. The problem one is confronted with, are feedback processes not only within the two branches, but also between the two branches.

Further, learning activities and knowledge diffusion play an important role when exploring feedback processes between innovation, market share and firm size. As mentioned by (Campagni, 1991), (Best, 2001), (Porter, 2000) and (Krugman, 1991) in a more spatial context, that inter-firm cooperation based on knowledge sharing can explain the predominance of small firms in the market. Learning can be described as a cognitive process of attaining new capabilities, to cope with not only the economic but also with the physical and social environment.[2]. Learning curves have both strategic

[1] See (Cohen and Levinthal, 1989) for a summary or more recently studies from (Nickell, 1996), (Nickell et al., 1997) and (Blundell et al., 1995) for instance which show unambiguously negative correlations between market structure and innovation.

[2] Refer to (Asheim, 1996).

and competitive implications for firms as mentioned by (Spence, 1981) and a more strategic dimension for planning decisions as highlighted by (Chand and Sethi, 1990).

Empirically, numerous studies have found that learning curves differ on the one hand on an inter-industrial level and on the other hand on an intra-industrial level. [3] Because innovations are produced by firms, knowledge is the presumption for this task. Thus knowledge is generated and transmitted in firms and between firms by human being, a micro view learning curve concept which is focused on personal learning seems to be appropriate. (Anderson and Schooler, 1991) for instance showed in psychological designed laboratory experiments that learning curves with diminishing returns are consistent with hyperbolic, square root, exponential and power functions. It is assumed that, knowledge generation depends first on not directly observable components such as talent, which is a proxy for the apprehension and second on the historically given stock of knowledge of an agent as highlighted by (Florida, 2002). Thus, knowledge generation can be interpreted as a separate production process in firms with input factors talent, grasp and of course time which is needed to accumulate knowledge. As mentioned by (Machlup, 1980) the creation and diffusion of knowledge is a core element of the production process and finally for the market structure in which the firm operates.

The aim of this study is to combine the effect of firm size, innovation and the effect of market structure on innovation with the effects of knowledge diffusion and learning. Thus this model is an extension of the work of (Mazzucato, 2000) in that way, as it explicit introduces a channel of knowledge diffusion, which is endogenously determined by learning activities. To integrate both aims, the so called replicator dynamics approach is disposed. The tool itself stems from evolutionary economics and is based on Darwin´s principle of natural selection. Particularly, on the basis of simulation experiments it will be investigated how learning and knowledge diffusion affect market structure. With this model it will be proofed whether and if yes learning activities need a dilution of one of the stylized facts regarding firm size dynamics which states, that early stages of an industry life cycle is characterized by instable market patterns.

The reminder of this chapter is structured as follows: In the first section a replicator dynamics model of market structure, innovation and firm size is introduced. After simulating the model it is expanded by the aspect of learning and inter firm knowledge transfer. Section four deals with the simulation of the before expanded model. In section five a conclusion of the derived results is given.

[3] Refer for instance to this topic on (Hayes, 1986), (Dutton and Thomas, 1984) and (Pisano et al., 2001).

3.2 The basic model

3.2.1 Setup

In this section, a replicator dynamics model is introduced which is based on the works of (Mazzucato, 2000), (Cantner and Hanusch, 1998), (Malerba and Orsenigo, 1993) and (Noailly et al., 2003) for instance and the seminal work of (Dosi, 1982), who introduced the so called paradigm-trajectory-approach and hereby identified stylized facts for the evolution of an industry.

In the model economy, it is assumed that set of agents or firms $\mathcal{B}$ exist, who produce under n strategies with $i = \{1, 2, ..., h-1, h, h+1, ..., j-1, j, j+1, ..., n-1, n\}$. To keep the model simple, it is further assumed that every agent with strategy i produces with a linear production technique with the only input N_t. Hereby N_t can be considered as a regenerative but exhaustable resource[4] which is growing with a certain exogenous and time independent degree ξ in every period t, $t \in T$.

In contrast to the old neoclassical theory, it is not assumed a production technique which uses input factors available ad infinitum in extremum. It is more realistic that the production decision depends on scarce resources, especially in the short run.

The reason for this assumption is to endogenize the production decision which is directly linked to the resource dynamics via the cost of production. It is referred later to this point.

In this way, it is followed (Noailly et al., 2003), with the exception that we give a wider definition of the evolution of the input factor N_t. It is worth mentioning that one has to harvest the resource before using it as the input factor[5].

For this reason, it is assumed that the agents harvest a specific stock N_t of a natural resource in every period of time t. The maximum carrying capacity of our resource is defined by M, which is obviously time independent and exogenously given. As usual in resource economics, it is considered a logistic growth of the resource N_t:

$$\frac{dN_t}{dt} = \xi N_t \left[1 - \frac{N_t}{M}\right] - \psi E(N_t). \tag{3.1}$$

Equation 3.1 is often called the "Schaefer equation", which is gathered from the Gordon-Schaefer model (Gordon, 1954), which is often used to discuss issues stemming

[4] See for instance (Dasgupta and Heal, 1979).
[5] The cost of harvesting are not considered in this model.

from resource economics. As we can see from equation 3.1, it is proposed that a fixed quantity of the natural resources is removed in every period of time, for example in every year. Furthermore, ξ represents the exogenous growth rate of the resource as mentioned before and $E(N_t)$ represents the aggregate harvesting function, which depends on the stock N_t. ψ represents the exogenous catchability coefficient, which is of no further interest.

To make the discussion easier, in the following it is focused on the two strategy case $i = \{h, j\}$. Hence, strategy i is associated with firm i. In this manner, a channel is created to introduce agent specific heterogeneity (Noailly et al., 2003).

The two strategies can be formulated as follows: the first strategy h we label the "green strategy", which means that this strategy is less productive but less resource intensive than the strategy j, $h \neq j$. The other strategy j is called the "black strategy", because it is more resource intensive but more productive than the first one. Hence, one can conclude that

$$E(\cdot)^h C(\cdot)^h > E(\cdot)^j C(\cdot)^j \tag{3.2}$$

must hold. In equation 3.2, C^i stands for the cost of production and E^i stands for the effort of strategy i, which is given exogenously.[6]

It is further implied that we can use the resource as input factor directly, which means that we do not include an intermediate good production sector in the model. The implication of this assumption is that the cost of production must include the cost of harvesting and furthermore, the cost of harvesting must equalize with the cost of production since no other costs of production are included in the analysis. Subsequently, in the following the terminology "cost of production" is used.

As mentioned before, we postulate a linear production function with input N_t. Thus, one can write for the production in period t, $\forall t$:

$$F_t(N_t)^i = E^i N_t, \; i = \{h, j\} \,, \forall t \tag{3.3}$$

As usual in resource economics, the costs per product are defined as $c_t^i(N_t) \equiv \frac{C_t^i}{N_t+1}$,

[6] Unless it is necessary, I leave the time index t for convenience.

which implies the more resource intensive the production is, the more expensive is the extraction of a fraction of the stock N_t in the next period. Therefore, an implicit forward looking agents´ behavior is assumed. Of course, if $N_t = 0$, then $C_t^i(0) = 0, \forall t$ per assumption. Consequently, N_t has to be treated as a necessary input factor for production. Again the reader has to bear in mind that the only purpose of the above mentioned assumption is to endogenize the production decision via endogenous cost of production.

With the above mentioned assumption it is now possible to deduce the profit per period t of the agents strategies $i = \{h, j\}$ which depends exclusively on the stock of the resource N_t, as one easily can derive from the next equation:

$$\Pi_t^i \equiv E^i N_t(p - c^i),\ i = \{h, j\}. \tag{3.4}$$

From that equation it is easy to obtain the profit per unit of strategy i in period t π_t^i as follows:

$$\pi_t^i \equiv \left[\frac{\Pi^i}{F_t^i}\right] = p - c^i, \tag{3.5}$$

where p stands for the exogenous price level.

Additionally and similar to (Noailly et al., 2003) it is postulated that the aggregate harvesting function is a convex combination of the single harvesting functions F^i. If we assume during a certain period of time a fraction β of the total population $\mathcal{B}, \beta \in \mathcal{B}$ explicitly decides to use the strategy i with $\sum_i s^i = 1$ we can formulate the aggregate harvesting function as

$$E(N_t) = \sum_i s^i \beta F^i(N_t). \tag{3.6}$$

s^i stands for the market share of using strategy i.[7] With the last paragraph we have described the production sector totally.

To sum up, the main purpose of this section is to describe the dependence of the

[7] One can set the $\mathcal{B} = 2$ so that one strategy i corresponds to an agent i in the two strategy case $i = \{h, j\}$.

evolution of the scarce resource N in the production sector and the influences of the evolution of N_t on the cost of production C^i under usage of a certain strategy i from a pool of strategies n. In the next section, I proceed with some comments concerning the market evolution in the model.

3.2.1.1 Market share evolution

To expand the dynamic dimension in the model, it is presupposed that the market share s^i under usage of strategy i will change over time. Therefore, we have to acknowledge the time aspect in the expression of s^i. To model the dynamic dimension of s^i, we recur to some facts from the field of population genetics, on which evolutionary economics is mainly based.

In the year 1908, (Hardy, 1908) published a striking article, which can be treated as a cornerstone for mathematical orientated population genetics. In this article it is assumed that

- some genetic frequencies, which he labelled (q, p) of two allele of a certain gene position have not to be unchanged by reproduction over the generations, belong to a certain population. The implication of this assumption is that the possibility of selection is excluded.
- the probabilities of belonging to a certain genotype (x, y, z) is exclusively defined by the initial co-generation in the way that: $x = p^2,\ y = 2pq,\ z = q^2$.

(Fisher, 1930) formulated a general equation of population genetics which bases on the ideas of (Hardy, 1908):[8]

$$\dot{x}^i = x^j \left(\sum_i \omega^{ij} x^j - \sum_{r,s} \omega^{r,s} x^r x^s \right), \tag{3.7}$$

with $\sum_i \omega^{ij} x^j$ as the fitness of reproduction of all genotypes $A^i A^j$, ω as the advantage of survival and $\sum_{r,s} \omega^{r,s} x^r x^s$ as the average fitness for all other genotypes from $A^r, A^s, \in \mathcal{H}, \{r, s\} \neq j$. The expression $\sum_j \omega^{ij} x^j - \sum \omega^{r,s} x^r x^s$ can be interpreted as the advantage of survival, as a consequence.

Next, the ideas of the (Fisher, 1930) equation on our problem of how to model the market share evolution are adopted. From the above gained facts, we can conclude that s^i depends solely on the comparison between the fitness f^i and the average fitness

[8] In this equation i and j denote the generations of genotypes.

$\bar{f}$ of the chosen strategy i of an agent.

In general, fitness depends on a n-dimensional vector $\mathbf{s}$ which contains the relative frequency of all possible replicators. Accordingly, one can write:

$$\dot{s^i} = s^i[f^i(\mathbf{s}) - \bar{f}(\mathbf{s})]. \tag{3.8}$$

Now we are able to adopt this general equation for our purpose. If we assume that the rate of capacity enlargement g^i corresponding to the usage of strategy i is positive related to the profit per unit π_t^i, we can write:

$$g^i = \gamma\pi^i = \gamma(p - c^i) = \gamma\left[p - \frac{C^i}{N+1}\right], \tag{3.9}$$

with the reaction coefficient $\gamma > 0$.

Next the average costs per product are defined as $\bar{c} = \sum_n s^i c^i$, the average capacity enlargement rate or the average growth rate of the population of firms using a profitable strategy as $\bar{g} = \sum_n s^i g^i$ and set $f^i(\mathbf{s}) = g^i(\mathbf{s})$. Together with the derived equation we can write

$$\dot{s}^i = s^i(g^i - \bar{g}). \tag{3.10}$$

After doing some algebra, we can rewrite equation 3.10 together with equation 3.8 and equation 3.9 for strategy h for instance as follows:

$$\dot{s}^h = \gamma s^h(\bar{c} - c^h) = \gamma s^h\left(\frac{\bar{C} - C^h}{N+1}\right) = (1 - s^h)\left(\frac{C^j - C^h}{N+1}\right)s^h, \tag{3.11}$$

whereas in the last step $\gamma = 1$ is assumed. What can we gain from this last derived equation? By a given stock of N_t the evolution of strategy s^i depends only on the cost relation to a competing strategy j. If the cost difference $\Delta(C) \equiv C^h - C^j = 0$, then, the agents should be indifferent between these two strategies from the pool n by a given level of E^i or the agents have no incentive to change their strategy. Otherwise we have a strictly dominating strategy $s^h \succ s^j$ for $\Delta(C) \equiv C^h - C^j < 0$ et vice versa. Further, one can derive the following relationship of s^h and C^h for a given stock of N_t:

$$\frac{\partial \dot{s}^h}{\partial C^h} < 0 \tag{3.12}$$

and

$$\frac{\partial \dot{s}^h}{\partial C^j} > 0. \tag{3.13}$$

Subsequently, for a steady state of s^h, s^{h*} (which means that $\dot{s}^h = 0$) we can conclude:

$$\dot{s}^h = 0 \Leftrightarrow s^{h*} = 1 \wedge C^{h*} = C^{j*} \text{ for } s^{h*} \in (0,1). \tag{3.14}$$

Therefore, market survival depends on the fitness of a firm i which is exclusively governed by the firms´s cost structure, as one can see from equation 3.11. The main purpose of this section was to give a guess on how the evolution of the market share s^i depends on the resource N_t and the cost of production C^i. Until now, we have an idea about the market structure and the production sector. In the next section, technological progress is introduced.

3.2.1.2 Technological progress and market selection

As mentioned before, agent specific heterogeneity via different cost regimes has been modeled. It is plausible to assume that agents invest in a less cost intensive producing technology.

In this way, a further dimension of what is called the structural dynamic aspect in the model is highlighted. It is easy to see why. For a moment let us assume that an agent uses a strategy h with cost of production tending zero in the long run. Compared to a competing strategy j it is straightforward that this strategy j is ruled out of the strategy field of all agents which are producing in the market if $C^j \mapsto \bar{C} > 0$ for $t \mapsto \infty$. Hence, we have, looking at our previous results, a strictly dominating strategy h which monopolized the market. Consequently, the market itself is monopolized because the market share s^h by using strategy h tends to the value 1 in the long run. That‘s exactly the link between the existence of technology progress and how it influences the market structure in the long run. Of course, in the short run one can imagine some turbulences a propos the market evolution. This observation covers the industry life cycle assumptions (Malerba and Orsenigo, 1993).

Herewith, a direct link from the model to some ideas of Schumpeter on the subject of the dimension of structural dynamics and volatility is created. A wide known thesis proposed by Schumpeter is that creative destruction is a necessary condition for innovative firms. Of course, such firms have to dispose of financial potential to invest in R&D. Schumpeter assumes that the financial potential of firms is positive correlated to the market power of the firm. Thus, we realize a process of creative destruction mainly driven by R&D investments of innovative firms. The implication is that monopoly power is a necessary condition to create incentive for investments into a technology which itself drives technological progress. (Neumann et al., 1982) conclude "that larger firms ... acquire smaller firms in order to exploit the innovative potential originated in these firms"[9].

It is also obvious that a tradeoff between the static and dynamic characteristics of competition exists. On the one hand one can realize extra rents from monopoly power which ensure growth, on the other hand, we have to acknowledge an allocative loss of efficiency. Consequently, the size of firms, the degree of concentration and innovativeness are positive correlated. From this follows that a higher degree of concentrated industries must exhibit higher growth rates (Schumpeter, 1942). The "Schumpeter-Mark-II" is a major element in the frame work of the models of endogenous growth, which are mainly promoted by (Aghion and Howitt, 1992).

On contrary, (Arrow, 1962) showed in his article that the incentive of investing in R&D is negative correlated with the market power of an industry. He compares the gain of a cost reduction process innovation in a competitive world with the additional gain of cost reduction process innovation in a concentrated industry. He shows that the increase of profit in a competitive world is larger than in a monopoly.

The implication of the above mentioned is that many small firms are more innovative in a competitive world, while few but large firms are more innovative in a concentrated world.[10]

It is short mentioned that a mass of literature exists which aims to test the Schumpeter hypothesis empirically.[11]

The question arises how we can integrate some facts of the Schumpeter hypothesis in

[9](Neumann et al., 1982), p. 135.

[10]Refer to (Acs and Audretsch, 1987).

[11]For the relationship between the size of firms, the degree of concentration and innovativeness refer to (Cohen and Levinthal, 1989) and (Kamien and Schwartz, 1975) and for the relationship between the firm size and innovativeness refer to (Frisch, 1993). For German data refer to (Neumann et al., 1982), (Entorf, 1988), (Kraft, 1989), and to (Bertschek and Entorf, 1996) for Belgian, German and French data.

our model? If one recalls our formulation of equation 3.11, a direct link between market share development and the degree of concentration can be derived. It is straightforward that the higher the market share s^i is the more successful the strategy i must be. On that account, we can suppose that the incentive to invest in a more successful strategy i is higher compared to an inferior strategy j. It follows that the size of a firm and the power of investment are striking factors for the market structure as well as for the evolution of the market share s^i. Therefore, they both are positive correlated to the success of a strategy i.

The implication is, if we follow (Phillips, 1971), that the positive correlation of the firm size and the innovativeness follows a "success-breeds-success" hypothesis.[12] This implies that a positive dependence of successful innovation activity in the current period and investment endeavours in the next periods exist. Following (Cantner and Hanusch, 1998) or (Malerba and Orsenigo, 1993) this interpretation is in line with the so called "Schumpeter-Mark-II" hypothesis. On the other side, one could argue that smaller firms are more innovative because their behaviour regarding to investment decisions is more flexible (Malerba et al., 1997). This view is equivalent to the "Schumpeter-Mark-I" hypotheses.[13]

It is worth noting that the "Schumpeter-Mark-I" and the "Schumpeter-Mark-II" hypothesis are common patterns which can occur both through the industry lifecycle, whereas the early stage is characterized by the "Schumpeter-Mark-I" hypothesis, while the later periods are more in line with "Schumpeter-Mark-II". The implication is that a "Schumpeter-Mark-I" regime should be more volatile than a regime based on the "Schumpeter-Mark-II" hypothesis. It is common measuring the stability with a so called "instability index" [14] which is defined as:

$$\Im = \sum_{i}^{n} |\dot{s}^i|, \tag{3.15}$$

with s^i as the market share of strategy i at time t.

To catch this interesting ideas, I follow (Malerba and Orsenigo, 1993), (Malerba et al., 1997), (Mazzucato, 1998), (Mazzucato, 2000) and (Cantner and Hanusch, 1998) and create a direct link from the "Schumpeter-Mark-I" hypothesis to increasing returns

[12] See for instance (Flaig and Stadler, 1994) who have found an empirical confirmation of the success-breeds-success hypothesis for West-Germany using a German panel data set.

[13] For instance refer to (Acs and Audretsch, 1987), (Malerba and Orsenigo, 1993) and (Malerba et al., 1997).

[14] This index was first introduced by (Hymer and Pashigan, 1962).

of investing in a new, less cost intensive, technology. Instead of the "Schumpeter-Mark-I" hypothesis, the "Schumpeter-Mark-II" hypothesis is directly associated with the assumption of decreasing returns of investing in a new cost reduction technology. Additionally, the case of constant returns to scale for the sake of completeness has been incorporated. The latter case is extensively discussed by (Metcalfe, 1994).

Technological progress is reflected as intra-industry cost reduction over time. The parameter $\theta \in (0,1)$ denotes the speed of cost reduction. With respect to "Schumpeter-Mark-I", "Schumpeter-Mark-II" the cost evolution process can be separated in a dynamic constant return to scale case, dynamic increasing return to scale case and a dynamic constant return to scale case. Hence, it is possible to write for $\dot{C}^i$:

$$\dot{C}^i = \begin{cases} -\theta C^i \\ -\theta C^i s^i \\ -\theta C^i (1 - s^i) \end{cases} . \tag{3.16}$$

As one can see from equation 3.16, the first line represents the case of constant returns to scale, the second line the case of increasing returns to scale and the last line the case of decreasing returns to scale. In the next chapter, the formulation of a model of market selection which is driven by structural dynamics is given.

3.2.1.3 Summary

This section gives a summary of model elements. The model integrates the following aspects:

1. Resource dynamics N_t, which influence
2. the selection of a producing strategy s^i via an endogenous cost structure C^i,
3. which is itself driven by technological progress θ in the cost structure.

These three points can be summarized in a more mathematical manner as follows:

$$\begin{cases} \dot{N} & = & \xi N \left[1 - \frac{N}{M}\right] - \psi\beta \left[\sum_n s^h F^h(N)\right] \\ \dot{s}^h & = & s^h(-c^h + \bar{c}) \\ \dot{C}^h & = & \begin{cases} -\theta C^h \\ -\theta C^h s^h \\ -\theta C^h (1 - s^h) \end{cases} \\ \dot{C}^j & = & \begin{cases} -\theta C^j \\ -\theta C^j s^j \\ -\theta C^j (1 - s^j) \end{cases} \end{cases} . \tag{3.17}$$

Because of the fact, that the cost structure of firm h depends on the market share of firm j et vice versa, we have to analyze a system of four non linear differential equations for firm i.

Again, the first line of the system of equations 3.17 represents the evolution of the resource N_t, the second equation gives an impression of how the market share s^i is influenced by using strategy i. Whereas the last two line distinguishes between the different kinds of returns to scale respective to the investment into a new cost reduction technology.

The next step is to solve the model and examine its dynamic behaviour especially in the short run. The emerging question is how to study the dynamics system 3.17. As can be seen from above, we are confronted with a system containing non-linear differential equations. Therefore, in the next section some comments with respect to the dynamic behaviour are added.

3.2.2 Dynamic behaviour of the basic model

To discuss the dynamic behavior of our system, we have first to investigate, whether a steady state in the sense of a long run equilibrium exists. This is equivalent to the postulation that the partial derivatives over time of $\dot{N}_t$, $\dot{s}_t$ and $\dot{C}_t$ must be zero. Hence, we can formulate the following proposition:

Proposition 1: On behalf of the assumption that the partial derivatives of $\dot{N}_t$, $\dot{s}_t$ and $\dot{C}_t$ exist and that $\dot{N}_t = 0$, $\dot{s}_t = 0$ and $\dot{C}_t = 0$ hold simultaneously $\forall t$, the system 3.17 has a unique steady state vector $\mathbf{S}$, which contains N^*, s^{i*} and C^{i*} in the long run.$\square$

Next, I give a brief sketch of how to prove proposition 1.

Proof 1: A steady state vector exists, if and only if $\dot{N} = \dot{s}^i = \dot{C}^i = 0$ holds. This is realized, if

$$\begin{cases} \xi N^* \left[1 - \frac{N^*}{M}\right] & = & \psi\beta\left[\sum_n s^{i*} F^i(N^*)\right] \\ 0 & = & s^{i*}(-c^{i*} + \bar{c}) \\ 0 & = & \begin{cases} -\theta C^{i*} \\ -\theta C^{i*} s^{i*} \\ -\theta C^{i*}(1 - s^{i*}) \end{cases} \end{cases} . \tag{3.18}$$

As one can see from the system of equations 3.17, the equation for the evolution of C^i is influenced only by s^i but not by N for the increasing and decreasing returns to scale case each. For the constant returns to scale case the evolution of C^i is purely autonomous in the sense that it is not influenced by N_t or s^i. As a consequence the value of θ is an important determinant for the market structure evolution s^i. Thus, we can find the following, which holds asymptotically:

1. Assume now, that a value of θ exists which is greater than a threshold value of θ, $\tilde{\theta}$ and near to a maximum value of θ, called θ_{max} so that $\theta_{max} > \theta \gg \tilde{\theta}$ holds. Then, speed of cost reduction is very fast and as a result after a short period of time $\dot{s}^i = 0$. Additionally, from the second and third line of the steady state system follows immediately that $C^{h*} = C^{j*} = 0$ and $s^{i*} \in (0, 1)$, for every case of returns to scale assumption. We obtain $N^* = M\left[1 - \frac{\psi\beta[\sum_n E^i s^{i*}]}{\gamma}\right]$.
2. Further, assume that a smaller value of θ exists which is near to the minimum value of θ, called θ_{min}, and smaller than the threshold value $\tilde{\theta}$. Then $\tilde{\theta} \gg \theta > \theta_{min}$ holds, obviously. Accordingly, technological progress is very slow. For $C^h(0) > C^j(0)$[15] in $t = 0$ follows
   a) for the constant returns of scale case: $C^{h*} = C^{j*} = 0$ which means that $s^{h*} = 0$. In addition, we obtain $N^* = M\left[1 - \frac{\psi\beta E^j}{\gamma}\right]$.
   b) for the increasing returns to scale case: $C^{h*} \in \mathbb{R}_+ \wedge C^{j*} = 0$ which means that $s^{h*} = 0$. Once again, we obtain $N^* = M\left[1 - \frac{\psi\beta E^j}{\gamma}\right]$.
   c) for the decreasing returns to scale case with the further assumption that a $\epsilon \mapsto 0$ exists, for which one assume that $\epsilon < \tilde{\epsilon} \equiv |\theta - \theta_{min}|$: $C^{j*} \in \mathbb{R}_+ \wedge C^{h*} = 0$ which means that $s^{h*} = 0$. Once more, we obtain $N^* = M\left[1 - \frac{\psi\beta E^j}{\gamma}\right]$.

[15] Of course, one can assume $C^h(0) < C^j(0)$. If $C^h(0) = C^j(0)$ we cannot observe any dynamic of s^i and N_t right from $t = 0$. Then $s^i(0) = s^{i*}$ follows.

From point 2 of proof 1 follows that $\tilde{\theta} \gg \epsilon \gg \tilde{\epsilon}$ must hold. For that reason, we obtain $N^* = M\left[1 - \frac{\psi\beta[\sum_n E^i s^{i*}]}{\gamma}\right]$ for $s^{i*} \in (0,1)$.

3. If 1 or 2 of proof 1 holds, then $\exists\, N^* \in \mathbb{R}_+ \setminus \{0\}$.

4. If one assumes $N^* = 0$, then a set $\mathcal{I}$ of degenerated equilibria is realized for $s^{i*} \in (0,1)$ and $C^{h*} = 0$ because one degree of freedom is left to set s^{i*}.

■

Hence, a set of steady state values must be taken into account, which could all exist. But what follows from proof 1 intuitively?

First, the dynamic is only driven by the parameter θ, which is purely exogenous per assumption. Consequently, we obtain different scenarios regarding to our market structure depending only on the parameter of technological progress which is not explained by our model.[16]

Second, if one sets $\theta = 0$ we obtain a two-dimensional system consisting only in the development of N_t and s^i.

Thus, this model can be described as a variant of the model of (Noailly et al., 2003).

On the other hand, if we handle N_t as constant $N_t = N$, $\forall t$, we obtain a market structure model similar to (Cantner and Hanusch, 1998).

Now we will proceed with the stability analysis of the model.

3.2.3 Stability analysis of the basic model

As mentioned before, system 3.17 consists of four non-linear differential equations for each firm i. To prove the local stability of system 3.17, we can linearize the system around the steady states[17]. We can follow this way, because the Hartman-Grobman[18] theorem states, that the behaviour of a non-linear dynamical system near a hyperbolic equilibrium point is qualitatively the same as the behaviour of its linearization near the origin. Further, an equilibrium is called non-hyperbolic if one of the Eigenvalues of the linearized system 3.17 has a real part equal to zero. If this is the case, linearization cannot be applied to proof local stability of system 3.17.

After the linearisation of system 3.17, the Jacobi-Matrix for the constant returns to

[16]Technological progress falls like "manna from heaven". See for instance (Frenkel and Hemmer, 1999), p. 113.
[17]For this topic refer to appendix 1.
[18]Refer to (Guckenheimer and Holmes, 1983) for instance.

scale reads as:

$$Jac^{CRS} \equiv \begin{bmatrix} \frac{\partial \dot{N}}{\partial N} & \frac{\partial \dot{N}}{\partial s^h} & \frac{\partial \dot{N}}{\partial C^h} & \frac{\partial \dot{N}}{\partial C^j} \\ \frac{\partial \dot{s^h}}{\partial N} & \frac{\partial \dot{s^h}}{\partial s^h} & \frac{\partial \dot{s^h}}{\partial C^h} & \frac{\partial \dot{s^h}}{\partial C^j} \\ \frac{\partial \dot{C^h}}{\partial N} & \frac{\partial \dot{C^h}}{\partial s^h} & \frac{\partial \dot{C^h}}{\partial C^h} & \frac{\partial \dot{C^h}}{\partial C^j} \\ \frac{\partial \dot{C^j}}{\partial N} & \frac{\partial \dot{C^j}}{\partial s^h} & \frac{\partial \dot{C^j}}{\partial C^h} & \frac{\partial \dot{C^j}}{\partial C^j} \end{bmatrix} =$$

$$= \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ -s^{h*}\frac{(C^{h*}-C^j)^*(s^{h*}-1)}{(N^*+1)^2} & \frac{(C^{h*}-C^{j*})(2s^{h*}-1)}{N^*+1} & \left[\frac{s^{h*}-1}{N^*+1}\right]s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right]s^{j*} \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & -\theta \end{bmatrix}. \quad (3.19)$$

Next, we have to evaluate the Jacobian at their steady state values for N^*, s^{h*}, C^{h*} and C^{2*} for the constant returns to scale case. Refering to point 1 and 2a) of proof 1[19] we can conclude that we obtain two different versions of the Jacobian Jac_f^u[20] matrix:

$$Jac_1^{CRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & 0 & \left[\frac{s^{h*}-1}{N^*+1}\right]s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right]s^{j*} \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & -\theta \end{bmatrix}, \quad (3.20)$$

$$Jac_{2a)}^{CRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & -\theta \end{bmatrix}. \quad (3.21)$$

The corresponding Eigenvalues for Jac_f^u coincides for both fixed points and read as[21]:

$$\Psi_{1,2a)}^{CRS} \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma\frac{N^*}{M} \\ -\theta \\ -\theta \end{bmatrix}.$$

[19]The degenerate case (point 4) of proof 1 has been neglected, because sustainability has been assumed, which coincides with $N > 0$.

[20]The subscript f denotes to the numeration of proof 1, whereas the superscript u denotes to the cases of returns to scale: CRS stands for the constant returns to scale case, IRS stands for the increasing returns to scale case and DRS stands for the decreasing returns to scale case.

[21]For convenience, we stack the Eigenvalues in a vector Ψ_f^u each

Obviously, the fixed points are non-hyperbolic.

Now, we proceed with the computation of the fixed points for the increasing returns to scale case. The Jacobian now can be written as:

$$Jac^{IRS} \equiv \begin{bmatrix} \frac{\partial \dot{N}}{\partial N} & \frac{\partial \dot{N}}{\partial s^h} & \frac{\partial \dot{N}}{\partial C^h} & \frac{\partial \dot{N}}{\partial C^j} \\ \frac{\partial \dot{s^h}}{\partial N} & \frac{\partial \dot{s^h}}{\partial s^h} & \frac{\partial \dot{s^h}}{\partial C^h} & \frac{\partial \dot{s^h}}{\partial C^j} \\ \frac{\partial \dot{C^h}}{\partial N} & \frac{\partial \dot{C^h}}{\partial s^h} & \frac{\partial \dot{C^h}}{\partial C^h} & \frac{\partial \dot{C^h}}{\partial C^j} \\ \frac{\partial \dot{C^j}}{\partial N} & \frac{\partial \dot{C^j}}{\partial s^h} & \frac{\partial \dot{C^j}}{\partial C^h} & \frac{\partial \dot{C^j}}{\partial C^j} \end{bmatrix} =$$

$$= \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ -s^{h*}\frac{(C^{h*}-C^j)^*(s^{h*}-1)}{(N^*+1)^2} & \frac{(C^{h*}-C^{j*})(2s^{h*}-1)}{N^*+1} & \left[\frac{s^{h*}-1}{N^*+1}\right] s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right] s^{j*} \\ 0 & -\theta C^h & -\theta s^{h*} & 0 \\ 0 & 0 & 0 & -\theta s^{j*} \end{bmatrix}. \quad (3.22)$$

Evaluating the Jacobian at N^*, s^{h*}, C^{h*} and C^{j*} we obtain:

$$Jac_1^{IRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & 0 & \left[\frac{s^{h*}-1}{N^*+1}\right] s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right] s^{j*} \\ 0 & 0 & -\theta s^{h*} & 0 \\ 0 & 0 & 0 & -\theta s^{j*} \end{bmatrix}, \quad (3.23)$$

$$Jac_{2a)}^{IRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & \frac{-C^{h*}}{1+N^*} & 0 & 0 \\ 0 & -\theta C^h & 0 & 0 \\ 0 & 0 & 0 & -\theta s^{j*} \end{bmatrix}. \quad (3.24)$$

Once more, we define a column vector Ψ_f^u which contains the Eigenvalues for the first equilibrium as follows:

$$\Psi_1^{IRS} \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma\frac{N^*}{M} \\ -s^{h*}\theta \\ -s^{j*}\theta \end{bmatrix}.$$

In the same way we stack the Eigenvalues for the second equilibrium in a column

$$\text{vector: } \Psi_{2b)}^{IRS} \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma\frac{N^*}{M} \\ -\frac{C^{h*}}{1+N^*} \\ -s^{j*}\theta \end{bmatrix}.$$

Finally, we compute the fixed points for the decreasing returns to scale case with the corresponding Jacobian as:

$$Jac^{DRS} \equiv \begin{bmatrix} \frac{\partial \dot{N}}{\partial N} & \frac{\partial \dot{N}}{\partial s^h} & \frac{\partial \dot{N}}{\partial C^h} & \frac{\partial \dot{N}}{\partial C^j} \\ \frac{\partial \dot{s^h}}{\partial N} & \frac{\partial \dot{s^h}}{\partial s^h} & \frac{\partial \dot{s^h}}{\partial C^h} & \frac{\partial \dot{s^h}}{\partial C^j} \\ \frac{\partial \dot{C^h}}{\partial N} & \frac{\partial \dot{C^h}}{\partial s^h} & \frac{\partial \dot{C^h}}{\partial C^h} & \frac{\partial \dot{C^h}}{\partial C^j} \\ \frac{\partial \dot{C^j}}{\partial N} & \frac{\partial \dot{C^j}}{\partial s^h} & \frac{\partial \dot{C^j}}{\partial C^h} & \frac{\partial \dot{C^j}}{\partial C^j} \end{bmatrix} =$$

$$= \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ -s^{h*}\frac{(C^{h*}-C^j)^*(s^{h*}-1)}{(N^*+1)^2} & \frac{(C^{h*}-C^{j*})(2s^{h*}-1)}{N^*+1} & \left[\frac{s^{h*}-1}{N^*+1}\right]s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right]s^{j*} \\ 0 & \theta C^h & -\theta(1-s^{h*}) & 0 \\ 0 & 0 & 0 & -\theta(1-s^{j*}) \end{bmatrix}. \quad (3.25)$$

The evaluation of the Jacobian at her steady state values for N^*, s^{h*}, C^{h*} and C^{j*} result in:

$$Jac_1^{DRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & 0 & \left[\frac{s^{h*}-1}{N^*+1}\right]s^{h*} & \left[\frac{s^{h*}}{N^*+1}\right]s^{j*} \\ 0 & 0 & -\theta(1-s^{h*}) & 0 \\ 0 & 0 & 0 & -\theta(1-s^{j*}) \end{bmatrix}, \quad (3.26)$$

$$Jac_{2c)}^{DRS} = \begin{bmatrix} -\gamma(\frac{N^*}{M}) & -\psi\beta N^*(E^h - E^j) & 0 & 0 \\ 0 & \frac{C^{j*}}{1+N^*} & 0 & 0 \\ 0 & 0 & -\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (3.27)$$

We obtain two vectors Ψ_f^u containing the Eigenvalues of the two equilibria:

$$\Psi_1^{DRS} \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma(\frac{N^*}{M}) \\ -(1-s^{h*})\theta \\ -(1-s^{j*})\theta \end{bmatrix},$$

$$\Psi_{2c)}^{DRS} \equiv \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\gamma(\frac{N^*}{M}) \\ \frac{C^{j*}}{1+N^*} \\ -\theta \end{bmatrix}.$$

As we have seen from the conducted Eigenvalue analysis, all fixed points are non-hyperbolic. Therefore, the Hartman-Grobman theorem states, that we cannot rely on the linearized system to prove local stability of system 3.17. More important than the local stability analysis, is the proof of global stability. For a two-dimensional system, several methods for the proof of global stability can be found in the literature[22]. One often used method, providing sufficient conditions for the global stability of differential systems, is Dulac's criterion. Unfortunately, the application is restricted to two-dimensional systems. However, there is a generalization of Dulac´s criterion to three and higher dimensions in some special cases[23]. If possible, the dimension of the dynamic system could be reduced to two dimensions. For system 3.17, this is only the case if we assume CRS. Only for this case, the cost function can be described as an autonomous differential equation. For the IRS and DRS case, a reduction could be obtained by a possible aggregation of the cost and market structure. Consequently, the real dynamics of the four-dimensional system is then a perturbation of the aggregated model. If the aggregated model is structurally stable, then the dynamic behaviour of the aggregated model provides a good impression of the behaviour of the trajectories of model 3.17[24]. Hence, the perturbation method provides only an approximation of the model dynamics. The major drawback applying the perturbation method is the loss of the valuable feature of heterogeneity of model 3.17. To avoid this problem, (Noailly, 2008) suggests simulation methods, which should be used to obtain an impression of the long run behaviour of a complex system, due to the loss of the analytical tractability of system 3.17. This method is conform to the tradition of evolutionary economics. Therefore, in the next section, a simulation study will be conducted to obtain an intuition of the long-run behaviour of system 3.17.

[22]Refer to (Strogatz, 1994) for example.

[23]Refer to (Li, 1996) for this topic. The drawback of applying Dulac´s criterion is to find an appropriate Dulac function.

[24]Refer to (Mchich et al., 2007) for the application of this method for a predator-prey system.

3.2.4 Simulation study of the basic model

This section provides a simulation study of the before introduced basic model 3.17. The aim of this simulation study is to highlight market structure development, approximated via market share evolution which endogenously depends on firm size and innovation. For this reason, three different scenarios are simulated with respect to different returns to scale assumption. As mentioned before, decreasing returns to scale are often associated with small firms, whereas increasing returns to scale are associated with large firms induced by learning-by-doing, which takes place within each firm. For every scenario and with the exception of the decreasing returns to scale scenario, three different technological progress regimes are proposed to represent firm specific innovative activity which appears as a cost reduction technology. For an overview of simulation scenarios refer to table 3.2 in appendix 2.

The model proposes heterogeneous firms or agents with respect to their innovative activity. Hence, they are confronted with different cost regimes. To keep the simulation study simple, the simulation is restricted to the duopoly case, thus only the market share evolution of two firms, $i = \{1, 2\}$ are investigated and reported following. It is assumed that firm $i = 1$ is more efficient acquire new resources for production process than firm $i = 2$, but this induces higher production costs for firm $i = 1$. The market share at $t = 0$ is set to $s^1(0) = 0.6$ for firm $i = 1$ and thus $1 - s^1(0) = s^2(0) = 0.4$ for firm $i = 2$. The resource stock is set to $M = 50$ and the harvest at $t = 0$ is set to $N_0 = 50$.

Table 3.3 in appendix 3 provides an overview of parameter setting for simulation purpose. The parameter values are chosen accordingly to the work of (Noailly et al., 2003).

What can we expect intuitively regarding to our simulation study? Independently of what returns to scale scenario is assumed, the firm that offers the lower cost regime right from the beginning will remain in the market with certainty. The interesting question is, what will happen with the firm, which offers a higher cost regime, with respect to its market survival? Well, it depends on θ, which indicates the speed of technological progress. The faster technological progress is, the faster costs fall, hence the long run cost convergence overwhelms the feedback dynamics, which cannot deploy its entire force which leads to an elimination of laggard firms. Thus, the probability to obtain a monopoly scenario tends to one, if technological progress is very low.

But what changes, if different regimes of returns to scale are incorporated. As mentioned above, DRS are associated with the "Schumpeter-Mark-I" hypothesis, IRS with the "Schumpeter-Mark-II" hypothesis. If IRS are assumed, than selection mechanism and innovative performance boost each other. This is due to the fact that learning-by-doing within large firms leads to a shake out of laggard firms, which leads to a monopoly structure. On contrast, if DRS are taken into consideration, then the selection mechanism and innovative performance exhibits reverse effects, because of the idea, that small firms are more innovative than large firms. Of course, this could lead to some early market turbulences. For instance, if one firm gains a market share disadvantage its costs begin to fall at a faster rate, which leads to a market share advantage compared to a firm which exhibits market share advantage at the same time. Thus an overtaking occurs. This switching pattern behaviour with respect to market structure can theoretically recur several times until a stable pattern appears in the long run.

Simulation has been conducted with `Mathematica 5.2.0.0` and further with `Anylogic 5.5` to robustify simulation results. `Mathematica 5.2.0.0` offers a non linear solving routine[25], as well as `Anylogic 5.5`[26] does.

The figures 3.5, 3.6 and 3.7 in appendix 4 provide an overview of simulation results for the CRS, IRS and DRS case each. For every returns to scale scenario, the impulse responses for the market share s^i , per unit production costs c^i and resource extraction N_t for a low, middle and high technological progress regime are depicted. Firm $i = 1$ is characterized with a red colour, whereas for firm $i = 2$ green colour is alloted. A summary of the results based on the simulation study are given below.

1. Increasing returns to scale
   If we turn back to figure 3.6 we observe the following. As mentioned above, an increase in market share of the leading firm leads to a further increase in cost reduction, for instance by learning-by-doing. This "success-breeds-success"[27] scenario is reflected also in figure 3.6. A slow economic progress causes the expected monopoly scenario. Although the market share of firm $i = 1$ overwhelms right at the beginning of the simulation the market share of $i = 2$, after a few periods firm $i = 2$ will become monopolist, which induces that firm $i = 1$ has to leave the market. The faster technological progress is, the more realistic is a duopoly scenario. With other words, a fast rate of cost reduction outweights the positive feedback of leading firm and the cost convergence. This result confirms the "Schumpeter-Mark-II" hypothesis. Further it confirms the empirical thesis, that laggard industries tend to be more concentrated as mentioned by (Mazzucato, 2000).

2. Constant returns to scale

[25] Please refer to the `Mathematica 5.2.0.0` package `NDsolve`.
[26] `Euler`, `RK4`, `RK45`, `RADAUS` etc. for differential equations.
[27] Refer to (Phillips, 1971).

The case of CRS is also pictured in figure 3.5. As one can see, there are no new insights compared to the IRS case. Hence, for the CRS, the implications derived from the IRS case still holds.

3. Decreasing returns to scale
   The DRS case is graphically replicated with figure 3.7. For a low value of technological progress, we again observe a monopoly tendency in the market. The less cost efficient firm $i = 1$ is shaked out of the market. An intermediate level of technological progress instead would lead to a coexistence scenario, whereas the fittest firm is, as in the IRS and CRS case, the market leader. Therefore, the higher θ is, the less concentrated is the market. But for a moderate level of technological progress, it can be shown, as done in figure 3.7, that market turbulence cannot be ruled out in the case of DRS. Firms with higher market shares are confronted with slower rates of cost reduction and thus they have been surpassed by smaller firms in terms of cost efficiency. The switching behaviour lingers until cost convergence has reached. The key point is, that instead of IRS and CRS cases, the prediction of the so called final ranking of firms is no possible. The latter unstable market structure observation, which replicates a stylized fact of firm-size dynamics, has been empirically confirmed by a bulk of studies[28].

3.3 Extension of the basic model: Learning and Knowledge Diffusion

A certain limitation of the basic model is that it does not include the possibility of learning and knowledge diffusion. As mentioned above, by (Campagni, 1991) for instance, inter-firm cooperation based on knowledge sharing can explain the predominance of small firms in the market. One implication of the basic model is that small firms will be shacked out of the market for certain parameter constellations, as seen above. Hence, it seems to be logical, to incorporate the aspect of learning and knowledge diffusion in the basic setup.

Learning in the model context is kept rather simple. It is assumed, that a firm i which exhibits an inferior cost structure with respect to a firm j may benefit from spillovers, which are send from the technological leading firm j. In the best case, the underperforming firm i will benefit from the entire pool of spillover, generated by firm j. If this rather old style neoclassical assumption is made, learning is more or less senseless. For this reason, it is assumed that an underperforming region i has an

[28]Please refer to (Mazzucato, 2000), p. 49 for an overview.

inherent incentive to learn, because knowledge is specific and cannot be understood right from receiving. For this reason, spillovers have to be integrated in the model before talking about how to model learning activities.

3.3.1 Integration of knowledge spillovers

To model the spillovers from using a strategy h, which is inferior in a strategy j to certain period of time t, to a common notational form of the so called knowledge gab literature is referred. To motivate the technological gap, it is assumed for the moment, that a firm h has to choose from a given pool I a cost reduction technology. Hence, every firm h can be described by a different level of technology $T(h)$. The possible spillover pool is then defined as the gap between different technological levels $T(i)$. Keeping this in mind, the spillover Γ_{hj} from h to j can be written as:

$$\Gamma_{hj} = \ln\left(\frac{T_h}{T_j}\right) \tag{3.28}$$

or to get an impression regarding the evolution of the spillover:

$$\dot{\Gamma}_{hj} = \hat{T}_h - \hat{T}_j, \tag{3.29}$$

with $\hat{T}_i$ as the growth rate of technology $T(i)$. As one can see from equation 3.28, the greater Γ_{ij} the greater is the technological heterogeneity which means the greater the possible spillover pool which can be used by j et vice versa.

Next Γ_{hj} has to be specified. First, we have to think about an explicit specification of $T(i)$. $T(i)$ in the basic model can be approximated with the cost structure. In the easiest way, we can assume, that costs C^h compared to the higher costs in the market, given by $\tilde{C} \equiv C_j + C_{Market,fix}$ defines the technological gap from which firm j may benefit. Please note, that $C_{Market,fix}$ cannot be reduced by innovative activity but it is an exogenous number.

Second, we have to assume, that only a $\phi \in (0,1)$ fraction of the spilloverpool Γ_{hj} can be understood. One assumption could be, to say, that this fraction $\phi \in (0,1)$ is constant over time. This implies, that a firm cannot learn during its spillover benefiting phase. Surely, this scenario is possible. But it should be treated as a special case of learning behaviour instead of assuming, that this holds on general, as assumed by (Cantner and Hanusch, 1998) for instance. Because of this reason, $\phi_t^h, \in (0,1)$ is time dependent and labeled as the degree competence, which can be influenced by learning activities. Thus

$\phi_t^h, \in (0,1)$ itself is an endogenous number which influences the endogenous spillover function Γ_{hj}.

Keeping this two aspects in mind, the spillover function Γ_{hj} can be rewritten as follows:

$$\Gamma_{hj} = \begin{cases} 0, & C_h \geq \tilde{C} \\ \phi_t^j \ln(\frac{C_h}{\tilde{C}}), & C_h < \tilde{C}, \quad \phi_t^h \in [0,1] \end{cases}. \tag{3.30}$$

Of course one can expand the formulation 3.30 in terms of integrating of so called "absorptive capacities" as done by (Verspagen, 1992a) and (Verspagen, 1992b). The next section deals with the specification of $\phi_t^h, \in (0,1)$.

3.3.2 Integration of learning aspects

Learning aspects have been widely discussed in traditional learning curve literature (Yelle, 1979) and were introduced by (Wright, 1936). To the best to my knowledge, until today there is no formulation which combines aspects from the psychological motivated learning curve literature and technology gap literature. From this point of view this is rather remarkable because understanding knowledge requires the correct cognitive structure to make sense of a particular piece of useful information[29].

As mentioned above, the learning curve concept referred to in this context, is based on the ideas of models of time allocation[30], because competence evolvement and time allocation are closely related.

From the relevant psychological studies, two learning curve concepts have been prevailed. First, a concave learning curve which covers the fact of diminishing returns of time investment for learning. Second, a learning curve, which is used in the theoretical work of (Hull, 1943), (Van Gert, 1991) and (Newell et al., 2001) et al.. The latter concept assumes a logistic learning curve. The idea behind the logistic learning curve is, that learning at the beginning is slow, than the learning progress is increasing rapidly in the middle and is slabbing at the end of the learning progress.

The question now is, which learning curve covers which learning behaviour? Logistic or S-shaped learning curves are often used to map a complex skill learning event, learning a language or complex avenues of approach for instance. Empirical evidence of

[29]Refer to (Nooteboom, 1992) and (Nooteboom, 1999).
[30]Refer to (Metcalfe, 2002), (Nelson and Narens, 1990) and (Nelson and Narens, 1994).

S-Shaped learning curves are found by (Rice et al., 1998) who outlined that "inspection of the individual curves shows slow growth at the beginning [...] followed by rapid acceleration, and then a final period of leveling off"[31]. Additionally (Frey and Sears, 1978) have mentioned that curves in conditioning "are typically S-shaped, with a period of positive acceleration followed by one of negative acceleration"[32]. An exponential or concave learning curve is associated to a rather ideal learning process. Thus, under a regime of an exponential learning curve the learning subject can be described as less complex regarding to a regime of a S-shaped learning curve. In this work, the sigmoid learning curve concept has been considered.

Figure 3.1 provides a sketch of the idea, which lies behind the sigmoid learning curve concept. Consider for the moment A is constitute by a degree of competence c_A and a given time t_A which denotes the time reaching the competence level c_A. In this way a higher degree of competence c_A can be reached simply by allocate more time to learning activities, because $t_A < t_B$.

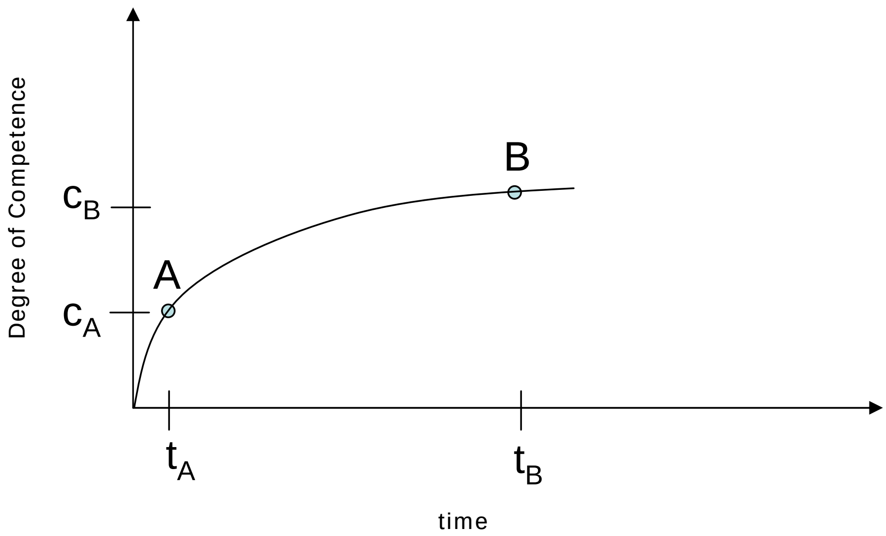

Figure 3.1: One-dimensional sigmoid learning curve

Of course this one-dimensional learning curve concept is too simple to cover the complex aspect of learning. To give other facts which influence the learning activity consideration, the aspect of talent is also included in the learning curve concept. But aspects such as talent are rather difficult to implement directly in a learning curve environment. For this reason, a proxy for the vector of unobservables characteristics is required. The proxy should exhibit the feature that it replicates the fact that a person

[31](Rice et al., 1998), p. 1425.
[32](Frey and Sears, 1978), p. 324.

should reach a higher degree of competence by spending the same time for learning activity compared to another, perhaps the first is more talented than the latter.

Given, every firm i is endowed with a different knowledge stock χ_i and its unobservable learning characteristics are embodied by ι_i, then the degree of competence ϕ_t^i based on a sigmoid multidimensional learning curve can be written as:

$$\phi_t^i \equiv \varpi(\chi_i, \iota_i, t) = \frac{1}{1 + \exp[\nu_i - \iota_i t]}, \tag{3.31}$$

with $\nu_i \equiv \frac{1-\chi_i}{\chi_i}$.

Equation 3.31 exhibits moreover the desired attribute that a higher endowment χ_i leads to an earlier start of the learning process. Thus, in terms of probabilities, we can say that the probability, to reach a degree of competence of one is given by:

$$P[\phi_t^i = 1] = \varpi(\chi_i, \iota_i, t) = \frac{1}{1 + \exp[\nu_i - \iota_i t]}. \tag{3.32}$$

Figure 3.2 provides a graphical representation of equation 3.31 for given χ_i, whereas figure 3.3 represents 3.31 for a given value of ι_i.

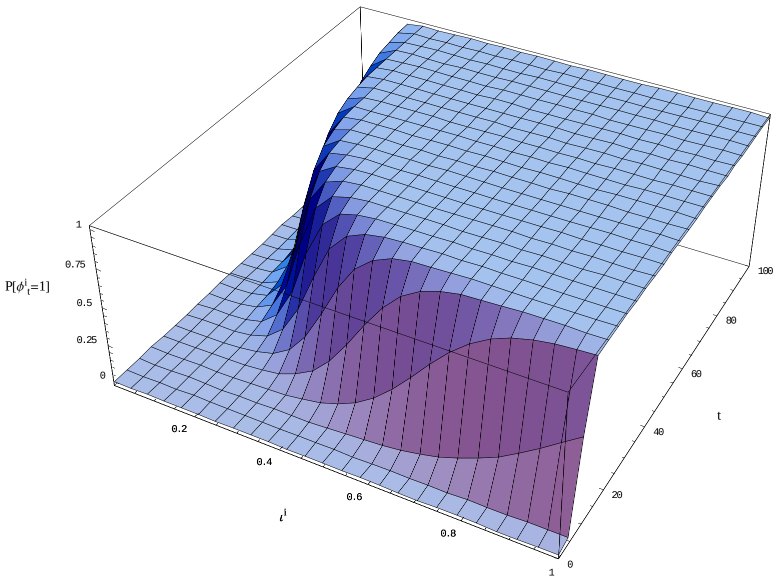

Figure 3.2: $P[\phi_t^i = 1]$ expressed by ι_i and t

Further, it is assumed that idosyncratic learning activities should be treated as a stochastic event rather than to assume that learning endeavor is deterministic. In this context, the idiosyncratic learning curves may fluctuate around the deterministic learning curve. Hence, the stochastic version of 3.31 is given by:

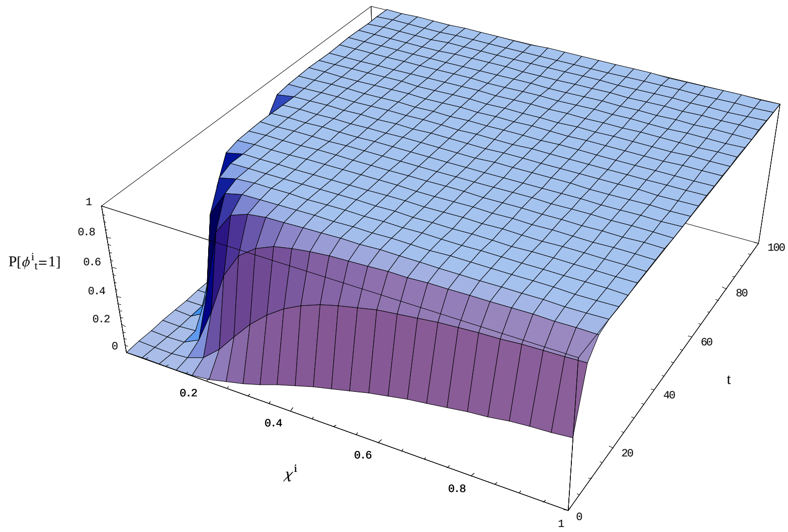

Figure 3.3: $P[\phi_t^i = 1]$ expressed by χ_i and t

$$\phi_t^i \equiv \varpi(\chi_i) = \frac{1}{1 + \exp[\nu_i - (\iota_i + \zeta_i)t + \epsilon_i]},\ \epsilon_i \sim (0, \sigma_\phi^2),\ \zeta_i \sim (0, \sigma_\zeta^2), \tag{3.33}$$

with $\nu_i \equiv \frac{1-\chi_i}{\chi_i}$.

Figure 3.4 provides some realizations of the stochastic version of 3.31. As one can see, the uncertainty of learning progress is largest around the inflection point of 3.31.

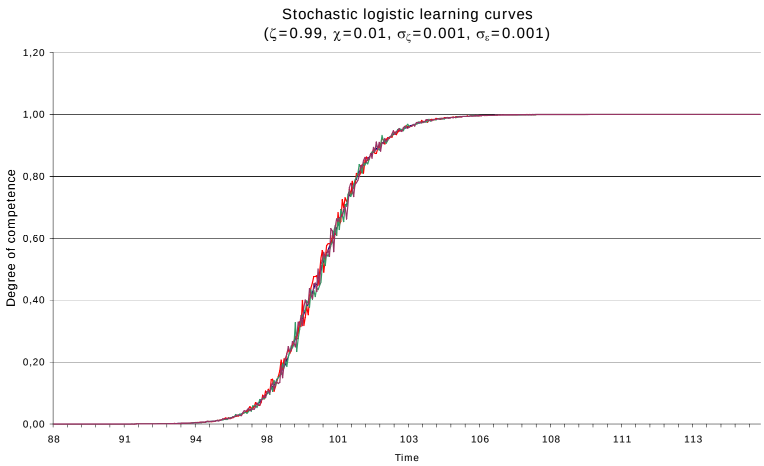

Figure 3.4: Some realizations of the stochastic sigmoid learning curve

After motivating the learning curve concept, the basic model 3.17 can be expanded as follows:

$$\begin{cases} \dot{N} & = & \gamma N\left[1-\frac{N}{K}\right]-q\beta\left[\sum_N s_i F_i(N)\right] \\ \dot{s}_i & = & s_i(c_i-\bar{c}) \\ \dot{C}_i & = & \begin{cases} -\theta C_i-\Gamma_{ij} \\ -\theta C_i s_i-\Gamma_{ij} \\ -\theta C_i(1-s_i)-\Gamma_{ij} \end{cases} \end{cases} . \tag{3.34}$$

As one can easily see, the extension of model 3.17 cannot be solved analytical. Even, a conventional stability analysis cannot be conducted due to the inherently high complexity of this model. But it is possible to derive a steady state condition for this model.

Proposition 6: Given proposition 1 holds than the partial derivatives of $\dot{N}_t$, $\dot{s}_t$ and $\dot{C}_t$ still exist. Provided $\dot{N}_t = \dot{s}_t = \dot{C}_t = 0$ holds simultaneously than system 3.34 has an unique steady state vector S which contains N^*, s_i^* and C_i^* in the long run. Thus

$$\begin{cases} \xi N^*\left[1-\frac{N^*}{M}\right] & = & \psi\beta\left[\sum_I s_i^* F_i(N^*)\right] \\ 0 & = & s_i^*(c_i^*-\bar{c}) \\ 0 & = & \begin{cases} -\theta C_i^*-\Gamma_{ij} \\ -\theta C_i^* s_i^*-\Gamma_{ij} \\ -\theta C_i^*(1-s_i^*)-\Gamma_{ij} \end{cases} \end{cases} \tag{3.35}$$

must hold.□

We know, that a steady state must exist. The steady state is reached, if cost convergence of both firms has been occurred, thus $\Gamma_{ij} = 0$ is realized. Then we are automatically back to model 3.17, from which we know, that a steady state exists. The next section deals with the simulation of the extended version of model 3.17, model 3.34.

3.3.3 Simulation study of the extended model

The simulation setup for model 3.34 is the same as for model 3.17. Thus, it is referred to the same parameter setting as in the simulation of model 3.17. The learning curve parameter have been chosen as follows:

The simulation study has been conducted with Anylogic. To avoid redundancies with respect to simulation results discussion, in the following it is referred mainly on

Parameter	Value
χ	0.50
ι	0.50
σ_ζ^2	1×10^{-6}
σ_ϵ^2	1×10^{-6}

Table 3.1: Learning curve parameter setting

simulation meanderings induced by integration of learning aspects. As done before, three simulation scenarios have been performed, for the DRS, CRS and the IRS case. For the parameter setting of the technological progress it is therefore referred to table 3.2.

With respect to the DRS, CRS, and IRS the inclusion of learning aspects leads to the following observations, based on the simulation study. Again, one can find the impluse responses for the simulation of the extended model in appendix 5, 6 and 7. The CRS case is depicted in appendix 7 in figures 3.15, 3.16 and 3.17. The IRS case is graphically replicated in appendix 6 with the corresponding figures 3.12, 3.13 and 3.14, whereas the DRS case can be found in appendix 5 in figures 3.8, 3.9, 3.10 and 3.11.

1. Increasing returns to scale
   As argued before, the feedback of market selection and the feedback of cost convergence is weightout by inducing a high rate of technological progress with respect to cost reduction. On the other side, the lower θ the more weight is laid to the positive feedback of innovative activity and thus the more cost efficiency firm is crowding out the laggard firm. Now we have to account for a third effect: the positive spillover effect which is driven by learning activities for the laggard firm. With respect to the simulation results, on the first sight, there is no significant changing with respect to the simulation results based on model 3.17. But if we look more closely, then we observe, that for no feasible parameter constellation of θ the before mentioned monopoly scenario occurs. Soonest this is the case for a small value of technological progress, because the herfindahl takes its largest value for this case. Hence, the higher θ, the lower the herfindahl index HI[33], and thus the more realistic a coexistence of both firms is. For the case of $\theta = 0.5$ a nearly uniformly market segmentation is obtained with a corresponding herfindahl index of $HI = 0.50$. An interesting results is obtained for a high speed of technological progress, $\theta = 0.99$. In this case without learning effects, the convergence effects outweighs the selection effects, because technological progress is so fast that the

[33] The herfindahl index HI is defined as $HI := \sum_i s_i^2$.

lowest cost level is reached until selection effects or diseconomics of scale have time to taken effect in the market. Finally, the less cost efficient firm $i = 2$ becomes market leader, because the high speed of technological progress leads to a temporal cost leadership of firm $i = 2$ in a negative sense until cost convergence has reached. Now positive learning effects from which exclusively the laggard firms can benefit, lead to a more turbulent market evolution in the beginning of the simulation study, as one can easily see from the stability index and herfindahl index. But at the end no significant difference can be observed compared to the "no learning" case. Hence, the inclusion of learning effects in the CRS scenario of model 3.17 results in no significant changes regarding to the simulation results, with the exception that no monopoly scenario occurs. This could be due to the smoothing effects induced by the knowledge spillovers.

2. Constant returns to scale
   As for the IRS case, learning effects to not exhibit a significant change on simulation results, compared with the case of CRS for model 3.17. But, also for the IRS case mentioned above, no smoothing effects of knowledge spillovers lead to the exclusion of monopoly scenarios.

3. Decreasing returns to scale
   For the DRS scenario, we do not observe any significant changes regarding to the DRS case of model 3.17, with two exceptions. First, as mentioned before for the IRS and CRS case respectively, no monopoly scenario occurs for any reasonable value of θ. For any given parameter value of θ, firm $i = 2$ will remain the market leader at the end. Second, and the more interesting is the special case of $\theta = 0.02$. As argued before, the weak negative feedback causes a slower rate of cost reduction for the leading firm and thus will lead to a surpass by smaller firms. Thus a switching occurs until cost convergence is realized. This mentioned market instability leads to the conclusion, that a final ranking of firms cannot be predicted. If we now integrate learning behaviour we still observe market instability, which is distinct at most with respect to time. But again, the smoothing effects of spillover lead to a more stable structure. No switching phenomena is observed and thus and on contrary as before, the prediction of a final firm ranking seems to be more feasible.

On summary, we can conclude, that integrating learning aspects leads to a more stable market structure at all. Further, this model supports the empirical finding by (Campagni, 1991), which states that inter-firm cooperation based on knowledge sharing can explain the predominance of small firms in the market. This conclusion can be drawn because of the fact that for no simulation scenario, a monopoly market structure occurs. It is worth to mention, that unreported simulation studies, which are based on a "constant learning" scenario, reveals that for low speed of technological progress monopolistic market structures occur. Hence, only the inclusion of de facto learning effects leads to an exclusion of monopoly market structures.

3.4 Summary

The early stages of an industry life cycle are characterized by instability and a relatively competitive market environment. This awareness is also labeled in the relevant literature as a stylized fact regarding firm-size dynamics. (Mazzucato, 2000) has shown in a simulation study, based on the replicator dynamics approach, that in fact the before mentioned stylized fact can be replicated by the model assuming decreasing returns to scale which corresponds to the "Schumpeter-Mark-I"' hypothesis. The argument of this pretty simple and easy to retrace: firms with a high market share would expire a slower rate of cost reduction potential and thus those firms will be lurched by smaller firms. This process leads to a switching behaviour of market structure, particularly at the beginning of the life cycle of an industry. For other parameter constellations, small firms will be shaked out of the market, especially when increasing returns to scale are assumed.

The lack of the model (Mazzucato, 2000) is that it is not able to cover the fact that a high knowledge transfer intensity, for instance due to cooperations which are based on knowledge transfer, enhances firm innovativeness and hence induces a feedback on market structure. As mentioned by (Campagni, 1991), (Best, 2001), (Porter, 2000) and (Krugman, 1991) in a more spatial context, the exchange of ideas and knowledge lead to a predominance of small firms in the market. For this reason, the question arises how affects learning behaviour the market structure.

For this reason, a model similar to the work of (Mazzucato, 2000) and (Noailly et al., 2003) is setup which is able to replicate the before mentioned stylized fact. After a simulation has been conducted the model was extended by learning effects and knowledge diffusion. Knowledge diffusion in this context is treated as an endogenous event, driven by psychologically motivated learning endeavours of firms. A further simulation study of the extended model has shown, that for any degree of technological progress small firms still remain in the market, also for the case of IRS, where large firms are in advance. Hence, this model is able to replicate the fact, that small firms are more likely to benefit from knowledge networks and thus from spillovers which define a source of innovativeness, from which large firms cannot profit.

Of course, there are several revenues for further research. For instance, it is planned to embed this rather simple model in a spatial model framework to cover explicit cluster effects of small firms.

3.5 Appendix

3.5.1 Appendix 1

To linearize a non linear system, as a first step, it is common to write for a n-dimensional non linear system in general:

$$\dot{\mathbf{x}} = F(\mathbf{x}_t, \mathbf{v}_t). \tag{3.36}$$

Hereby $\mathbf{F}(\cdot)$ is a $(n \times 1)$-dimensional vector containing n vectors $f_n(\cdot)$ of non linear functions, $\dot{\mathbf{x}}$ is a $(n \times 1)$- dimensional vector which contains the partial derivatives of x with respect to t and $\mathbf{v}_t$ is a $(n \times 1)$-vector of time dependent values. For our purpose I set $\mathbf{v}_t = \mathbf{0}$ without loss of generality. Therefore, equation 3.36 can be rewritten and thus one obtains[34]:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, ..., x_n) \\ f_2(x_1, x_2, ..., x_n) \\ \vdots \\ f_n(x_1, x_2, ..., x_n) \end{bmatrix}. \tag{3.37}$$

To discuss the dynamic behaviour of our system in a ϵ-neighbourhood of the steady state values $x^* = [x_1^*, x_2^*, ..., x_n^*]'$, we have to linearize our system around its steady state vector x^* using a Taylor expansion or approximation, respectively.

The intuition behind a first order Taylor expansion is to express the deviations Δ of the variables of interest $\mathbf{x}$ from their steady state values $\mathbf{x}^*$. Thus, we obtain:

$$\begin{aligned} \dot{x}_1 &= f^1(x^*) + f^1_{x_1}(x^*)(x_1 - x_1^*) + ... + f^1_{x_n}(x^*)(x_n - x_1^*) + \vartheta_1, \\ \vdots & \quad \vdots \quad \vdots \\ \dot{x}_n &= f^n(x^*) + f^n_{x_1}(x^*)(x_n - x_1^*) + ... + f^n_{x_n}(x^*)(x_n - x_n^*) + \vartheta_n. \end{aligned} \tag{3.38}$$

If $|x - x^*| \to \epsilon$ then $\vartheta_i = 0$, $i = \{1, 2, ..., n\}$. The advantage of the Taylor approximation in the neighbourhood of the steady state is that the first elements of every equation i vanish because of the existence of a steady state. The implication is that $\dot{x}_i = 0\ , \forall\, i$.

[34]I set the subscripts t in notational form to indicate time dependent variables.

In matrix algebra we can write:

$$\frac{\partial \Delta \mathbf{x}}{\partial t} = F(\mathbf{x}^* + \Delta \mathbf{x}). \tag{3.39}$$

Or, if we apply the Taylor expansion on $F(\cdot)$ around the steady state values $\mathbf{x}^*$, one can derive:

$$\frac{\partial \Delta \mathbf{x}}{\partial t} = F(\mathbf{x}^*) + \frac{\partial F(\cdot)}{\partial \mathbf{x}}|_{\mathbf{x}=\mathbf{x}^*}[\mathbf{x} - \mathbf{x}^*] + \vartheta(\Delta \mathbf{x}), \tag{3.40}$$

whereas the residual vector $\vartheta(\Delta \mathbf{x})$ can be treated as a redundant variable, as mentioned before. It is easy to see that we have to compute n-partial derivatives for each $f_i(\cdot)$ such we get at all together $n \times n$ derivatives for the matrix $\mathbf{F}(f_1, f_2, ..., f_n)$. Subsequently, we may concentrate our facts and write in matrix algebra in a more convenient manner:

$$\dot{\mathbf{x}} = A\overline{\mathbf{x}} \tag{3.41}$$

with $\overline{\mathbf{x}} \equiv \mathbf{x} - \mathbf{x}^*$. Here we define

$$A \equiv \left[\frac{\partial F(\cdot)}{\partial \mathbf{x}}|_{\mathbf{x}=\mathbf{x}^*}\right] = \begin{bmatrix} \frac{\partial \dot{x}_1}{\partial x_1} & \frac{\partial \dot{x}_1}{\partial x_2} & \cdots & \frac{\partial \dot{x}_1}{\partial x_n} \\ \frac{\partial \dot{x}_2}{\partial x_1} & \frac{\partial \dot{x}_2}{\partial x_2} & \cdots & \frac{\partial \dot{x}_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \dot{x}_n}{\partial x_1} & \frac{\partial \dot{x}_n}{\partial x_2} & \cdots & \frac{\partial \dot{x}_n}{\partial x_n} \end{bmatrix}. \tag{3.42}$$

One can verify that in expression 3.42 A is called the Jacobian of the system. Herewith, we have transformed a non linear system into a linearized non linear system, which is not homogeneous.

3.5.2 Appendix 2

Case I: constant returns to scale	Value
1. case: low technological progress	$\theta = 0.010$
2. case: middle technological progress	$\theta = 0.500$
3. case: high technologcial progress	$\theta = 0.999$
Case II: increasing returns to scale	**Value**
1. case: low technological progress	$\theta = 0.010$
2. case: middle technological progress	$\theta = 0.500$
3. case: high technological progress	$\theta = 0.999$
Case III: decreasing returns to scale	**Value**
1. case: low technological progress	$\theta = 0.0001$
2. case: relative low technological progress	$\theta = 0.0200$
3. case: middle technologcial progress	$\theta = 0.5000$
4. case: high technologcial progress	$\theta = 0.9999$

Table 3.2: Returns to scale scenarios for basic model simulation

3.5.3 Appendix 3

Parameter	Value
M	50
$s^1(0)$	0.6
N(0)	50
E^1	0.7
E^2	0.2
C^1	15
C^2	10
β	10
q	1
ζ	8
ψ	8
γ	1

Table 3.3: Parameter setting for basic model simulation

3.5.4 Appendix 4

Please refer to the next pages.

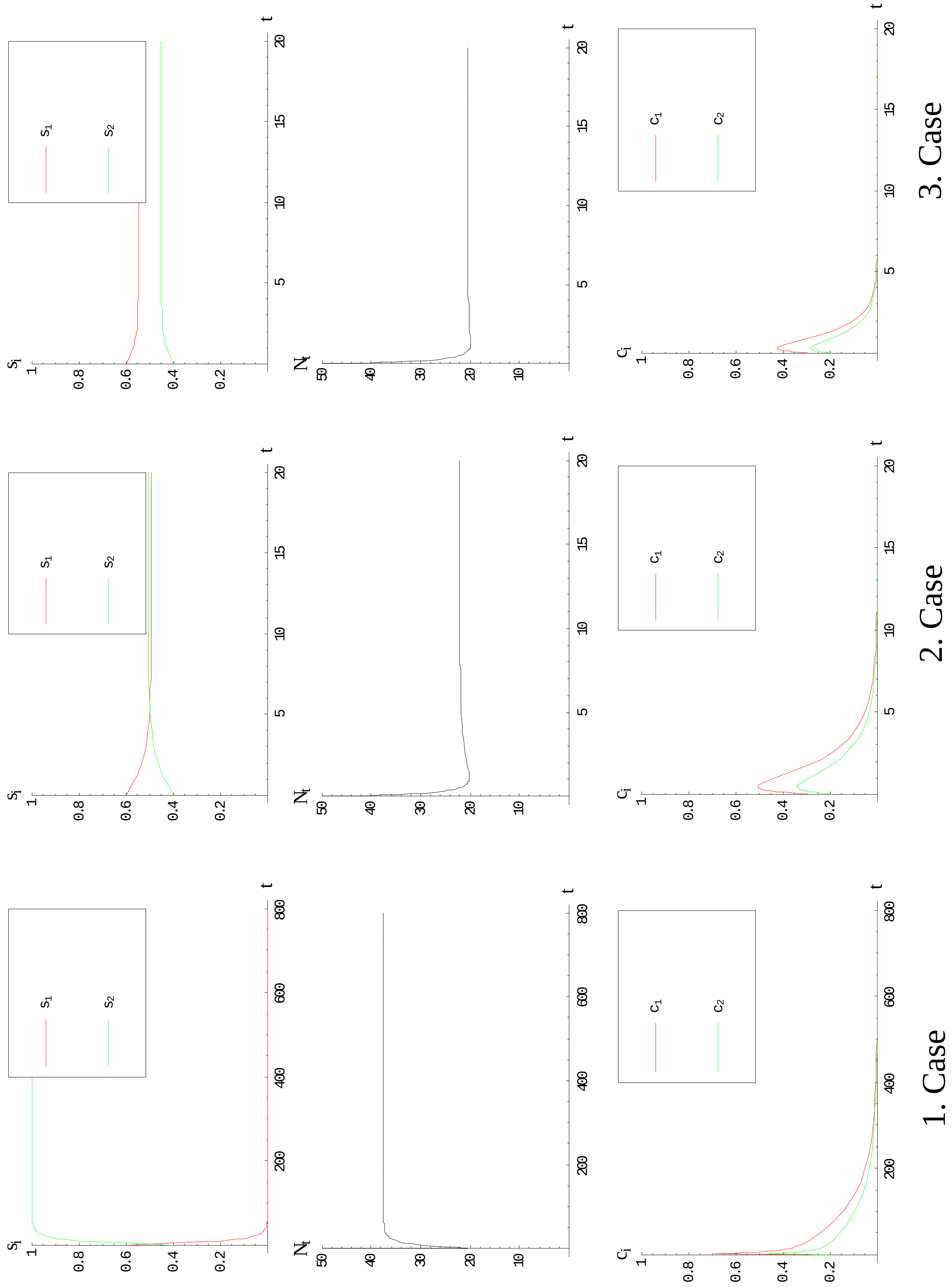

Figure 3.5: CRS impulse responses for the basic model

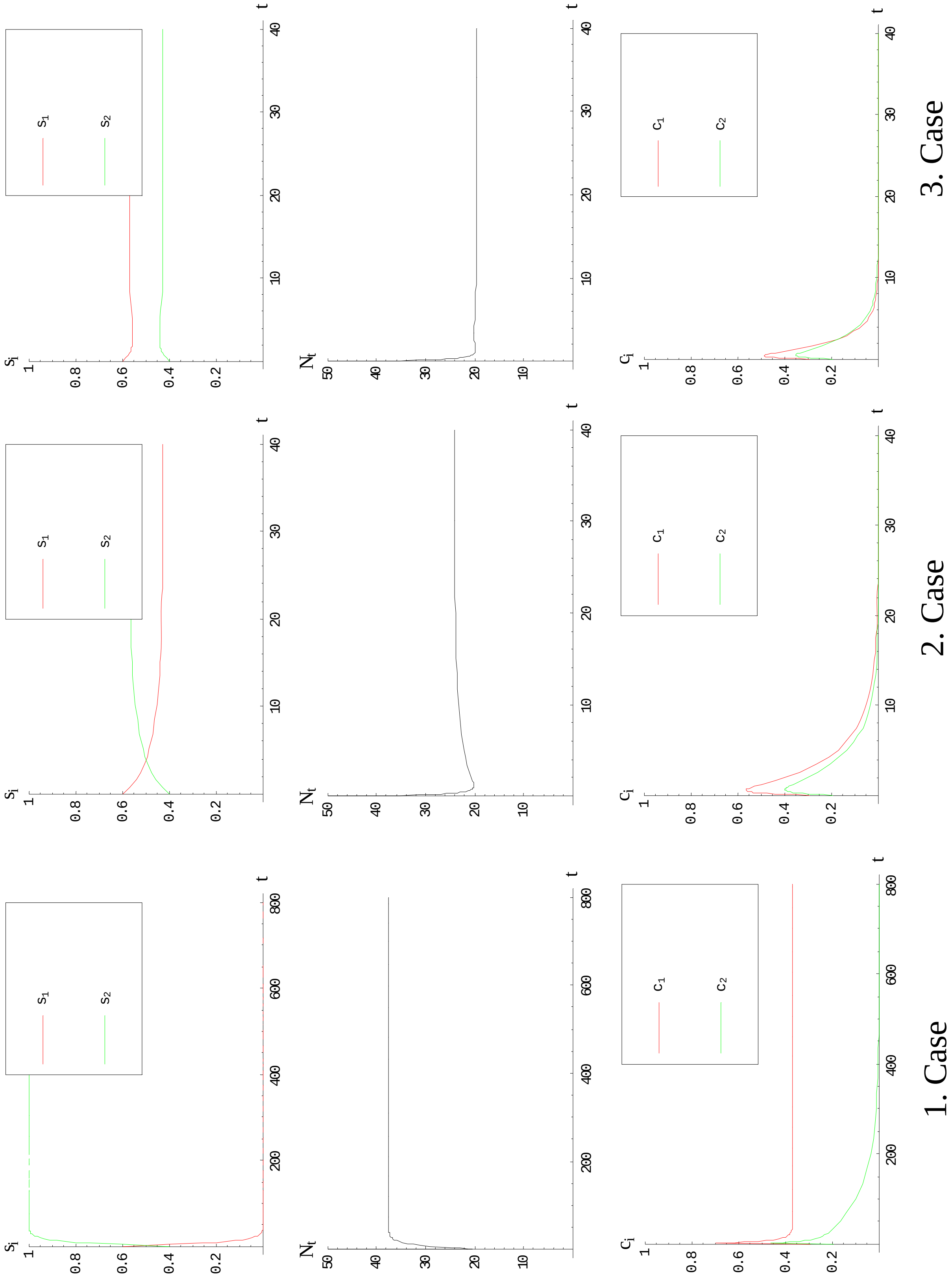

Figure 3.6: IRS impulse responses for the basic model

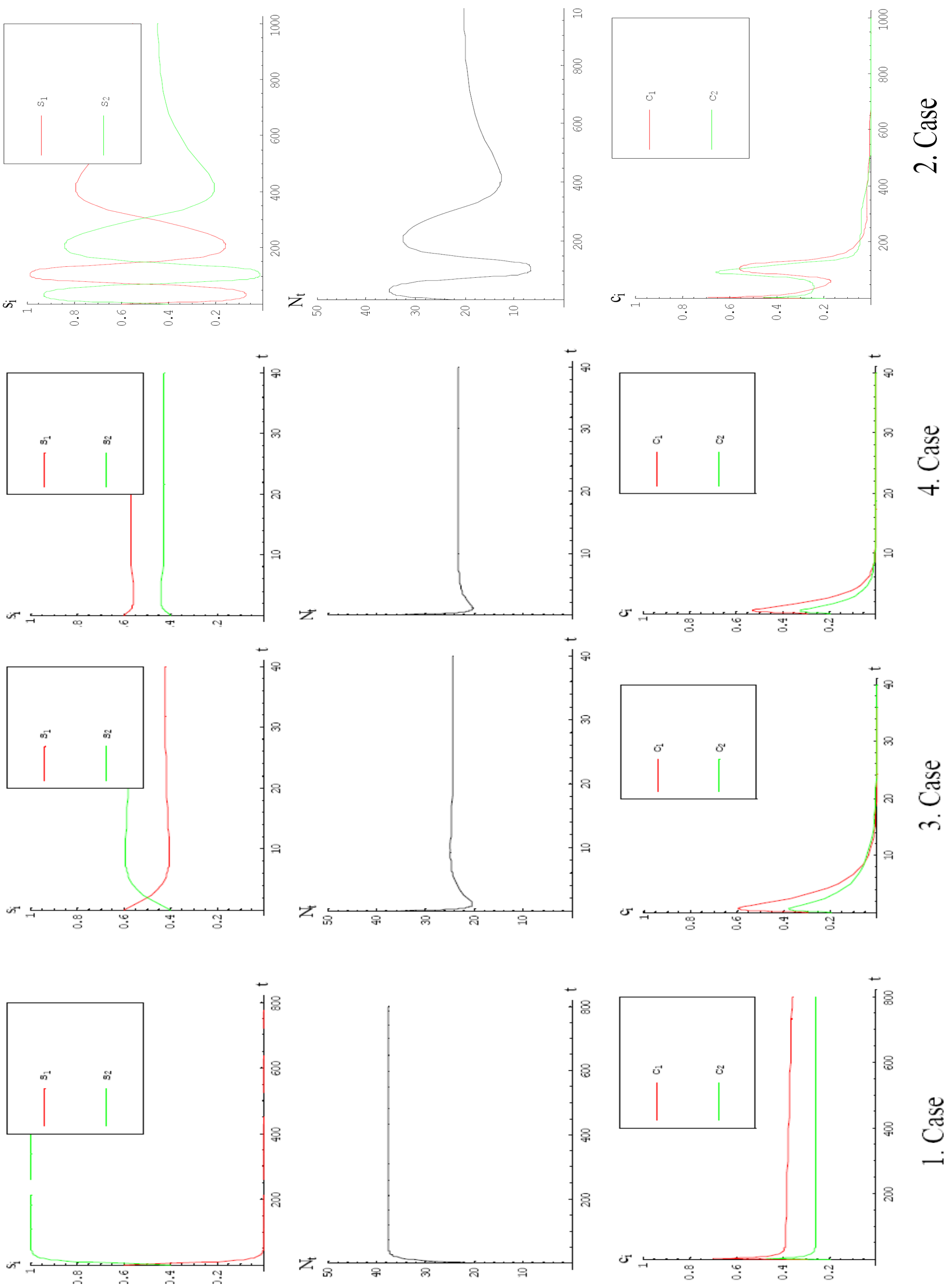

Figure 3.7: DRS impulse responses for the basic model

3.5.5 Appendix 5

Please refer to the next pages.

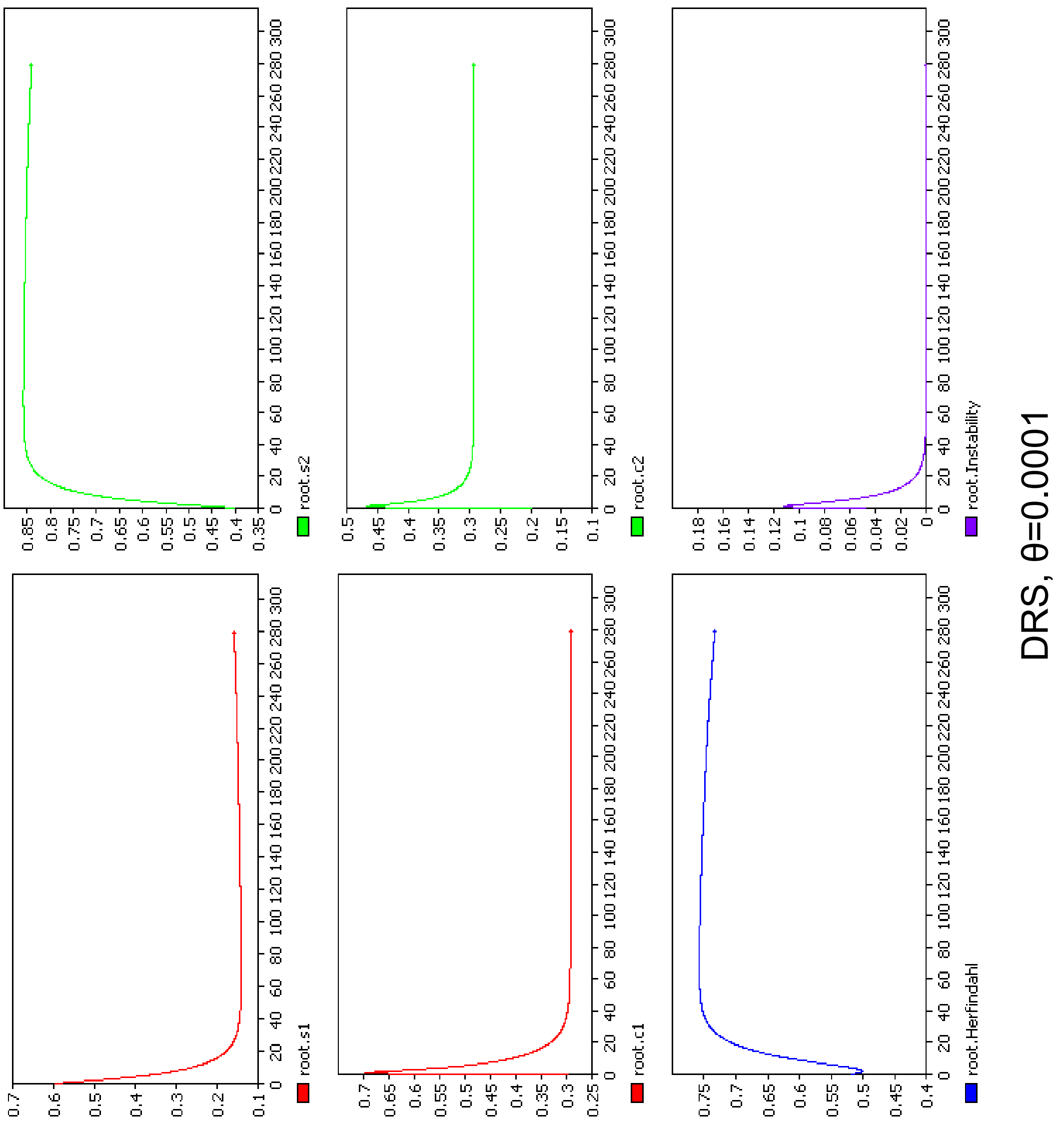

Figure 3.8: DRS impulse responses for the extended model

Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

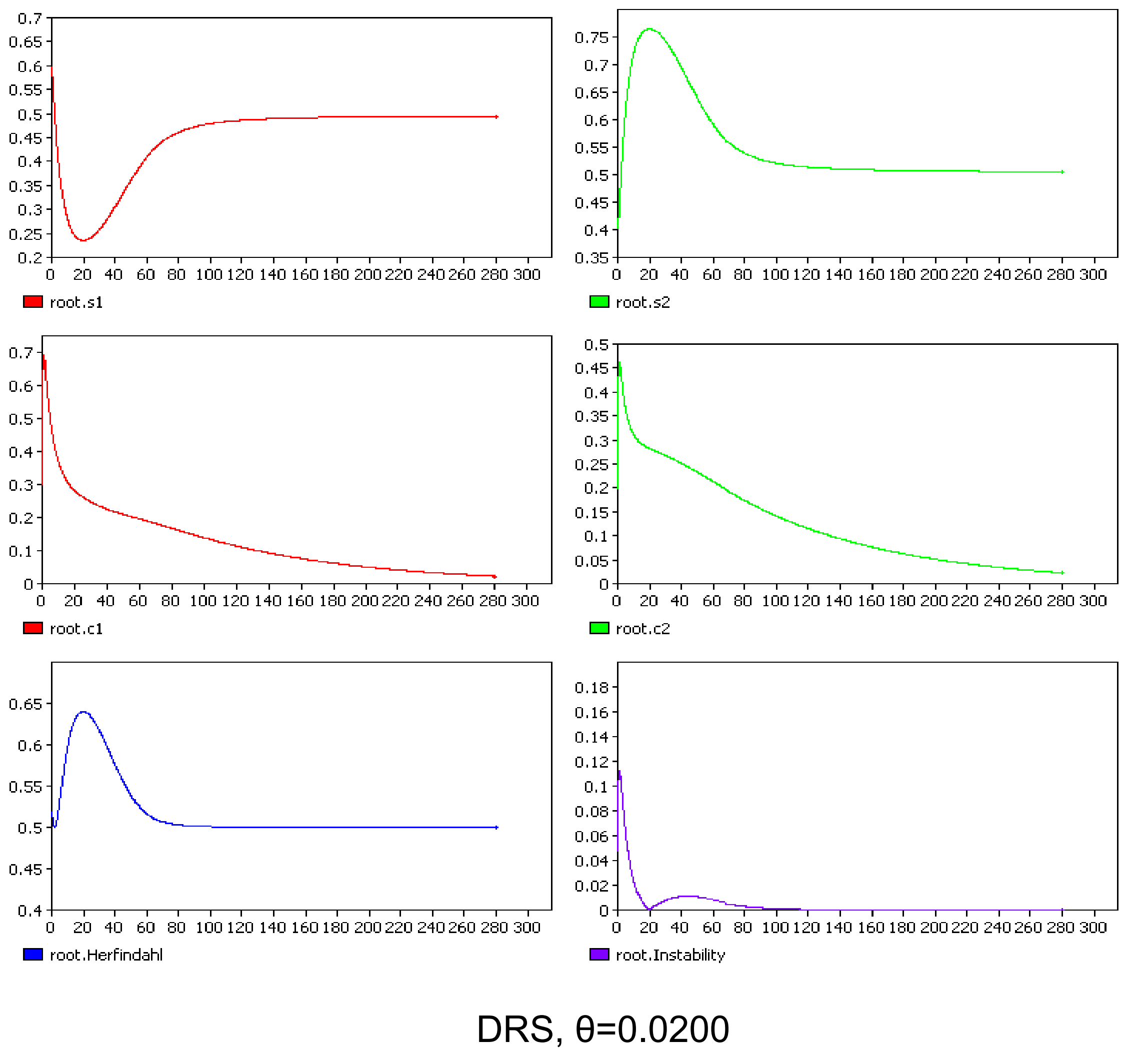

Figure 3.9: DRS impulse responses for the extended model

Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

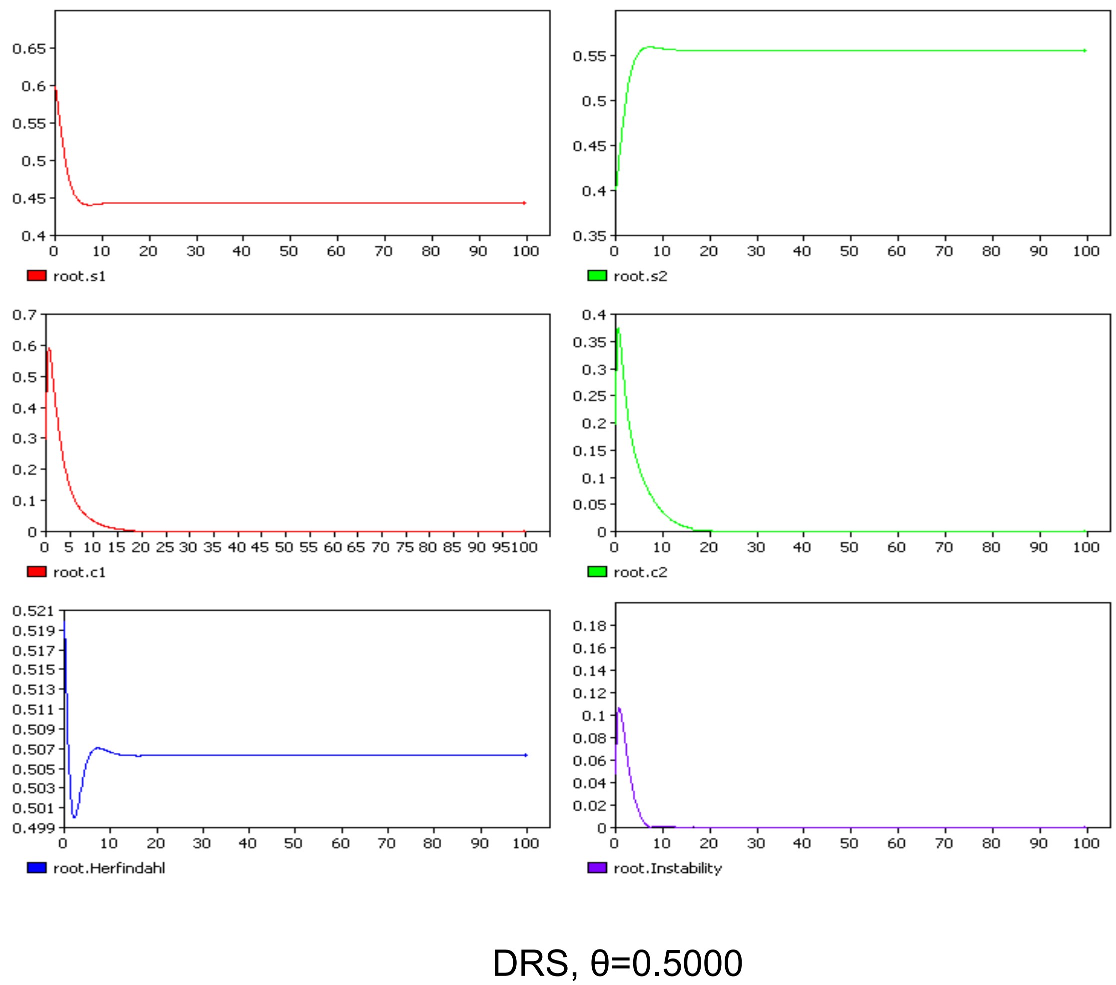

Figure 3.10: DRS impulse responses for the extended model
Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

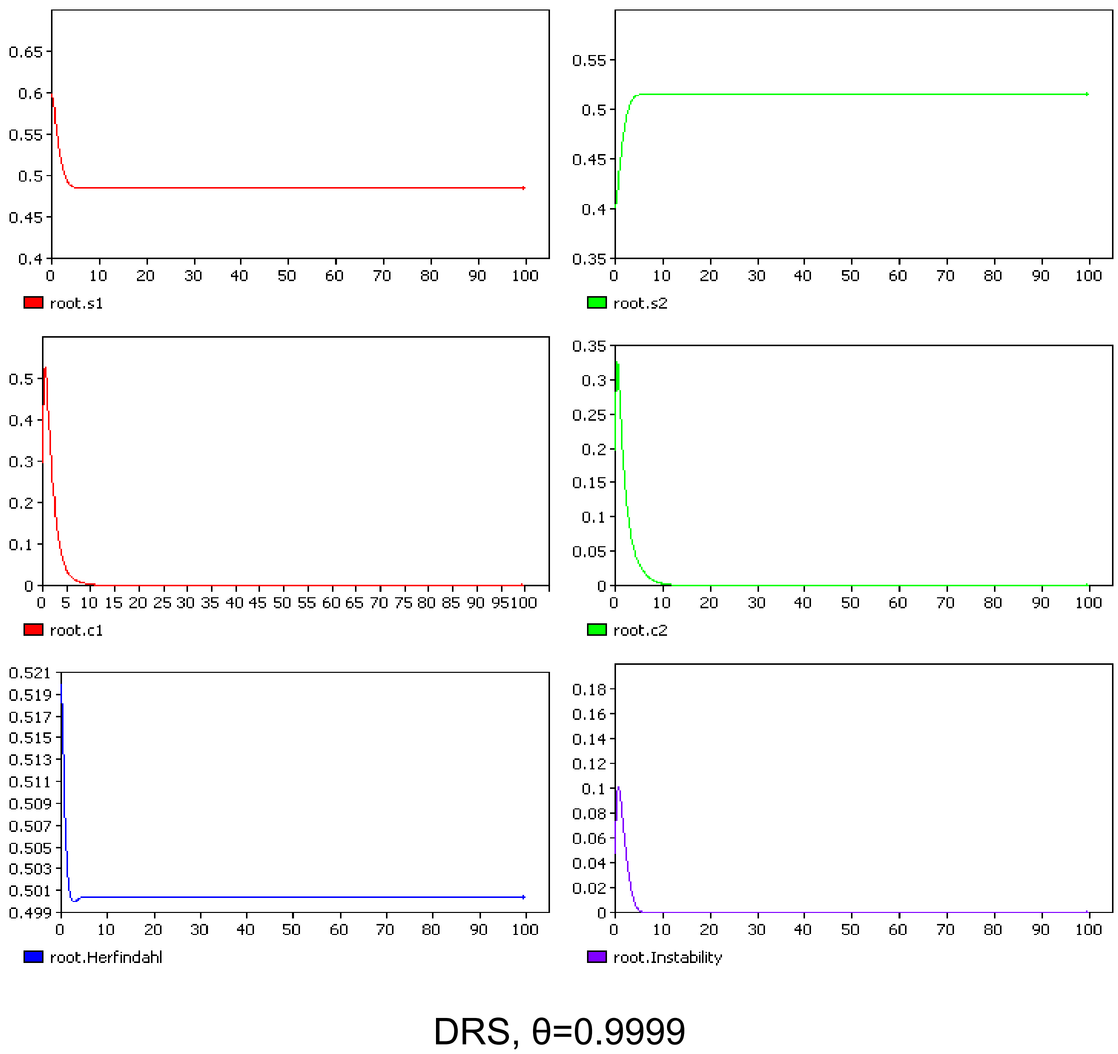

Figure 3.11: DRS impulse responses for the extended model
Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

3.5.6 Appendix 6

Please refer to the next pages.

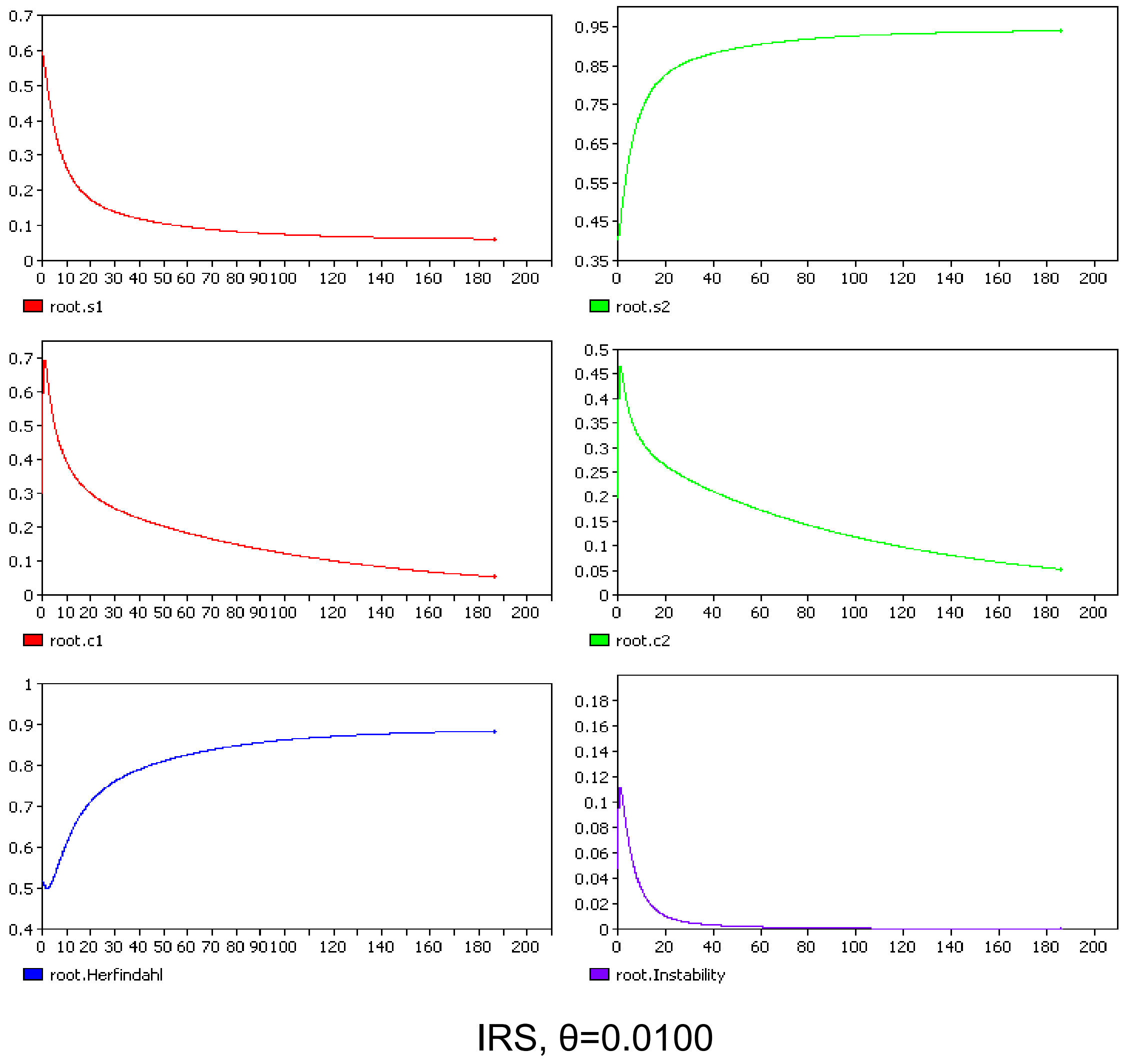

Figure 3.12: IRS impulse responses for the extended model

Note: "root.s*i*" stands for the market share of firm i. "root.c*i*" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

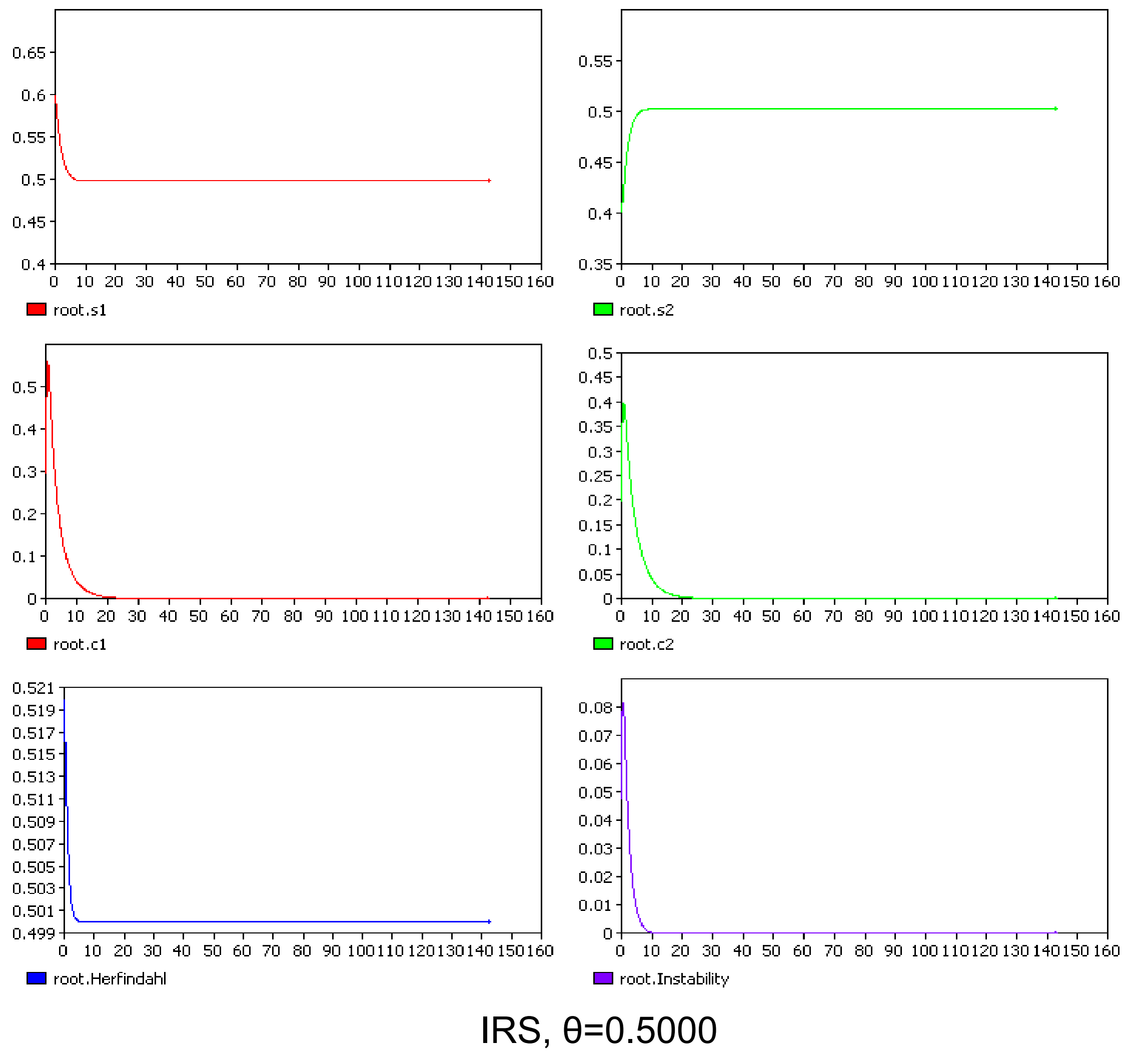

Figure 3.13: IRS impulse responses for the extended model
Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

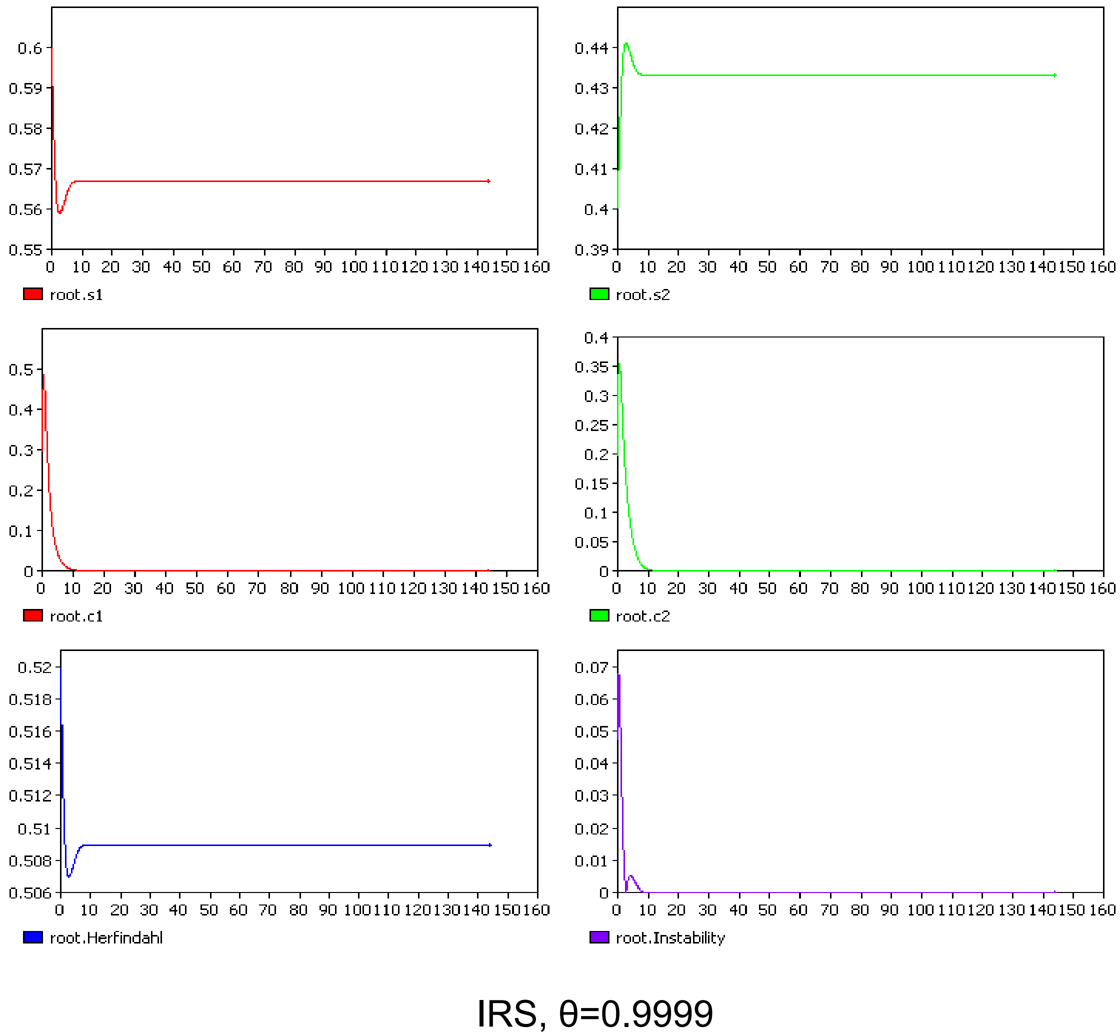

Figure 3.14: IRS impulse responses for the extended model

Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

3.5.7 Appendix 7

Please refer to the next pages.

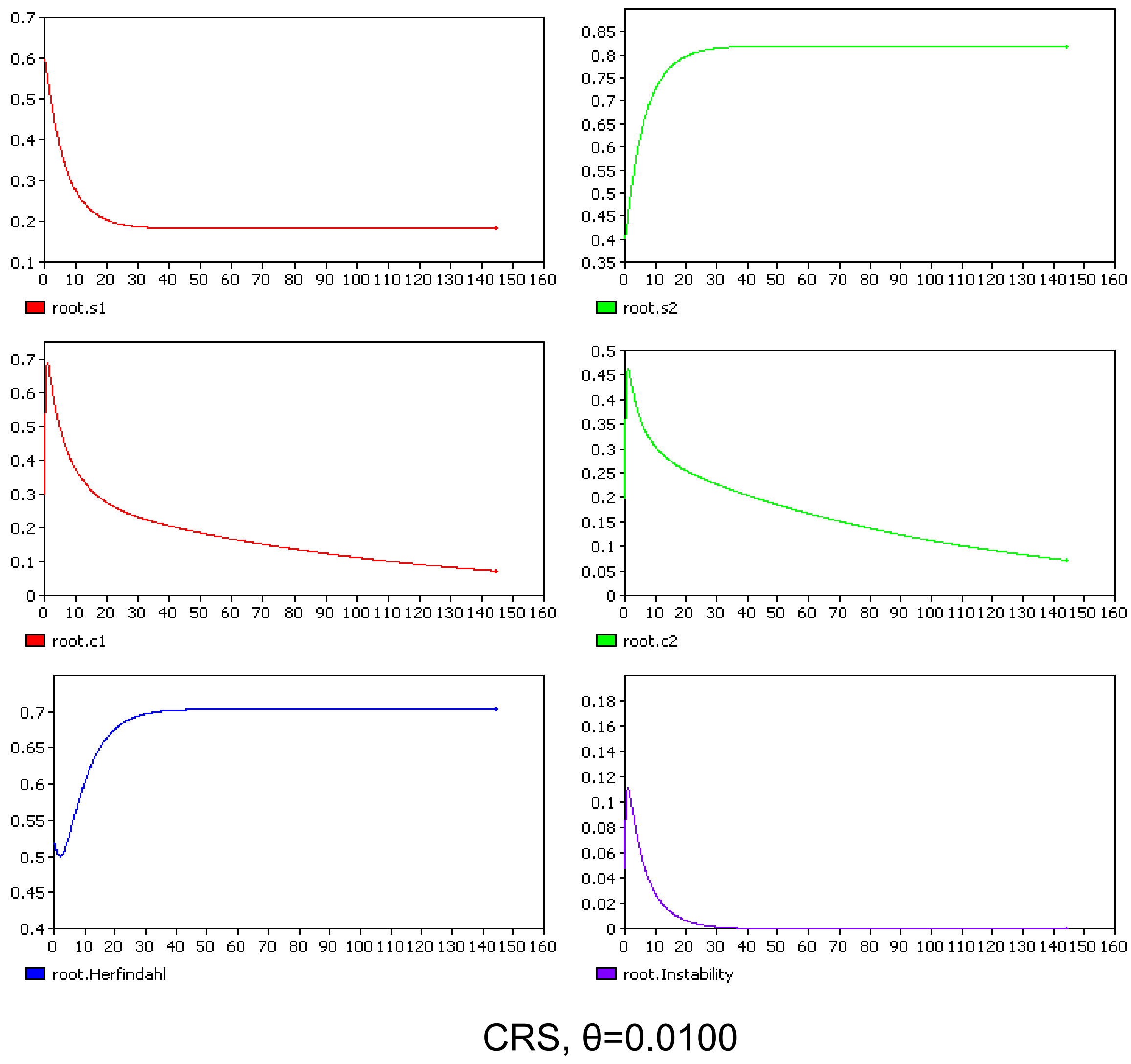

Figure 3.15: CRS impulse responses for the extended model
Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

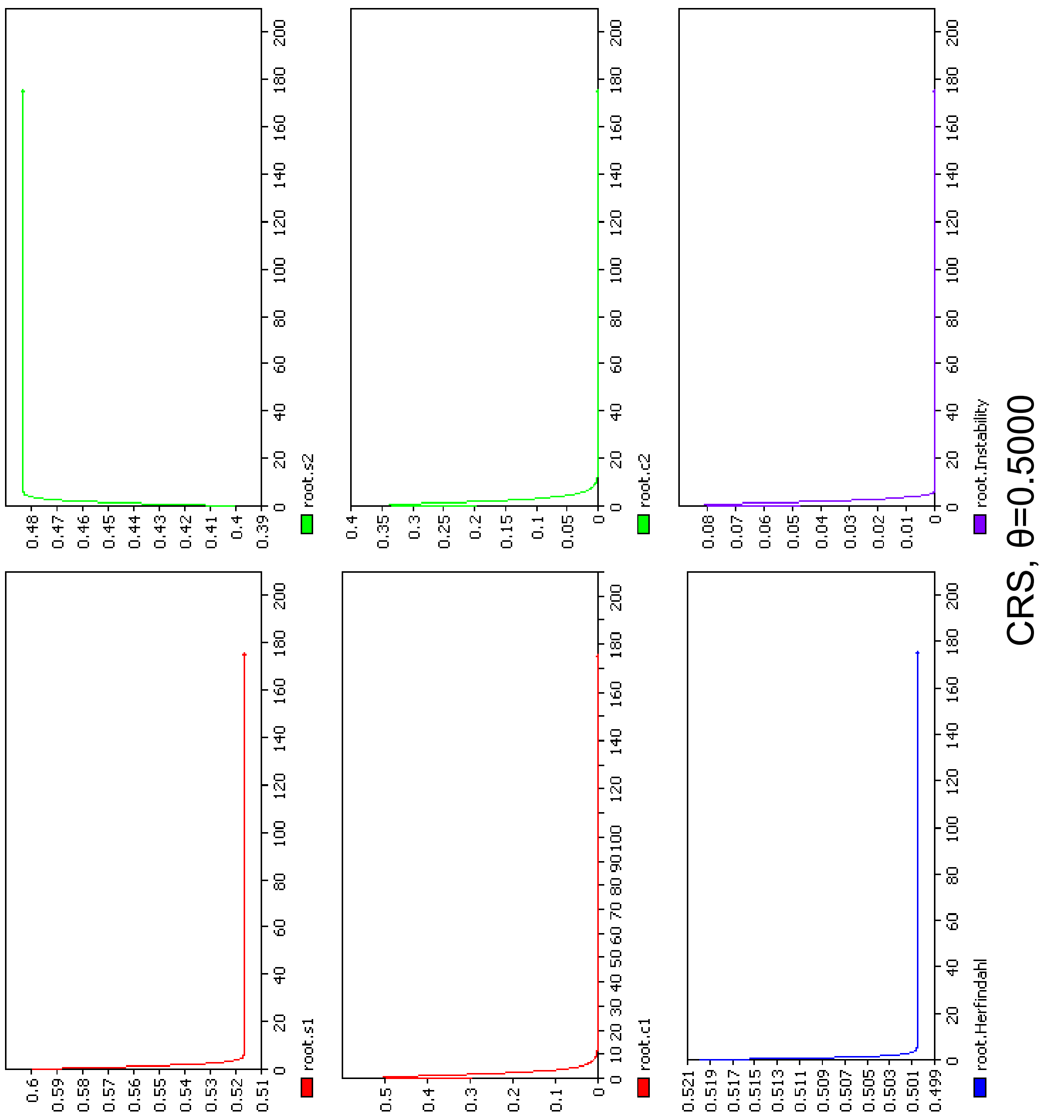

Figure 3.16: CRS impulse responses for the extended model

Note: "root.s*i*" stands for the market share of firm *i*. "root.c*i*" stands for the average costs of firm *i*. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

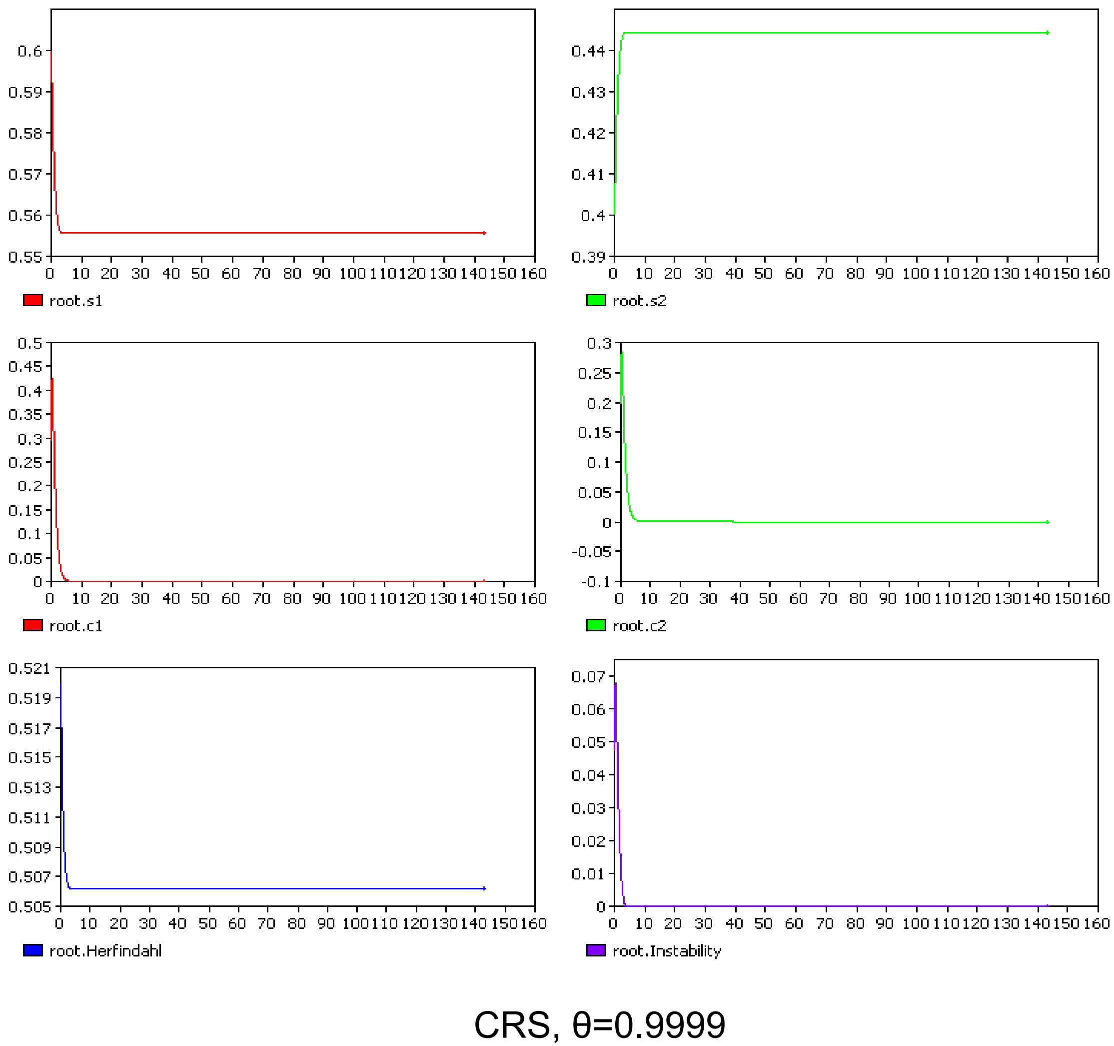

Figure 3.17: CRS impulse responses for the extended model
Note: "root.si" stands for the market share of firm i. "root.ci" stands for the average costs of firm i. "root.Herfindahl" stands for the evolution of the herfindahl index HI. "root.Instability" stands for the evolution of the instability index $\Im$.

4 The spatial dimension of knowledge diffusion

4.1 Introduction

It is an undisputable fact that knowledge and technological change are the driving forces for long run economic growth. Additionally, endogenous growth theory tells us that knowledge spillovers are necessary for long term growth of high-income regions. Several contributions regarding this topic have been published during the last years. (Lucas, 1988), (Krugman, 1991) and (Romer, 1986), for instance have explicitly focused on the accumulation of new knowledge in context of new growth theory. Their key finding is, that endogenous accumulation of knowledge is the surety of per capita income growth. These approaches have in common that they focus on convexities in the production process[1]. For instance, convexities in production can arise from positive externalities caused by learning-by-doing, human capital accumulation and the supply of governmental goods.

As argued by (Keilbach, 2000), knowledge spillovers can be treated as a special type of positive externalities and, moreover, is one motivation for positive returns to scale in an aggregate production function approach which was first used by (Griliches, 1979).

At the latest as European leaders met in Lisbon 2003 and defined the goal of becoming "the most dynamic and competitive knowledge-based economy in the world" by 2010 the term it can be said without any limitations that the knowledge-based economy has gained much attraction, not only in research but also in politics. Today, the creation and diffusion process of knowledge is the focal point of research, because "knowledge is the most important strategic resource and learning the most important process"[2]. But what is knowledge? Well, the term knowledge is often used in scientific publications, but it is sometimes confounded with the term "information". It must be clear that knowledge comprises the individual specific abilities which can be used to solve more or less strategic problems underpinned with a pool of information. As

[1] Refer on (Krugman, 1991) for this topic for instance.
[2] (Morgan, 1997), p. 493.

pointed out by (Krugman, 1991) "[k]nowledge flows are invisible; they leave no paper trail by which they may be measured and tracked[...]". Information instead, is more or less visible. It can interpreted as the collection of knowledge, for instance the collection of data. Hence, when talking about knowledge, we often don't know what we know. Thus, knowledge cannot be measured directly, as other production inputs such as the stock of capital, for instance. The consequence is, that we have to find proxies for this knowledge, for instance human capital or data of patent citations. But doing so, we have to macerate the strict distinction between information and knowledge. That should be kept in mind when talking about the outstanding role of knowledge for economic growth.

Additionally, it is difficult to extract the incentives and resources of knowledge creation and diffusion. As argued by (Rosenberg, 1982), the so called "black box" of innovation which can be described by inherent loops and feedback processes, is also suitable to describe the difficulties of how to identify the source of knowledge creation and dissemination. Given we know the source of knowledge creation, how can we describe concisely the way of how knowledge is transfered from sender to receiver? Is it always the case, that transmitted knowledge can be interpreted correctly by the receiver and more important, is it possible at all to transfer knowledge? The questions we have to ask are therefore, first, is it always true that knowledge diffusion is an unlimited process regarding space, and second, does knowledge transmission depend also on the kind of knowledge?

To answer these questions, we have to think about the kind of knowledge we are talking about. For example, if knowledge is tacit than face-to-face communication or spatial proximity is a necessary condition for knowledge diffusion. On the other hand, if knowledge is codified, then modern communication facilities can be used to transfer knowledge from sender to receiver. Thus codified knowledge is less space depended than tacit knowledge as highlighted by (Anselin et al., 1997). Therefore, we should expect that tacit knowledge dissemination is different from explicit knowledge dissemination with respect to time and space. As mentioned by (Maskell and Malmberg, 1999) tacit knowledge is a key factor for new innovations and thus spatial proximity, which is closely related to tacit knowledge should be acknowledged.

From this point of view, it is plausible not only to focus on time and the kind of knowledge, when integrating knowledge diffusion in a growth model context for example, but also to consider a possible space limitation of knowledge transfer.

It is rather intuitive, that spatial barriers of knowledge diffusion can be used as an argument for income and production differentials between regions. That should be considered as one reason why we observe cluster and agglomeration in economic long run growth. Regions (take cities for example) which are more productive and supply a higher life quality are more attractive for innovative companies. Consequently, these regions become more attractive again and this process leads to a more and more decreasing productive differential. It is not a surprising fact, that economic growth and agglomeration are positive correlated (Baldwin and Martin, 2003). Hence, growth differentials are enforced by knowledge capital concentration. As mentioned by (Fujita and Thisse, 2002), knowledge spillovers can be interpreted as a source for sustainable regional growth, given decreasing returns of learning are excluded.

If we argue that spatial patterns are worth investigating, it is necessary to ask the question how knowledge spillovers affect agglomeration. To answer this question we could argue that cities or densely populated regions may have positive effects on their productivity due to so called Marshallian externalities. (Marshall, 1920) mentioned, that so called externalities are necessary for economic agglomeration and therefore create a so called look-in effect[3]: "When an industry has thus chosen a location for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas."[4] Of course, the justification of agglomeration by Marshall is primarily based on trade arguments but can easily be expanded to other factors, which influence the decision of where to situate a location, as mentioned above. (Kahnert, 1998) found that knowledge intensive processes are agglomerated in dense regions, while less knowledge intensive processes are situated in more peripheral regions. Thus, knowledge spillovers cause externalities and force agglomeration and as a consequence, as pointed out by (Scotchmer and Thisse, 1992) leads to uneven geographical distribution of economic activity.

Hence, from a theoretically driven view, increasing returns to scale, agglomeration and distribution of economic numbers, for instance per capita productivity are closely linked with space. Although, the link of technological innovations and knowledge diffusion for technological growth is acknowledged in growth literature[5], the role of

[3](Fujita and Thisse, 2002), p. 7.
[4](Marshall, 1920), p. 225.
[5]Refer to (Romer, 1986), (Romer, 1990) and (Krugman, 1991) for instance.

knowledge diffusion is only partly considered. Some of the North-South trade literature on diffusion and technological progress[6] consider feedback effects between the North and the South in the steady state, but an analysis of the transitional dynamics for either region is missed. (Barro and Sala-I-Martin, 1997) indeed derived transitional dynamics for the South but feedback effects are excluded due to the effect of no trade of intermediate goods. Thus, a transition path for the North cannot be derived. The communality of this strand of literature is only focused on two country or two region models, which consists of a rich North and a poor South or a core and a peripheral country. From this perspective, those types of models are less suitable to investigate the link of increasing returns to scale, agglomeration and distribution of economic numbers because of the simple reason: in a two country framework, it is not reasonable to investigate agglomeration effects when referring to regions. One of the factors, why multiple country or regional focused growth models are less attractive or gained less attention could be the fact that such growth models become very complex and cannot solved analytically and only numerically solutions remain.

For this reason, the relevant literature which investigates the link between increasing returns to scale, agglomeration and distribution of economic numbers is heavily empirical orientated and is sometimes more or less ad hoc. To investigate spatial agglomeration effects empirically, one has to refer to tools from a toolbox which can be summarized with "spatial econometrics", a term widely used in New Economic Geography (NEG)[7]. (Anselin, 1988)'s book can be described as the first comprehensive introduction to spatial econometrics. In contrast to spatial statisticians, where pure data or data based approaches are in the front, the spatial econometricians deal with model-funded approaches, based upon a theoretical model. However, the commonality of the two perspectives is the acceptance of the existence of spatial stochastic processes.

Although, from an empirical view, there has been made much progress in explaining the link between increasing returns to scale, agglomeration and distribution of economic numbers. But there are still limitations especially when talking about the grasp of knowledge spillovers and knowledge diffusion.

First, less attention is concentrated on the fact, that knowledge diffusion is not a constant process over space. Often it is assumed that only the nearest neighbour has a significant influence on economic growth, whereas farther away neighbours do not exert any economic influence, or more technically spoken, often it is assumed that knowledge diffusion follows a spatial AR(1) or spatial MA(1) process and second or higher order effects or a combination of both are neglected. This assumption seems to be to strict.

[6] Refer to (Krugman, 1979), (Dollar, 1986), (Grossman and Helpman, 1991b), (Grossman and Helpman, 1991a), (Rivera-Batiz and Romer, 1991), (Barro and Sala-I-Martin, 1997) and (Glass, 1997).

[7] For an overview of NEG refer to (Krugman, 1998a) and (Krugman, 1998b) for example.

Instead of ignoring higher order effects of spatial influence, one should insert them into a model framework, because neglecting them could lead to an underestimating of spatial influence. Further, this second or higher order processes should not be treated as a constant extrapolation, but rather as non constant function over space. Hence, it is reasonable to assume that more contiguous neighbours have a direct and stronger influence than less contiguous neighbours.

In most of the existing empirical studies the grasp of knowledge spillovers has only gained limited attention. (Anselin et al., 1997) and (Anselin et al., 1997) are two of the few studies how mentioned concrete numbers of knowledge spillover scope. (Anselin et al., 1997) found by investigating the influence of university related research and private research and development (R&D) effort on of knowledge transfer that a significant positive effect can be detected within a 50 mile radius of metropolitan statistical areas (MSAs) only for the university research. For private R&D such an significant effect could not be detected. (Anselin et al., 1997), with a similar setup as (Anselin et al., 1997) additionally have shown, that not only spillovers within MSA but also between MSA can be found. The key cognition of the latter mentioned study is, that without exact geographical distance measures, it can be shown that spatial influence is bounded locally. (Audretsch and Mahmood, 1994) have shown on patent basis for 59 US metropolises, that knowledge spillovers are limited towards the metropolises' boarders. They come to this conclusion because they found that only for research institutes which are settled within a metropolis, significant knowledge spillovers can be detected, whereas for research institutes, settled in each metropolis related country, no such effects could be found.

Second, within the specification of spatial models, spatial heterogeneity is mostly missed. It is sometimes ignored, that spatial effects can appear as two types: the one type is spatial dependence, the other is spatial heterogeneity. Spatial dependence, which is consistently assumed in the above mentioned studies, is mainly caused by problems of measuring that are caused by spatial spillovers and spatial externalities. In contrast to spatial dependence, spatial heterogeneity means that spatial effects are not uniformly distributed across space and outliers could exist. From a standard econometricians toolbox, this could be seen as a spatial kind of heteroscedasticity. Although several arguments militate in favour that spatial heterogeneity matters[8], this aspect is not "seen as a serious problem in spatial regression"[9]. One reason could be, that spatial

[8](Anselin, 1988) for instance comment on page 13 with respect to importance of spatial heterogeneity in econometricians work, that "several factors, such as central place hierarchies, the existence of leading and lagging regions, vintage effects in urban growth [...] would argue for modeling strategies that take into account the particular features of each location (or spatial unit)."

[9](Keilbach, 2000), p. 122.

econometrics, if we refer to theoretical econometrics, is still a developing discipline.

But what should be done, if spatial dependence, spatial heterogeneity or a combination of both types is relevant and further a set of possible AR(p), MA(q) or ARMA(p,q) processes with order *p*and *q* respectively, are suitable in model context? Given, our model is correctly specified, than standard econometrics tells us, that parameter estimates are inefficient if spatial heterogeneity is ignored, although it is relevant. But given, the model is based on a wrong choice of AR(p), MA(q) or ARMA(p,q) terms, then our model is wrong specified. Of course, the latter problem is the more serious one.

Although, model selection should be taken seriously, we frequently find that empirical based studies using tools from spatial econometrics, based on ex ante conceptions of a spatial model. This means, a model selection is often defaulted or, if done, it is based mainly on a limited class of spatial processes, which commonly include the decision of relying on a spatial AR(1) or spatial MA(1) process based on the assumption of spatial homogeneity. There are, to best of my knowledge only a few papers which cover the aspect of spatial model choice.[10]

Thus, traditional or frequentest econometrics approach suffers for two reason in the context of spatial econometrics: first, the models and the underlying estimation methods assume spatial homogeneity, and second, model selection is rather heuristic. For these reasons, Bayesian methods have been prevailed and proved in spatial econometric application. The key difference between frequentest and Bayesian methods are that the latter treat the coefficient vector of estimators itself as random, whereas frequentest say that the resulting estimates of the coefficient vector is random. Bayesian methods hold a great deal for several reasons: for instance, first, it is possible to model hierarchy of place or regions, second, one can integrate a more or less systematic change of variance over space, and thus spatial heterogeneity and third it is possible to acknowledge a hierarchy of regions or places. Bayesian methods can incorporate these ideas because of their underlying concept as prior information complements existent sample data information, whereas frequentest methods can solely rely on latter mentioned. As mentioned before, although Bayesian methods seem to be very attractive, their usage in application is very limited. On the other side, frequentest methods are, if they only limited to the spatial dependence case, and therefore assume spatial homogeneity, lead to insufficient parameter estimates. Anyway, a more or less large research agenda for both, spatial econometrics and spatial statistics remains.

From the discussion above, we see that two different arguments regarding productiv-

[10]For instance refer to (Hendry, 1979), (Florax et al., 2003) and (Hendry, 2006) for an intensive discussion regarding model selection methods.

ity growth are discussed in the relevant literature: on the one hand, the (theoretically) role of technological innovations and knowledge diffusion for technological growth[11], and on the other hand the (empirical) role of spatial agglomeration on long run productivity growth[12]. The point is, that the first mentioned strand does discuss growth implications of knowledge diffusion in a less suitable frame when focusing on distribution questions and agglomeration, while the latter strand suffers more or less from theoretical fortification.

Hence, these two approaches are more or less discussed in isolation rather to be combined and to investigate the relationship between knowledge diffusion, agglomeration and growth. This topic has gained less attention in relevant literature, although (Fujita and Thisse, 2002) mentioned that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[13].

There is to best of my knowledge only one study, which tries to bridge the two approaches: (Keilbach, 2000) has investigated the role of knowledge for German "Kreise"[14] both empirically and theoretically within a (Romer, 1986) context. He found, that increasing returns to scale lead to significant cluster effects. Further, he found on basis of several production functions estimations, that spatial dependence has a significant influence on labour productivity. But it has to be mentioned, that (Keilbach, 2000) assumes explicitly spatial homogeneity and only first order spatial effects, both in his theoretical and empirical studies. Further, using "Kreise" as regions could lead to spatial dependence per definition, due to the fact that "Kreise" are the smallest entity of regions for the case of Germany, and thus stream of commuters can lead to biased estimations of spatial effects by construction.

Thus, one intention of this chapter is, to include the economic variable space in a simple theoretical hybrid growth model, which core is based on to the model of (Mulligan and Sala-I-Martin, 1993), (Uzawa, 1965) and (Lucas, 1988). The purpose of the theoretically derived model is to derive a theoretical growth orientated justification of the "folk theorem of spatial economics"[15], that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[16].

In the theoretical model it is assumed that regions are learning regions which means that low-income regions can catch up to high-income regions. This spatial catch up process has not been acknowledged in growth theory so far. The implication is, that

[11]Refer to (Romer, 1986), (Romer, 1990) and (Krugman, 1991) for instance.

[12]Refer to (Keilbach, 2000), (Bottazzi and Peri, 2003), (Greif, 1998) and (Frauenhofer, 2000) for instance.

[13](Fujita and Thisse, 2002), p. 342.

[14]"Kreise" is a German administration unit which is equivalent to NUTS-3 level.

[15]Refer to (Scotchmer and Thisse, 1992).

[16](Fujita and Thisse, 2002), p. 342.

knowledge is not completely tacit but contains a certain public good character as highlighted by (Brezis and Krugman, 1993). On the other site, one has to acknowledge the fact, that spatial influence is limited and not constant over space. This is a consequence of the (Fujita and Thisse, 2002) thesis explaining economic clusters. Thus, the aim of this chapter is first, to investigate the role of knowledge and agglomeration which is a logical combination of the role of growth and knowledge and the role of growth and agglomeration in a theoretical growth model context.

Second, on the basis of the developed theoretical model, it is investigated, given spatial influence is limited and not constant over space, whether spatial spillovers are more local or more global and thus, the "folk theorem of spatial economics"[17] can be justified also empirically. If knowledge spillovers are more local, then this would be an explanation of agglomeration or cluster effects and a confirmation of the "folk theorem of spatial economics"[18]. This empirical study is based on a spatial cross section production function approach, proposed by (Griliches, 1979) which should measure the effects of innovativeness, measured by knowledge capital, such as human capital, patents or *R&D* and spatial spillovers on output for German NUTS-2 regions. NUTS-2 regions are used to exclude spatial dependence by construction.

Further, a new model choice mechanism is introduced which on the one hand is based on traditional econometric tools and on the other hand integrates Bayesian model choice criteria. This mechanism also controls for spatial heterogeneity. Finally, under the condition that spatial processes can be detected in the data, a filter method is applied to remove spatial influence and thus to identify own and neighbour productivity effects of regions and to discuss political implications against the background of obtained results.

4.2 Theoretical model

The aim of the theoretical model is to find support or not for the fact, that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[19] and thus to justify the "folk theorem of spatial economics"[20]. The model further assumes that spatial dependence over space is not constant. Because the model could become very complex, a *Cellular Automaton* programming technique is consulted to simulate spatial patterns. The next section deals with the empirical conversion of

[17] Refer to (Scotchmer and Thisse, 1992).
[18] Refer to (Scotchmer and Thisse, 1992).
[19] (Fujita and Thisse, 2002), p. 342.
[20] Refer to (Scotchmer and Thisse, 1992).

the theoretical model context.

4.2.1 Setup

This section deals with the setup of a discrete spatial growth model, which links knowledge creation, spatial knowledge diffusion and productivity to investigate the link of knowledge, agglomeration and growth. For this purpose, a two sector model which is similar to the model proposed by (Lucas, 1988) is set up and expanded in several ways as laid out in the this section.

Assume a world of $i = \{1, 2, ..., N_i\}$ regions which are distributed randomly over the entire space of the world. Every region is heterogeneous in the sense that it can be characterized by a specific labour productivity y_i which is different in every region i. Furthermore, every region i has different neighbours $j = \{1, ..., N_j\}$.

As mentioned above, two sectors are considered in the model. The first sector is the knowledge production sector. This sector produces exclusively knowledge with a specific neoclassical production technique Q. Moreover it is assumed, that every region i produces its own knowledge stock W^i. For the production technique we can write for region i in $t = \{1, 2, 3, ..., T\}$

$$Q^i_t(K^i_t, W^i_t, L^i_t) = B[a_K K^i_t]^\gamma [a_W W^i_t]^\phi [A^i_t L^i_t]^\kappa, \tag{4.1}$$

with W^i_t as the knowledge stock, K^i_t as the capital stock and L^i_t as unskilled workforce of region i. $B > 0$ is a global shift parameter and $a_K \in (0, 1)$ and $a_W \in (0, 1)$ stand for the global fractions of capital and knowledge stock used for production of new knowledge. $\gamma \in (0, 1)$, $\kappa \in (0, 1)$ and $\phi \in (0, 1)$ are the corresponding production elasticities. Thus, every region i produces with the same production technique Q^i_t in the knowledge production sector. A^i is a time dependent shift parameter with constant growth rate g^i_a, so that $A^i_{t+1} = (1 + g^i_a)A^i_t$.

As one can easily see from equation 4.1 is that unskilled workforce is entirely used in the sector of knowledge creation and cannot be used in the goods sector. This assumption seems to be strict at first glance, but the focus on this model is to work out the link of knowledge, agglomeration and growth. Of course, we can expand the model in this sense, that a fraction, say a_L can also be employed in the goods sector. But the implications of this model remain unaffected by this modification.

The goods sector is formulated similarily to the knowledge producing sector with the exception that only knowledge and capital are needed to produce output Y_t^i. For that reason one can write the production function Y_t^i as follows:

$$Y_t^i(K_t^i, W_t^i) = [(1 - a_K)K_t^i]^\alpha [(1 - a_W)W_t^i]^\beta. \tag{4.2}$$

Thus, every region i produces with the same production technique Y_t^i in the goods sector. As one can see from equation 4.2 the good is produced via "transformed" labour through knowledge capital generation and capital stock K_t. For the labour productivity in efficiency units y_t^i[21] we can write:

$$y_t^i \equiv \frac{Y_t^i}{A_t^i L_t^i} = [(1 - a_K)k_t^i]^\alpha [A_t^i L_t^i]^{\alpha - 1} [(1 - a_W)W_t^i]^\beta, \tag{4.3}$$

with $k_t^i = \frac{K_t^i}{A_t^i L_t^i}$. As usually, it is further assumed that labour is growing with constant rate g_n^i so that $L_{t+1}^i = (1 + g_n^i)L_t^i$.

In the next step we have to think about the integration of space in our model. This is done in several ways. First we have to formulate a rule for the unskilled labour. It is assumed that unskilled labour is not very mobile and mostly bounded to its origin region due to social connections as family, friendship relations etc.. Labour from a region i is only emigrating if it offers the lowest wage payed in the goods sector in the set of neighbours. More technical, a fraction θL_t will leave region i in t. On contrary, if region i offers the highest wage in the set of neighbours, then labour force from neighbouring regions is immigrating in region i. Again more technical, a fraction $\theta \sum_j L_t^i$ will immigrate to region i. Otherwise due to strong social ties, no migration movement occurs. Therefore we can formulate the following transition rule:

$$L_{t+1}^i = \begin{cases} \left\{\theta \sum_j L_t^j + L_t^i\right\}(1 + g_N) & \text{if } w_t^i = w_t^{max} \\ L_t^i(1 - \theta)(1 + g_N) & \text{if } w_t^i = w_t^{min} \\ L_t^i(1 + g_N) & \text{otherwise} \end{cases}. \tag{4.4}$$

with w_t^{max} as the maximum wage payed in the set of neighbours and region i, and with w_t^{in} as the minimum wage payed in the set of neighbours and region i. Assumption 4.4 can also be interpreted as the fact that an unskilled worker is not perfectly informed

[21] In the following "labour productivity" and "labour productivity in efficiency units" are used as synonyms.

about wage conditions in the entire world but only within the neighbourhood of his home region i. If the wage is situated between w^{min} and w^{max}, then there is no incentive to leave the home region i. Of course, if $\theta = 0$ no migration can be observed, the states of the system are entirely absorptive with respect to space but not with respect to time, because L_t^i is constant over space, but not over time. If g_n^i is also set to zero, L_t^i is constant over time and over space.

On contrary to the labour market which is more local, the capital market is organized globally and capital is mobile over the entire space of our world. This means that an a priori fraction of the investments φ from region i flows in that region j which exhibits a higher net capital productivity r_t^j compared to the mean capital productivity $\bar{r}$. The fraction $(1-\varphi)$ is invested in the region of origin. Although, the flow is not regionally bounded, the factor $\varphi \in [0,1]$ weights the neighbouring investments $sY_i\varphi$ of region i to acknowledge possible capital transfer restrictions, which may be imposed by politics or can be intrinsicly motivated. Thus, the transition rule for the capital is formulated as follows:

$$K_{t+1}^i = \begin{cases} s\left[\left(\sum_j Y_t^j \varphi \chi_i\right) + (1-\varphi)Y_t^i\right] + (1-\delta_K)K_t^i & \text{if } r_t^i > \bar{r} \\ (1-\varphi)sY_t^i + (1-\delta_K)K_t^i & \text{if } r_t^i < \bar{r} \end{cases}, \tag{4.5}$$

with $\varphi \in [0,1]$ as the fraction of investment which is made in neighbouring regions, $\delta \in [0,1]$ as the depreciation rate on capital and χ_t^i represents the weighting measure for capital flows. To obtain a weighting measure of how much capital flows a priori to neighbouring regions we construct an endogenous weighting measure which depends on the relationship of own marginal product of capital and the sum of neighbouring marginal products of capital. This can be transfered into the following equation:

$$\chi_t^i = \begin{cases} \frac{r_t^i}{\sum_j r_t^j} & \text{if } r_t^i > \bar{r} \\ 0 & \text{if } r_t^i < \bar{r}, \end{cases}, \tag{4.6}$$

which implies $\chi_t^i \in (0,1)$. From equation 4.6 we can see that even if $\chi \in (0,1)$ is positive for a region i, capital restriction in other regions j may hinder the flow to the own region i. For example, set $\varphi = 0$, then region i can reinvest only its own savings, even if $\mathrm{r}_t^i > \bar{r}$.

If we assume, that further increase of investment I_t^i is associated with higher investment expenditures, we may have to think about capital costs $\phi\left(\frac{I_t}{K_t}\right)$. A priori, capital costs should play a crucial role not only for home investments but also for neighbhouring investments. For that reason, we formulate

$$\phi\left(\frac{I_t^i}{K_t^i}\right)^i = \begin{cases} \left(\frac{1}{1-\zeta}\right)\left(\frac{s(\sum_j Y_t^j \varphi\chi_i + (1-\varphi)Y_t^i)}{K_t^i}\right)^{(1-\zeta)} & \text{if } r_t^i > \bar{r} \\ \left(\frac{1}{1-\zeta}\right)\left(\frac{sY_t^i(1-\varphi)}{K_t^i}\right)^{(1-\zeta)} & \text{if } r_t^i < \bar{r} \end{cases}, \tag{4.7}$$

with $\zeta > 0$. Thus for the transition rule of capital stock K_t^i we have to choose the following notational form:

$$K_{t+1}^i = \phi\left(\frac{I_t}{K_t^i}\right)K_t^i + (1-\delta)K_t^i. \tag{4.8}$$

We have to note, that $\phi(\cdot)$ is a concave and decreasing function its relevant argument and if one sets $\phi\left(\frac{I_t}{K_t}\right) \equiv \left(\frac{s(\sum_j Y_t^j \varphi\chi_i + (1-\varphi)Y_t^i)}{K_t^i}\right)$ or $\phi\left(\frac{I_t}{K_t}\right) \equiv \left(\frac{sY_t^i(1-\varphi)}{K_t^i}\right)$ one obtains equation 4.5 together with equation 4.8.

In the next step we have to create a direct link between knowledge spillovers and labour productivity. For this scope, we assume that a region i will benefit from "knowledge creation" of other regions j. Hence, the knowledge stock W_{t+1}^i is determined by the production of knowledge Q_t^i and the weighted spillovers $\sum_j W_t^j$ from neigbouring regions j. Therefore, we can formulate a transition rule for the knowledge stock W_{t+1}:

$$W_{t+1}^i = Q_t^i + \iota_i^t \sum_j W_t^j + (1-\delta_W)W_t^i, \tag{4.9}$$

whereas $\iota \in [0,1]$ is an endogenous measure of degree of spillovers and δ_W represents the deprecation rate on knowledge. It is assumed that the degree of spillovers ι_t^i can be modeled as a function of the maximum stock of knowledge which is available in the economy W_t^{max} and the region specific knowledge stock W_t^i. Thus, the spillover degree is the greater the smaller the difference of W_t^{max} and W_t^i is. Accordingly, we can write

$$\iota_t^i = 1 - \left\{\frac{W_t^{max} - W_t^i}{W_t^{max}}\right\}, \tag{4.10}$$

which is $\in [0,1]$. If $W_t^{max} - W_t^i = 0$ then ι_t^i takes its maximum level of one. On contrary, $\iota_t^i = 0$ if $W_i^t = 0$.

Not only W_t^i accounts for spillovers but also Y_t^i itself. It is known from the convergence debate that emerging countries should grow faster if they have not reached their balanced growth path. If we define an endogenous technological gap as $\Theta_t^i = \frac{\bar{Y}_t - Y_t^i}{\bar{Y}_t}$, then, in every period of time t a fraction of the technological gap $\Theta_t^i \in (0,1)$ can be reduced by $\vartheta \in (0,1)$ if and only if the region i is innovative. Whether a region i is innovative or not depends solely on a normal distributed random variable $\varpi \in [0,1]$. If this parameter $\varpi \in [0,1]$ exceeds a given threshold $\pi \in [0,1]$ then a region is innovative. In this way the tacitness of knowledge has been integrated. Remember, if $\pi \to 1$ then knowledge tends to be completely tacit and the probability of innovativeness is very small. This scenario induces a kind of knowledge which is hard to understand and therefore cannot be used with a high probability to reduce the technological gap. Otherwise, if $\pi \to 0$ the probability of tacit knowledge tends to zero and hence a large proportion of regions is innovative. In notational form, we can write for the technological gap Θ_t^i:

$$\Theta_t^i := \begin{cases} -\frac{Y_t^i - \bar{Y}_t}{\bar{Y}_t} & \text{if}: \; Y_t^i < \bar{Y}_t = \frac{1}{|H|}\sum_j Y_t^j \varpi > \pi \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Note that $\vartheta \in (0,1)$ and $\varpi \in [0,1]$ are treated with this formulation as global parameters. For the production of region i at the beginning of the next period $t+1$ we can write Y_t^i:

$$Y_{t+1}^i := \begin{cases} Y_t^i + \vartheta\Theta \text{ if}: \Theta > 0, \\ Y_t^i, \text{ otherwise.} \end{cases} \tag{4.12}$$

In this section we have defined a hybrid spatial growth model which should give a first hint of how knowledge creation, production and knowledge diffusion interact, not only in time, but also in space. As one can see, due to its complexity, this model cannot be solved analytically but numerically. The complexity stems particularly from the fact, that knowledge diffusion can be characterized with feedback rules. In the next section the simulation frame for the hybrid model is set up. *Cellular Automaton* (CA) is very attractive for simulation spatial models owing due its construction.

4.2.2 Cellular Automaton

A *Cellular Automaton* (CA) is a simple mathematical system, which shows highly complex behavoir[22]. It consists, loosely spoken, of a number of cells. Every cell checks for every period of time its own and its corresponding activities of their neighbours and updates if necessary its state based on given rules. On general, a *Cellular Automaton* consists of a d-dimensional grid D, cells and neighbourhoods of cells H and a transition rule κ. Usually, time is discrete and the transition rule is deterministic but may be influenced by stochastic global and local parameters Γ and Φ, respectively. The transition rule is responsible for the dynamic behaviour of the defined system.

The charme of the (CA) technique is that spatial effects or space itself can be modeled in an explicit way, because region and neighbourhood structures can be modeled. Another way of modeling space is referring on so called *Agent-based modeling* (ABM), which has attracted significant attention in social science during the last years. Prima vacie, (ABM) provides several advantages, such as controlling for heterogeneous entities, it encounters in fact several seriously methodological problems, especially the massive parameter space and the problem of validation. The implication of the first problem is, that we do not know which parameter settings leads to the desired behaviour of our system. Parameter setting is heuristic and not based on selection mechanism. Further, it is not possible to exclude singularities and discontinuities in the entire model space. Some regions could exhibit chaotic behaviour, whereas other regions do not. The implication of the second problem is, that it is not possible to derive an empirical model from the (ABM) structure. (CA) instead of (ABM) only provides a (spatial) framework, in which model behaviour can be discussed. As seen below, also (CA) is suitable to discuss heterogeneous phenomena.

Let us start with the definition of the dimension of (CA). It is a regular 2-dimensional and quadratic $n \times m$ grid. Thus, we can write:

$$D := \{(i,j)|i,j \in \mathbb{Z}, 0 \leq i < N_i, 0 \leq j < N_j\}. \tag{4.13}$$

Next, we have to make some remarks regarding a given state Z of our model. At first glance, we could think we could assume that the state vector Z is a τ-tupel and can be formulated in general as

$$Z^\tau = \{0, 1, 2, 3, ..., \tau - 1\}. \tag{4.14}$$

But in the model context we identify several states for the variables L, K, W and Y

[22]For an overview of *Cellular Automaton* please refer to (Wolfram, 1994).

due to the fact that $Z \in \mathbb{R}^0_+$.

In this model, agglomeration of labour productivity is in the focus of investigation. Thus, if we consider a 2 dimensional grid, we can stack each region specific labour productivity y in a $n \times m$ matrix D. In this way, it is possible to observe the evolution of labour productivity over time t and over space which is defined via D. In this way, in every time step t a Gini-coefficient with respect to y with respect to regions can be computed. In addition, the evolution of spatial correlation of y can be measured [23].

Further, we have to consider the neighbour relationship of each cell i. Usually, refering on (CA) we distinguish between *von-Neumann* (vN) and *Moore* (M) non absorptive but periodic neighbourship relations. Let us define a so called immediate neighbour cell h which does not consider itself as a neighbour. Thus the neighbour relations for a cell i located on the two dimensional grid with coordinates $\{a, b\} \in D$ in t are:

$$i_t^{a,b} = \begin{cases} (i_t^{(a-1,b-1)}, i_t^{(a-1,b)}, i_t^{(a-1,b+1)}, i_t^{(a,b-1)}, i_t^{(a,b+1)}, i_t^{(a+1,b)}, i_t^{(a-1,b-1)}, i_t^{(a+1,b+1)}) & \text{if (M),} \\ (i_t^{(a-1,b)}, i_t^{(a,b-1)}, i_t^{(a,b+1)}, i_t^{(a+1,b)}) & \text{if (vN).} \end{cases} \quad (4.15)$$

Thus, if one refers to (M), then a region i has 8 direct neighbours, whereas a (vN) world implies 4 direct neighbours for a given region i under the condition $r = 1$. These different kinds of first order neighbourships ($r = 1$) can also be graphically demonstrated as in figure 4.1.

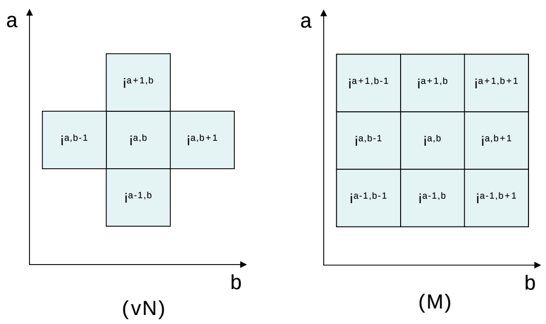

Figure 4.1: Representation of (vN) and (M) neighbourship relations with ($r = 1$)

In this model, we rely on the *Moore* (M) relationship. We can see, that the (M) relationship builds a "ring" of neighbours with radius $r = 1$ round the cell of interest $i^{a,b}$. At this point, it should be kept in mind, that we have to integrate assumption 4 in our model, which means that we have to think about a more explicit space dependency. The easiest way to do this, is to create a second ring round the neighbour cell $i^{a,b}$ with

[23]With `Matlab 6.5.0` one can visualize this simulation experiment with `spy(D)` for instance.

radius $r = 2$. Of course, one can go further to integrate higher degrees of r, but this should be enough do see the difference if one acknowledges the so called "neighbours of neighbours" influence. On general, we can write for $r = \{1, 2, ..., R\}$:

$$i_t^{a,b} = \begin{cases} (i_t^{(a-1,b-1)}, i_t^{(a-1,b)}, i_t^{(a-1,b+1)}, i_t^{(a,b-1)}, i_t^{(a,b+1)}, i_t^{(a+1,b)}, i_t^{(a-1,b-1)}, i_t^{(a+1,b+1)}) & \text{if } r = 1, \\ (i_t^{(a-1,b-1)}, i_t^{(a-1,b)}, i_t^{(a-1,b+1)}, i_t^{(a,b-1)}, i_t^{(a,b+1)}, i_t^{(a+1,b)}, i_t^{(a-1,b-1)}, i_t^{(a+1,b+1)}, \\ i_t^{(a-2,b-2)}, i_t^{(a-2,b)}, i_t^{(a-2,b+2)}, i_t^{(a,b-2)}, i_t^{(a,b+2)}, i_t^{(a+2,b)}, i_t^{(a-2,b-2)}, i_t^{(a+2,b+2)}) & \text{if } r = 2, \\ \vdots \end{cases} \tag{4.16}$$

Thus with this notation the (CA) represents an economy which is divided into several regions and which allocates an identical number of neighbours to each region. We can therefore represent the economy as a so called circular city.[24]

As mentioned, it is assumed that spillovers are not treated as constant over space and further it is assumed that they are limited over space. More concrete a region i benefits more from the nearest regions than from farther away regions regarding knowledge spillovers. Thus we have to introduce a spatial weighting scheme of neighbourhood potential regarding. Further, we have to acknowledge home effects of a given region i. In this way, we have to discriminate region specific effects and neighbour effects which affects a given knowledge specific economic variable $\tilde{V}_t^i \in \mathbb{R}_0^+$. Label V_t^{spill} the spillover potential of neighbourhood and V_t^i the region specific economic variable then overall effect can be written as

$$\begin{aligned} \tilde{V}_t^i &= \left[\left(\xi^1 \sum_{k \in N_j^1} V_t^k + \xi^2 \sum_{k \in N_j^2} V_t^k + ... + \xi^R \sum_{k \in N_j^R} V_t^k \right) + V_t^i \right] \\ &= \left[V_t^{spill} + V_t^i \right] \end{aligned} \tag{4.17}$$

with $\xi^1 \geq \xi^1 \geq ... \geq \xi^R$, and $N_{i,j}^r \subset N_{i,j}$ for $r = \{1, 2, ..., R\}$ and $\xi^r \in (0, 1)$ which act as a weighting parameter for higher order neighbour influence. If $r = 1$ only nearest neighbour relations matter. The latter assumption is the common assumption which has been made in empirical literature when talking about spatial effects.

Now we are able to set up the dynamic behaviour of the CA. For that purpose, we need a mapping scheme to integrate the dynamics into our system. Please note, that a given variable Z_t is endogenous because it is influenced through the neighbours H_t and global and local parameters Φ_t and Ψ_t[25]. Therefore, let us write $Z_t(H_t, \Psi_t, \Phi_t)$[26]

[24] Refer to (Tirole, 1988), (Hotelling, 1929) and (Krugman, 1995).
[25] The vectors contain the depreciation rate, the saving rate etc..
[26] Φ and Ψ may be time variant or not.

To map the dynamics a mapping function κ is required. This function reads as follows:

$$\kappa := Z_t^{H_t} \to Z_{t+1}. \tag{4.18}$$

4.2.3 Model simulation

As easily can be seen from above, the model is not restricted to have constant returns to scale, which means that $\alpha + \beta = 1$ and $\gamma + \psi + \kappa = 1$. For instance, if the goods sector exhibits increasing returns to scale $\alpha + \beta > 1$ even in a competitive environment, if knowledge spillovers are introduced as done by (Lucas, 1988). As known, the results obtained in a competitive environment are generally not Pareto optimal. In this case, governmental subsidize schemes have to be initialized to subsidize activities with positive spillovers. Further it should be noted, that large spillovers could create multiple equilibria which can be ranked by the Pareto criterion.[27]

As highlighted by (Lucas, 1988) knowledge spillovers lead to increasing returns to scale in the goods sector. Of course, such a condition is compatible with endogenous growth, but it is not a necessary condition. The model of (Lucas, 1988) can also generate endogenous growth without knowledge spillovers from knowledge sector. Although, the focus on this analysis is not in first line tend to discuss the conditions of endogenous growth in this model framework, this should fact should be kept in mind.

If we turn back to our simulation exercise and if we further follow (Eicher and Turnovsky, 1999), three simulation scenarios are distinguished: first, both the goods sector and the knowledge good sector exhibit increasing returns to scale, second, the goods sector and the knowledge good sector exhibit constant returns to scale, and third, both sectors exhibit decreasing returns to scale. All scenarios are run for first order and second order spatial influence.

For the simulation study, it is assumed, that labour is mobile, which means that $\theta > 0$ and it is growing with a constant rate g_L. Further it is assumed, that capital is mobile and capital restrictions are close to zero ($\varphi = 0.99$). For the capital adjustment costs a value of $\rho = 0.5$ has been chosen. The savings rate is set to $s = 0.10$ which reflects a ten year average saving rate for Germany[28]. It is further assumed that in every period the technological gap of a region i can be reduced by $\vartheta = 0.10$. This is a very small value, but it is in line with the assumption that knowledge is tacit which means that $\pi = 0.8$. Furthermore, first order ($r = 1$) and second order influence of neighbourhood ($r = 2$) is not constant over space but decreasing, hence we set $\xi = 0.1$.

[27]Refer to (Barro and Sala-I-Martin, 1995), p. 199.

[28]Refer to the homepage of "Statistische Bundesamt": http://www.destatis.de for further information regarding the development of the German saving rate.

Values for elasticities of production have been chosen according to the works of (Lucas, 1988) and (Jones, 1995a). Data for depreciation rates both for human and physical capital have been taken from (Kydland and Prescott, 1982). Table 4.1 provides a summary of the parameter settings.

Sector	Parameter	CRS	DRS	IRS
Goods sector	α	0.360	0.300	0.400
Goods sector	β	0.640	0.500	0.700
Goods sector	(1-a_K)	0.500	0.500	0.500
Goods sector	(1-a_W)	0.500	0.500	0.500
Goods sector	ϑ	0.100	0.100	0.100
Goods sector	π	0.800	0.800	0.800
Knowledge sector	γ	0.100	0.100	0.100
Knowledge sector	ϕ	0.300	0.200	0.400
Knowledge sector	κ	0.600	0.200	0.600
Knowledge sector	a_K	0.500	0.500	0.500
Knowledge sector	a_W	0.500	0.500	0.500
Knowledge sector	δ_W	0.005	0.005	0.005
Capital market	δ_K	0.025	0.025	0.025
Capital market	ζ	0.500	0.500	0.500
Capital market	φ	0.990	0.990	0.990
Labour market	θ	0.300	0.300	0.300
Labour market	g_A	0.001	0.001	0.001
Labour market	g_N	0.001	0.001	0.001
Neighbour relations	ξ	0.100	0.100	0.100
Neighbour relations	r	1/2	1/2	1/2

Table 4.1: Parameter setting

Further, one has to choose arbitrary starting values for the stock of knowledge, labour and capital. With the exception of knowledge W_0, which is random and distributed uniformly in the interval $[0, 0.5]$ all variables of interest are set to $K_0 = L_0 = 1$ for all regions i. Thus, the regions differ only with their initial endowment of knowledge $W_0^i \neq W_0^j$.

4.2.4 Simulation results

This section provides an overview of the simulation results. Results are presented both for first order ($r = 1$) and second order ($r = 2$) spatial influence. Simulations have been performed using `Matlab 6.5.0`. [29]

[29]The program is available on request.

4.2.4.1 First order spatial influence

The first simulation has been run for the case of decreasing returns to scale (DRS) scenario. As we can see from figure 4.2 we do not observe an agglomeration tendency for this case after 200 periods.[30] As a consequence of that, the Gini-coefficient as well as the spatial concentration should be rather low for labour productivity, which can be seen from figure 4.2. As a result, decreasing returns to scale do not display relevant agglomeration tendencies within our framework.

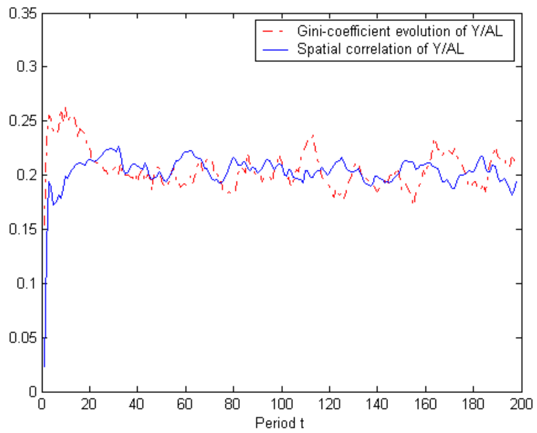

Figure 4.2: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for DRS and $r = 1$

For the second simulation (figure 4.3) we assume constant returns to scale (CRS). On contrary to the before discussed case, we observe a spatial concentration of the per capita income after 200 iteration steps. The Gini-coefficient exhibits a higher value on average compared to the DRS scenario, which means that distribution of per capita income tends to be more unequal as in the DRS scenario.

The last simulation (figure 4.4) has been done for the increasing returns to scale case (IRS). The conspicuous fact is, that we can observe a strong agglomeration tendency right from the beginning of the simulation. After 200 simulation runs we observe only a few regions which exhibit a high per capita income relative to the rest of the world. This is in line with the fact that the Gini-coefficient indicates a strong uneven income per capita distribution.

[30] Spatial correlation is measured similarly to time series analysis context with the so called Moran's-$\mathcal{I}$.

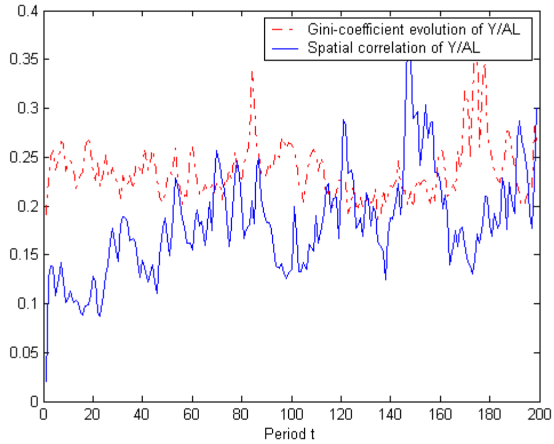

Figure 4.3: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for CRS and $r = 1$

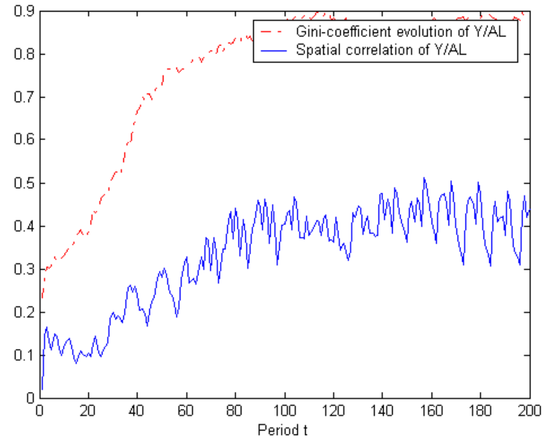

Figure 4.4: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for IRS and $r = 1$

4.2.4.2 Second order spatial influence

In this section, we perform the same simulations as done before in the preceding section with respect to the fact that second order neighbour influence matters. The intuition is, that second order spatial influence leads to a stronger spatial correlation of per capita income, because of the fact, that more regions benefits from knowledge spillovers. Further, the Gini-coefficient should exhibit a more equal distribution, also due the fact, that more regions can benefit from knowledge spillover pool. Simulation scenarios can be found in figures 4.5, 4.6 and 4.7.

First, the simulation of the DRS case has been performed. Compared to DRS scenario with $r = 1$, we observe, that spatial correlation is higher but at the same time income per capita is more evenly distributed as for the case of first order spatial effects.

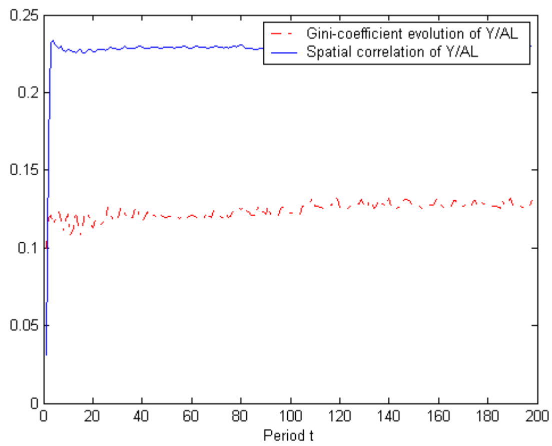

Figure 4.5: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for DRS and $r = 2$

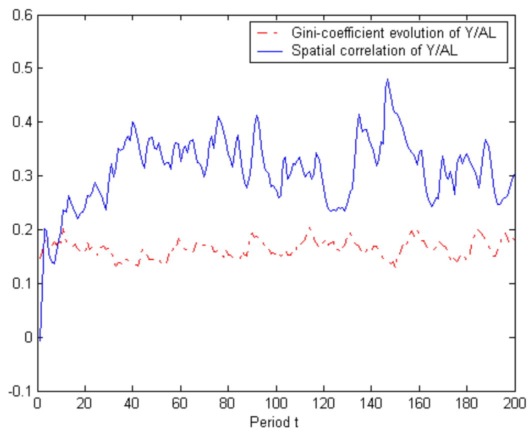

Figure 4.6: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for CRS and $r = 2$

Second, if we compare the CRS scenario for $r = 1$ with the CRS scenario with $r = 2$ we conclude, that income per capita distribution is more evenly distributed for the case of second order spatial influence.

Third, only for the IRS case, we observe no relevant differences between the first and second order spatial influence scenario. Although the obtained results are based on one particular parameter constellation, unreported sensitivity analysis indicate that the obtained results hold more generally.

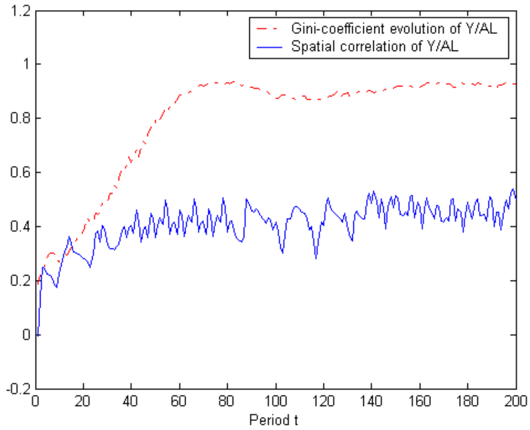

Figure 4.7: Evolution of Gini-coefficient and spatial correlation of $\frac{Y}{AL}$ for IRS and $r = 2$

4.2.5 Conclusion

The aim of the model derived above is to investigate the relationship between knowledge diffusion, agglomeration and growth. From a theoretical growth literature view, only the link of technological innovations and knowledge diffusion for technological growth is widely discussed[31], while the role of knowledge diffusion is only partly considered. Ex ante, the so called North-South trade model seems appropriate to cope with this research question. Some of the North-South trade literature on diffusion and technological progress[32] consider feedback effects between the North and the South in the steady state, but an analysis of the transitional dynamics for either region is missing. (Barro and Sala-I-Martin, 1997) indeed derived transitional dynamics for the South but feedback effects are excluded as there is no trade of intermediate goods. Thus, a transition path for the North cannot be derived.

The communality of this strand of literature is only focused on two country or two region models, which consist of a rich North and a poor South or a core and a peripheral country. From this perspective, those type of models are less suitable to investigate the link of increasing returns to scale, agglomeration and distribution of economic numbers because in a two country framework, it is not reasonable for instance to investigate agglomeration effects over regions. From this point of view, those North-South models are not appropriate to give a justification of the "folk theorem of spatial economics" which states that increasing returns to scale are essential for explaining agglomeration effects and thus uneven geographical distribution of economic numbers.

To investigate the relationship between knowledge diffusion, agglomeration and growth

[31]Refer to (Romer, 1986), (Romer, 1990) and (Krugman, 1991) for instance.

[32]Refer to (Krugman, 1979), (Dollar, 1986), (Grossman and Helpman, 1991b), (Grossman and Helpman, 1991a), (Rivera-Batiz and Romer, 1991), (Barro and Sala-I-Martin, 1997) and (Glass, 1997).

one has to refer to a multi country framework. One of the reasons, why multiple country or regional focused growth models are less attractive could be that such growth models become very complex and cannot solved analytically. For computational reason, a *Cellular Automaton* framework has been used to simulate the before established model. This environment has been selected because of its ability to visualize spatial effects.[33].

The aim of the theoretically derived model, which is based on the works of (Uzawa, 1965) and (Lucas, 1988), is to derive a theoretical growth orientated justification of the "folk theorem of spatial economics"[34], that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[35]. For this reason, a world consisting of 100 regions has been simulated to study the effects of decreasing returns to scale, constant returns to scale and increasing returns to scale, both in the goods sector and in the R&D sector on the per capita production in each region. To measure inequality over regions, we refer to the Gini-coefficient. Further it was distinguished between first order and second order spatial effects to control for different grasps of knowledge spillover.

After performing two simulation scenarios, it was found that productivity is more evenly distributed the higher the degree of spatial effects is, et vice versa. Second, spatial dependence is higher, the higher the degree of spatial effects is. Third, a strong unevenly productivity distribution results only for the case of increasing returns to scale, for any degree of spatial effects. Thus, the "folk theorem of spatial economics"' seems to be justified within this model framework.

Of course, there are various avenues for further research. One of the possible research fields is, to embed the (CA) modelling technique in a general equilibrium framework. Further, the question how (weak) scale effects in per capita production affects the per capita production distribution of regions should be investigated deeper in further research.

In the next section, an empirical model is set up which tries to identify spatial agglomeration effect in German regions.

[33]Refer to (Keilbach, 2000).
[34]Refer to (Scotchmer and Thisse, 1992).
[35](Fujita and Thisse, 2002), p. 342.

4.3 Empirical model

In the foregoing section we have argued that spatial knowledge spillovers can explain agglomeration phenomena in real economics. Increasing returns to scale play the key role in explaining spatial concentration that we call cluster in general. The aim of this chapter is to give an answer to the question, whether spatial spillovers can be identified in the data and if spatial heterogeneity matters. Further the questions, how do these knowledge spillovers, given they exist, affect labour productivity and are knowledge spillovers more local or more global are, should be answered. Finally we want to filter spatial effects, if necessary.

4.3.1 Motivation

The basic cross section regression model stems from a simple production function approach and can be written as follows:

$$y = X\beta + \epsilon, \tag{4.19}$$

where y is a stochastic $N \times 1$ vector of observation, X is a full rank $N \times K$ matrix of K non stochastic independent variables, β is a $K \times 1$ vector of regression coefficients and ϵ is treated as a normally and independently distributed $N \times 1$ vector of errors. The drawback of a formulation like equation 4.19 is, that it does not acknowledge spatial dependence. But if spatial dependence, especially spatial autocorrelation, exist in the data, and if they are neglected within the estimation setup 4.19, an estimation based on OLS may not be consistent[36]. This argumentation is familiar when talking about estimation problems within a pure time series approach.

Therefore, equation 4.19 has to be altered and expanded for spatial processes. Generally, spatial events appear in three forms: first, spatial dependence is only observed in the y vector. As a consequence of that, a spatial lag model or a spatial AR(1) model has to be estimated. Second, spatial dependence is only observed in the error term vector ϵ, which means that one has to model a spatial error or a spatial MA(1) model. Or third, a combination of both spatial events occur in the data. Then a mixture of a spatial lag and a spatial error model has to be used. Given the latter is true, then we can rewrite equation 4.19 as a spatial ARMA(1,1) model as follows:

[36] Refer to (Anselin, 1988) and (Anselin and Rey, 1991) and appendix 1.

$$y = \rho W y + X\beta^{X} + \tilde{X}\beta^{\tilde{X}} + \lambda W \epsilon + \kappa, \tag{4.20}$$

with $X = [x_1, x_2, ..., x_K]$, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_M]$ and the $K \times 1$ vector $\beta^X = [\beta_1^X, \beta_2^X, ..., \beta_K^X]$, and the $K \times 1$ vector $\beta^{\tilde{X}} = [\beta_1, \beta_2, ..., \beta_M]$.

The parameter ρ is the so called spatial autoregression coefficient, W is a $N \times N$ matrix containing spatial weights, and κ is a $N \times 1$ vector containing errors. Often it is assumed that $M = K$. Thus, a close relationship between time series and spatial econometrics modeling can be observed. But it is worth to note, that the analogy regarding the labeling of such a process to time series is misleading sometimes because spatial spillovers are often described by feedback-processes, as mentioned before.

The $N \times K$ matrix X contains non spatial exogenous variables, whereas the $N \times M$ matrix $\tilde{X}$ contains the spatial lagged exogenous variables. Of course we can write $\tilde{X} = WX$. Stacking Wy, X, $\tilde{X}$ and $W\epsilon$ in $\tilde{X}^+ = [Wy, X, \tilde{X}, W\epsilon]$ and $\tilde{\beta} = [\rho, \beta^X, \beta^{\tilde{X}}, \lambda]'$ leads to

$$y = \tilde{X}^+ \tilde{\beta} + \kappa. \tag{4.21}$$

Although it is common to assume that $\kappa \sim N(0, \sigma^2 I)$, it is more plausible to assume that $\kappa \sim N(0, \sigma^2 \Omega)$ with $\sigma_i = \tilde{h}(f_i' \alpha)$ and $h(\cdot) > 0$ as unknown, continuous function which are treated as the diagonal elements of the error covariance matrix $\sigma^2 \Omega$.

Although (Keilbach, 2000) and (Klotz, 1996) argue that spatial heterogeneity is not seen as a serious problem in spatial econometrics context it should be in fact treated as a serious problem ex ante. Remember for instance that some regions do not follow the same spatial relationships as other regions. This "enclave effects" or in an econometric notation, these "outliers" could cause severe problems such as fat-tailed errors which are not normal of course. A t-distribution is more appropriate then. In such cases it seems more appropriate to acknowledge these outliers and use Bayesian methods for instance.

Only for the fact that $\tilde{h} = \sigma^2$ it follows that $\kappa \sim N(0, \sigma^2 I)$ which implies spatial homogeneity. The big problem estimating a heterogeneous spatial model is that allowing for heteroscedasticity we have to estimate N additional parameter for each σ_i. Of course, this leads to the so called "degree of freedom" problem, because we do not have simply spoken enough observations to compute an estimate for every point located in space. Therefore, we are confronted with a problem using the "traditional" econometricians toolbox. One way to deal with this problem is to refer to Bayesian econometrics. Bayesian methods in regression context do not encounter the similar degree of freedom problems, because informative priors are available. As seen later, the prior distribution for our N diagonal elements of Ω are independently $\frac{\chi^2(s)}{s}$ distributed. Note, that the χ^2-distribution is a single parameter distribution where we can represent this parameter as s. This allows us to estimate N additional parameter of the diagonal elements of Ω by adding a single parameter r to our regression procedure.

Hence, the estimation strategy is defined as follows: one should start with an estimation of a spatial ARMA-model with homogeneous errors based on equation 4.20. Of course, expression 4.20 can be considered also as a spatial ARIMA-model, if $|\rho| = 1$. If we do observe a significant coefficient of ρ close to one[37], one should estimate a spatial ARIMA-model to avoid results based on spurious regressions. Equation 4.20 can be consistently estimated via Maximum-Likelihood (ML) as mentioned by (Anselin and Rey, 1991). Please note again, that (ML) based models are not suitable to model spatial heterogeneity. For this reason, (ML) estimations implicitly assume spatial homogeneity. For this reason, Bayesian models with the additional assumption of heterogeneous errors should be estimated. After performing model selection mechanism, a direct model comparison of the (ML) based and the Bayesian model should be used, to find the model which best fits to the data generating process. If one detects dissimilarities between the two approaches, then one of course should rely on the Bayesian model than on the (ML) approach.

4.3.2 Spatial weight

Until today, there is no theory about how to find the "correct" spatial weight matrix W. Therefore, the choice of the spatial weights should be done on the basis of the specific research topic. The first question one has to ask is how to proxy spatial proximity. One approach is to say, that spatial proximity is best proxied by geographical distances. Another way is to say, that geographical boarders are less important for spatial proximity and for this reason one should better rely on non geographical data, such as trade shares[38] or data on FDI[39].

The latter strategy has two major drawbacks in this context: First, in this work, it is primarily focused on knowledge diffusion. When talking about this issue it is rather not intuitive to proxy spatial proximity by trade shares or FDI data for instance. Second,

[37] Naturally, ex ante it is difficult to decide, whether one is confronted with a highly persistent or an unit root process with respect to space.

[38] Refer to (Coe and Helpman, 1995).

[39] Refer to (Lichtenberg and van Pottelsberghe de la Potterie, 1996).

there is a methodological problem: using these weights it is very likely, that they can be endogenous and therefore lead to biased estimators if not using an IV or GMM approach.

Hence, the majority of the literature is refering to more geographical weights. It is common using geographical distances (Keller, 2001) or more precisely using great circle distances between regions' centroids (Anselin, 1988). But this has the inherent assumption that knowledge spillover sources are located in region´s centroids. Another way, which is also consulted in this study, is simply to refer to binary weighting schemes [40]. If a region i is a neighbour of another region j, then the i-th element of W, w_{ij} takes a 1, otherwise a 0.

Thus, we can write for the symmetric $N \times N$ matrix W with weights w_{ij}:

$$w_{ij} = \begin{cases} 1, & \text{if i and j have a common border and i}\neq\text{j} \\ 0 & \text{otherwise} \end{cases} . \tag{4.22}$$

Often, this matrix is weighted or standardized because this facilitates the interpretation of the estimated coefficients[41] and guarantees that the Moran's $\mathcal{I}$ is situated in the interval $[-1;1]$[42]. Using the weighting scheme, proposed by (Anselin, 1988), we write for the standardized elements w_{ij}^{+} of W^{+}:

$$w_{ij}^{+} = \frac{w_{ij}}{\sum_{j=1}^{N_j} w_{ij}}. \tag{4.23}$$

In this way we have created a row standardized spatial weighting matrix W^{+} which is used in the preceding estimation exercise.

4.3.3 Higher order spatial influence specification

One major drawback of model 4.20 is, that higher order spatial dependencies are not included. To obtain a higher order weighting matrix W^{+r} for $r = \{1, ..., R\}$ we should increase the power of the simple contiguity matrix [43]. Labelling the order of the spatial dependency with $r = \{1, 2, 3, ..., R\}$ then $\tilde{X}$ can be expanded as follows:

[40] Refer to (Tappeiner et al., 2008).
[41] (Anselin, 1988), p. 23.
[42] Refer to (Ord, 1975) and (Griffith, 1996).
[43] Refer to (Anselin, 1992).

$$\tilde{X}^{++} = \left[\begin{pmatrix} \tilde{x}^1_{11} & \tilde{x}^1_{12} & \cdots & \tilde{x}^1_{1M} \\ \tilde{x}^1_{21} & \tilde{x}^1_{22} & \cdots & \tilde{x}^1_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}^1_{N1} & \tilde{x}^1_{N2} & \cdots & \tilde{x}^1_{NM} \end{pmatrix}, ..., \begin{pmatrix} \tilde{x}^R_{11} & \tilde{x}^R_{12} & \cdots & \tilde{x}^R_{1M} \\ \tilde{x}^R_{21} & \tilde{x}^R_{22} & \cdots & \tilde{x}^R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}^R_{N1} & \tilde{x}^R_{N2} & \cdots & \tilde{x}^R_{NM} \end{pmatrix} \right], \tag{4.24}$$

or in short hand notation:

$$\tilde{X}^{++} = [\tilde{X}^1, \tilde{X}^2, ..., \tilde{X}^R]. \tag{4.25}$$

Defining $P = [\rho^1, \rho^2, ..., \rho^R]'$, $\tilde{y} = [W^{+1}y, ..., W^{+R}y]$, $\Lambda = [\lambda^1, \lambda^2, ..., \lambda^R]'$ and over the more $\tilde{\epsilon} = [W^{+1}\epsilon, ..., W^{+R}\epsilon]$, and $\beta^{++} = [\beta^{\tilde{X}^1}, \beta^{\tilde{X}^2}, ..., \beta^{\tilde{X}^R}]'$ with $\beta^{\tilde{X}^r} = [\beta_1^{\tilde{X}^r}, ..., \beta_M^{\tilde{X}^r}]$ we can rewrite our model 4.20 as:

$$y = \tilde{y}P + X\beta^X + \tilde{X}^{++}\beta^{++} + \tilde{\epsilon}\Lambda + \kappa \tag{4.26}$$

with $\kappa \sim N(0, \sigma^2\Omega)$. For $R = 1$ model 4.20 follows directly. From the general model 4.26 we can derive three major submodels for $r = \{1, ..., R\}$: the spatial lag (SAR(r)) and spatial error (SEM(r)) and a spatial model with exogenous spatial variables (SEV(r)). For the (SAR(r)) we can write:

$$y = \tilde{y}P + \kappa \tag{4.27}$$

with $\kappa \sim N(0, \sigma^2\Omega)$, for the (SEM(r)) we can write

$$y = X\beta^X + \tilde{\epsilon}\Lambda + \kappa \tag{4.28}$$

with $\epsilon = \tilde{\epsilon}\Lambda + \kappa$ and $\kappa \sim N(0, \sigma^2\Omega)$ and for the (SEV(r)) we notate:

$$y = X\beta^X + \tilde{X}^{++}\beta^{++} + \kappa \tag{4.29}$$

with $\kappa \sim N(0, \sigma^2\Omega)$.

It has to be pointed out, that the estimation of 4.26 and its submodels 4.27, 4.28 and 4.29 could lead to biased and inconsistent OLS estimates. Take submodel 4.27 for instance: $P\tilde{y}$ is correlated not only with κ but also with neighbourings κ. If all elements of $\tilde{y}P$ are zero OLS estimates are unbiased but inefficient. If submodel 4.29 is chosen, then the model contains only exogenous spatial lagged variables besides non spatial lagged exogenous variables. In this case OLS is only BLUE if $\kappa \sim N(0, \sigma^2 I)$. OLS is even more unbiased if estimating a spatial error model, thus referring on submodel 4.28.[44]

To test this spatial model, we regress the regional output, measured as gross value added on regional R&D-effort, human capital, regional number of patent applications, regional capital stock, regional number of low qualified labour force, regional infrastructure, spatial weighted gross value added, spatial weighted dependent variables and a West-East dummy, which covers the fact that East German regions are less productive than West German regions. Additionally, the number of patent applications are regressed on regional R&D output, as proposed by (Griliches, 1979). In this way it is possible to cover "articulated knowledge" and "tacit knowledge".[45]

The question which remained unanswered is, how to choose the order R. If one refers to the literature there is no hint how to choose the order R. Regarding this subject, (Anselin, 1992) argues that especially for small samples the order of the weighting matrix W should be chosen small. As mentioned above, in this investigation we base the order of R on the data, especially on Moran‘s $\mathcal{I}$. But before checking the data concerning spatial dependencies, we should throw a first glance at the data.

4.3.4 Data and variables

Before testing the model, which has been introduced in the preceding chapter, one has to give a short description of the data. As mentioned before, NUTS-2 data for all German regions for the year 2003 have been used. The reason why one should decide to base the empirical study upon NUTS-2 data is, that referring on so called "Kreisdaten" could result in spurious spatial dependence, which could be caused by streams of commuters, for example.[46] This problem is boosted by the empirical fact of suburbanization, which has increasingly appeared in the last years.[47] That is why most similar research field studies refer to so called "land use planning units", such as NUTS-regions, particularly for European studies or "Arbeitsmarktregionen" for German investigations. Whatever of the latter mentioned spatial unit one decides to use, the worth mentioning communality is, that a "land use planning unit" subsumes smaller subgroups, such as "Kreise". Thus, referring to "land use planning units", the

[44]Refer to appendix 1 for a proof.
[45]Refer to (Maurseth and Verspagen, 2002).
[46](Keilbach, 2000), p. 120-121.
[47]Refer to (Kühn, 2001) and (Kaltenbrunner, 2003) for a discussion.

spurious spatial dependence problem is from less importance or even canceled out. The year 2003 was selected because of reliability and accessibility of European patent data. Particularly the problem of missing data is serious for NUTS-2 data. Of course, if data would have been available for a longer period of time, then regression based on time averages would be the appropriate approach.
In the table 4.2 one finds listed the German NUTS-2 regions which are subject of this investigation. I have decided to mark Berlin as a West German NUTS-2 region, because of its both historic and economic exceptional position. If we look at table 4.2 then 31 West German and 8 East German NUTS-2 regions have been detected in the dataset.

Code	German NUTS-2 region	Location
de11	Stuttgart	West
de12	Karlsruhe	West
de13	Freiburg	West
de14	Tübingen	West
de21	Oberbayern	West
de22	Niederbayern	West
de23	Oberpfalz	West
de24	Oberfranken	West
de25	Mittelfranken	West
de26	Unterfranken	West
de27	Schwaben	West
de30	Berlin	West
de41	Brandenburg-Nordost	East
de42	Brandenburg-Südwest	East
de50	Bremen	West
de60	Hamburg	West
de71	Darmstadt	West
de72	Gießen	West
de73	Kassel	West
de80	Mecklenburg-Vorpommern	East
de91	Braunschweig	West
de92	Hannover	West
de93	Lüneburg	West
de94	Weser-Ems	West
dea1	Düsseldorf	West
dea2	Köln	West
dea3	Münster	West
dea4	Detmold	West
dea5	Arnsberg	West
deb1	Koblenz	West
deb2	Trier	West
deb3	Rheinhessen-Pfalz	West
dec0	Saarland	West
ded1	Chemnitz	East
ded2	Dresden	East
ded3	Leipzig	East
dee	Sachsen-Anhalt	East
def0	Schleswig-Holstein	West
deg0	Thüringen	East

Table 4.2: List of German NUTS-2 regions

The data stem from the online database provided by Eurostat, from the online support of the German statistical office in Wiesbaden (genesis online), from the online representation of the "Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der Länder"" as well as from the INKAR-database CD-Rom published by the "Bundesamt für Bauwesen und Raumordnung".

In detail, the following variables are specified:

1. ***Output (Y)*** is approximated with Gross Value Added. The data are published annually on the CD-Rom "Statistik regional" by the "Statistische Ämter des Bundes und der Länder" and have been stated in Mio. Euros.
2. ***Human capital (H)*** is measured as the percentage of the employees on NUTS-2 level, subject to social insurance contribution, who obtained a high level degree, such as an university, a polytechnical or a technical college degree. With the exception of Sachsen-Anhalt, the data stem from the CD-Rom "Statistik regional" edited by the "Statistische Ämter des Bundes und der Länder".[48] To exclude the above mentioned commuter problem, the data correspond to the activity area, not to the place of residence of the employees. Naturally, this assumption comprises that added value is created at the activity area. Unfortunately, the data do not exhibit the desirable attribute that they are restricted to the employed human capital in production sector. Hence, as mentioned by (Keilbach, 2000) we have to bear in mind implicit spillovers of employed human capital from the non-producing sectors.
3. ***Labour (L)*** is measured as number of employees in thousands on NUTS-2 level subject to social insurance contribution less human capital, defined above. Data have been obtained from the CD-Rom "Statistik regional" edited by the "Statistische Ämter des Bundes und der Länder".
4. ***Capital (K)*** stock construction for the regional NUTS-2 manufacturing sector is a serious problem. By mischance, it is not possible to hark back to regional disaggregated stock of capital data for NUTS-2 regions from the official statistic suppliers. Only for the German "Bundesländer" the "Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der Länder"" offers capital stock data. Naturally, on this rather aggregated level, capital stock estimation via the perpetual inventory method (PIM) is rather easy to implement. The fundamental idea of PIM is that different vintages of the stock of capital exhibit different efficiencies

[48] The data of Sachsen-Anhalt have been obtained directly from the "Statistisches Landesamt Sachsen-Anhalt".

beeing used in the production process. This idea has to be acknowledged when calculating the stock of capital. Therefore, one has first to determine the average durability of an asset. Next, calculation of long term data regarding the annual investments is needed to initialize PIM. Basically, it is common to refer to gross fixed capital formation as a proxy, because PIM is nothing else than computing an average weighted sum of past investments. Long term data are necessary, especially in the case of Germany, because the cumulative investment data have to be corrected using a survival function and depreciation function to obtain an estimation for the capital stock. In the case of Germany the Gamma distribution has to be consulted to get a measure for the mortality function from which the depreciation function can be obtained directly[49]. Based on the gamma function, it cannot be ruled out ex ante that the service life of an asset oscillates more than twice of an average service life of an asset. That is exactly the reason why it is strongly recommended to use long investment data[50]. In this way it is possible to calculate the stock of capital K in period t using data of gross fixed capital formation I from the period $t+1$, a depreciation rate on stock of capital δ, obtained from the depreciation function and an average growth rate ζ of gross fixed capital formation. In a more formal manner the following relationship results[51]:

$$K_t = I_t \sum_{\kappa=0}^{\infty} \left(\frac{1-\delta}{1+\zeta} \right)^{\kappa} = \frac{I_{t+1}}{1+\zeta} \frac{1}{\left(1 - \left(\frac{1-\delta}{1+\zeta}\right)\right)} = \frac{I_{t+1}}{\zeta+\delta}. \tag{4.30}$$

As mentioned above, long term data for the gross fixed capital formation are needed to initialize PIM. Unfortunately, long term series of desired data are not available for Germany on NUTS-2 level. EUROSTAT offers data for gross fixed capital formation on NUTS-2 level for German regions only for the years 2002 and 2003.[52] Concerning the above mentioned, it is not reasonable to rely on PIM estimating the stock of capital for NUTS-2 regions.

Because of that reason, the estimation of the stock of capital is done with a method similar to the shift analysis. The basic idea of the shift analysis is to compute a so called structural factor and a location factor. The structural factor should provide information about the capital intensity of branches and

[49] It is suitable to set the dilation parameter of the mortality function to the value $p = 9$.

[50] The starting date for series of gross fixed capital formation is 1799 for buildings, 1899 for machinery and equipment and 1945 or later for intangible assets. For an deeper introduction of PIM, particularly for Germany, consult (Schmalwasser and Schidlowski, 2006).

[51] Refer to appendix 2 for a deviation of expression 4.30.

[52] The Statistische Landesamt Baden-Wurttemberg offers data for the gross fixed capital formation from 1998 onwards online.

furthermore, should give a hint, whether capital intensive branches are over- or underrepresented in a specific region. Assume for the moment[53], that we have $i = \{1, 2, ..., I\}$ branches and $h = \{1, 2, ..., N\}$ NUTS-2 regions in each of the $j = \{1, 2, ..., M\}$ states. It is worth mentioning, that a single NUTS-2 region can represent an own state[54]. Due to the fact that we do not analyse specific branches, we can set $i = 1$.

Hence, we can notate the structural factor SF for region h in a formal manner:

$$SF_t^h = \frac{\sum_i g_{t-1,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t-1,i}^M} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M}, \tag{4.31}$$

where $I_{t,i}^M$ stands for the gross fixed capital formation in the state M in year t, $I_{t,i}^N$ stands for the gross fixed capital formation in the NUTS-2 region N in year t and $g_{t,i} \equiv \frac{I_{t,i}^N}{I_{t,i}^M}$ is the weight for region h. Of course, $\sum_h g_{t,i} = 1$ and $\sum_j \sum_h g_{t,i}$ must equal the number of the states M. In the case of Germany $M = 16$.

Instead of the structural factor, the location factor assumes implicitly, that a specific region, which can be characterized by high investments in the past, must exhibit a high stock of capital relative to other regions in the present. For the location factor LOF one can write:

$$LOF_t^h = \frac{\sum_i g_{t,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t,i}^M}. \tag{4.32}$$

Multiplying the structural factor with the location factor one obtains the regional factor. More formally spoken, we obtain the regional factor RF:

$$RF_t^h = \frac{\sum_i I_{t,i}^N}{\sum_i I_{t-1,i}^N} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M} = \frac{\sum_i g_{t,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t-1,i}^M} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M}. \tag{4.33}$$

A value greater than one for the regional factor implies, that a specific region has grown faster than the average, a value less one means, that a specific region has grown less than the average.

To calculate the weights for RF_t^h and LOF_t^h for every region h we have to consult data of gross fixed capital formation for 2003. After calculating the region specific weights, the regional capital stocks for the German "Bundesländer" are weighted

[53]Please bear in mind that the national form is only valid for presentation of shift analysis.

[54]This is true for Hamburg, Bremen, Berlin, Mecklenburg-Vorpommern, Schleswig-Holstein, Saarland, Thüringen and Sachsen-Anhalt in the case of German "Bundesländer".

with these. In this way, we have estimated NUTS-2 specific stocks of capital in Mio. Euros.

5. ***R&D (R&D)*** effort is expressed as the total R&D expenditure (GERD). The expenditures include the business enterprise sector, the government sector, the higher education sector as well as the private non-profit sector. Data have been expressed in Mio. Euros and have been provided by EUROSTAT. Obviously, relying on this data, we cannot exclude spillovers from the non-producing sector to the producing sector. As mentioned by (Keilbach, 2000) this effect should be neglectable. Although it would be reasonable on the first sight, we should not use R&D employees as a proxy for R&D, because it is justified to assume that within the R&D sector, more than in the manufacturing sector, the majority of offered jobs requires a high skilled labour force, a subset of human capital, defined above.

6. ***Patent (P)*** applications to the European Patent Office (EPO) by priority year at the regional level have been gathered from EUROSTAT. The priority starts after the year filing the patent application. Data are expressed as total number of patent applications in a specific NUTS-2 region.

7. ***Infrastructure (I)***: Since (Aschauer, 1989) there has been a intensively leading debate about how to measure infrastructure and what effects public infrastructure has on output growth using a production function approach. In general, the studies can be grouped in national level studies and regional or state level studies. One traditional approach is to use information about undeveloped areas serving for streets, railways or airways and traffic on waterways (Keilbach, 2000). Additionally, other factors, such as political interest, friendship ties, basic trust and quality of life etc. should flow into the regression context. Regrettably, these data are not available on NUTS-2 regions. Therefore, for this study on has to refer to data on highway density per squared kilometre published by EUROSTAT.

8. ***Density (DEN)*** is measured as inhabitants per square kilometre. Data for the average population for 2003 per NUTS-2 area as well as details for the NUTS-2-areas in square kilometre have been obtained from the CD-Rom "Statistik regional".

9. ***Dummy***: The dummy covers East-Western productivity differences. It is defined as follows:

$$d = \begin{cases} 1 & \text{if region } i \text{ belongs to the group of West German NUTS-2 regions} \\ 0 & \text{if region } i \text{ belongs to the group of East German NUTS-2 regions} \end{cases}.$$

It is reasonable to include the dummy, because a bulk of papers have found empirical evidence that a significant difference regarding the capital intensity still exists between East and West German region. After initial continuous progress concerning the productivity of East German regions right after the German reunification and an observed stagnation in the years 1996 and 1997 this gap seems to widen again in recent years.[55] For instance (Smolny, 2003) has found that East German capital intensity is 80% of corresponding West German capital intensity.

4.3.5 A first hint for spatial knowledge diffusion: a descriptive view

After describing the data set, this section should provide us a first guess concerning the existence of knowledge diffusion phenomena in the data. The traditional way detecting spatial phenomena in the data is to compute the so called Moran‘s $\mathcal{I}$, which is defacto "the" standard instrument in spatial econometrics for detecting spatial correlation[56] coefficient.[57]

The interpretation of the spatial correlation coefficient based on Moran´s $\mathcal{I}$ is a priori similar to time series analysis context. But it is not the same: Autocorrelation in time series means proximity of variables in time. Autocorrelation in space instead means geographic proximity of variables which is often two-dimensional. The important difference between the time series and the spatial econometric context is that spatial correlation has the attribute that a spatial event can be described via feedback loops, whereas time series correlation goes only in one direction, that is time. The interpretation of spatial correlation is quiet easy: if negative spatial correlation is observed, then regions are dissimilar with respect to their economic performance, whereas if positive spatial correlation is observed, then regions are similar with respect to their economic performance. The aim of the Moran´s $\mathcal{I}$ analysis is to measure the strength of spatial correlation and to find a hint how far spatial correlation spreads.

The Moran´s $\mathcal{I}$ is defined as follows:

$$\mathcal{I} = \frac{N}{O}\frac{e'W^{+r}e}{e'e}, \tag{4.34}$$

with O as the sum over all elements in W^{+r} and N as the number of observations. Of course if $\frac{N}{O} = 1$ we have a row standardized weighting scheme. e are the residuals

[55]For the convergence debate of East German regions refer to the empirical based analysis of (Bellmann and Brussig, 1998), (Almus and Czarnitzki, 2003), (Klodt, 2000), (Smolny, 2003), (Sinn, 2000) and (Sachverständigenrat, 2005).

[56]Refer to (Moran, 1948) and (Moran, 1950).

[57]This is most used indice for detecting spatial phenomena. Despite Moran´s $\mathcal{I}$, other indices such as Geary´s $\mathcal{C}$ and Ripley´s $\mathcal{K}$. But the two latter are seldomly used.

obtained from an OLS estimation of a variable V on its spatial counterpart VW^{+r}. To center 4.34 around zero we follow (Ord, 1975) and standardize 4.34:

$$\tilde{\mathcal{I}} = \frac{\mathcal{I} - E(\mathcal{I})}{\sqrt{Var(\mathcal{I})}}, \tag{4.35}$$

with

$$E(\mathcal{I}) = \frac{N}{O}\frac{tr(V^{++}W^{+r})}{N-K},$$

and

$$Var(\mathcal{I}) = \left\{\frac{N}{O}\right\}^2 \frac{\{tr(V^{++}W^{+r}V^{++}W^{+r\prime}) + tr(V^{++}W^{+r})^2 + [tr(V^{++}W^{+r})]^2\}}{(N-K)(N-K+2)} - [E(\mathcal{I})]^2,$$

with $V^{++} = I - V(V'V)^{-1}X'$ as the projection matrix. In this way $\tilde{\mathcal{I}}$ is normal distributed.

Before computing Moran´s $\mathcal{I}$ for the desired variables, we should first have a look at the data. As mentioned above, we try to estimate a standard production technique to investigate the effects of spatial knowledge spillovers on labour productivity. Table 4.3 and 4.4 provide an overview of the data used in the analysis.

	Y	K	L	H
Mean	48795.52	266105.80	627841.30	8.49
Modus	–	–	–	–
Median	41022.01	228133.0	544004.00	8.38
Max	140902.40	895491.10	1603418.00	14.01
Min	9963.63	66538.54	135678.00	4.26
Std. Dev.	33057.33	177768.70	350356.20	2.66
Skewness	1.44	1.63	1.17	0.38
Kurtosis	4.23	5.66	3.62	2.15
Observations	39	39	39	39

Table 4.3: Table of descriptive statistics (I) of variables used for the analysis

From table 4.3 and 4.4 we can see that all variables exhibit positive skewness, what means that the distribution has a long right tail. This is especially true for the variable density (*Den*) but not astonishing, because we have a few high densely populated areas such as Berlin, Hamburg and Bremen. Additionally the distributions are peaked, which means they are leptokurtic relative to the normal distribution.

Additionally, we can see from table 4.5 the logarithmic variables which have been used in the regression analysis. Please remember that lower letters denote logarithmic

	P	I	R&D	Den	Dummy
Mean	330.06	43891.18	1333.87	432.76	–
Modus	–	–	–	–	1.00
Median	223.28	43586.00	612.69	211.60	–
Max	1486.63	76028.00	7035.16	3803.00	1.00
Min	26.16	4785.00	67.64	74.99	0.00
Std. Dev.	352.11	17383.47	1592.97	698.53	–
Observations	39	39	39	39	39

Table 4.4: Table of descriptive statistics (II) of variables used for the analysis

values: $v = ln(V)$, where $V = \{Y, K, L, H, P, R\&D, I, Den\}$ and $v = \{y, k, l, h, p, r\&d, i, den\}$ contains the corresponding logarithmic variables. As expected we find positive and high correlation between gdp per head, labour force and capital and positive correlation between human capital, R&D, infrastructure and patents. Furthermore, we can see from table 4.5 that the correlation between labour force and capital is nearly linear. Another interesting observation is that the ratio of capital to output is roughly constant. This observation reflects one of the Kaldor facts.

	y	k	l	h	p	i	r&d
y	1.0000						
	(0.0000)[◇]						
k	0.9719	1.0000					
	(0.0000)	(0.0000)					
l	0.9711	0.9527	1.0000				
	(0.0000)	(0.0000)	(0.0000)				
h	0.3519	0.3168	0.4006	1.0000			
	(0.0280)	(0.0494)	(0.0115)	(0.0000)			
p	0.8850	0.8840	0.8282	0.1944	1.0000		
	(0.0000)	(0.0000)	(0.0000)	(0.2357)	(0.0000)		
i	0.1811	0.1172	0.0704	0.1204	0.1432	1.0000	
	(0.2700)	(0.4774)	(0.6701)	(0.4653)	(0.3047)	(0.0000)	
r&d	0.8359	0.8069	0.7945	0.5741	0.8128	0.1687	1.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.7247)	(0.0000)

[◇]Variables in () denote the p-values of a test $H_0 := \rho = 0$. The p value is computed by a transformation of the correlation creating a t-statistic with $(N-2)$ df, with N as the number of rows of a matrix X containing the correlation coefficients.

Table 4.5: Variables scatter plot of correlation coefficients

From table 4.10 we find that the log values are rather normal distributed.

Cluster phenomena and spatial correlation are closely related. Therefore, coloured map plots are used to identify similar regions and separate them from more dissimilar regions. This is done also in this exercise. In upper left map of figure 4.8 we see some evidence that a difference regarding labour productivity between West German and

East German regions exist. But also West German regions itself are different from each other. Some regions in West Germany perform better than other regions in West Germany. Therefore, from an econometricians point of view it seems to be problematic to control only for West-East German differences. The question, whether German regions tend to converge or not is still unanswered. German research institutes found [58] that between 1993 and 1999 East German regions exhibit convergence tendencies, while (Bohl, 1998) found on basis of a panel unit root test for West German regions, that divergence cannot be ruled out. (Bröcker, 2002) concludes that the neoclassical convergence hypothesis cannot be disproved empirically. If we take a closer look at figure 4.8 we can find some high productivity clusters in the West and the South of Germany.

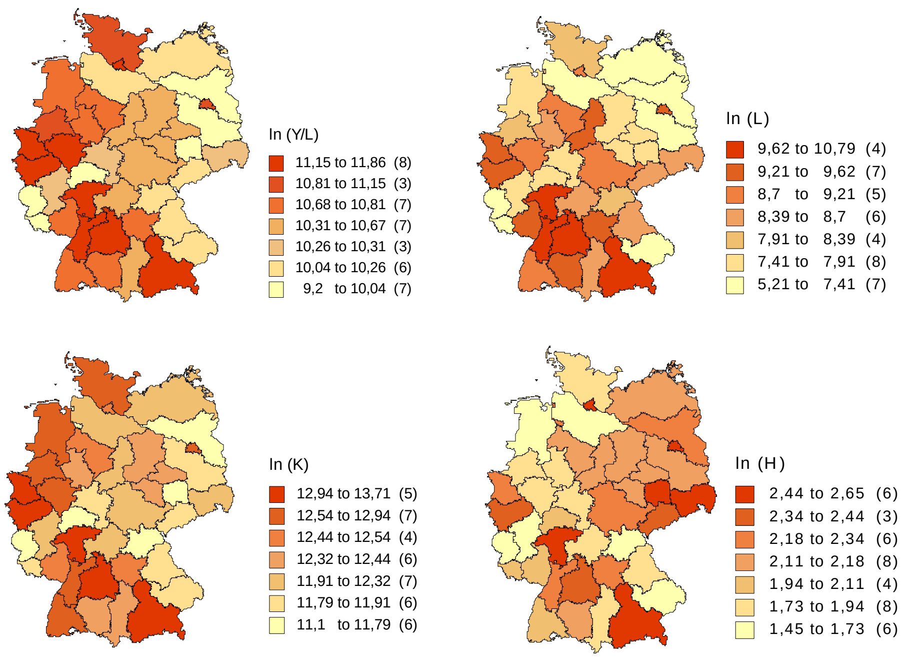

Figure 4.8: Spatial distribution of V (I)

In figure 4.8 we see that West German regions provide a higher stock of physical capital than East German regions. Again, some West German regions in the South and in the West can be characterized by the highest stock of physical capital. Picture 4.8 visualizes the labour force endowment of German NUTS-2 regions. From this picture we observe a very heterogeneous labour force endowment across German regions with a slight East-West differential. A rather astonishing and a priori contra intuitive impression we obtain, if we look at the human capital distribution over German re-

[58] Between 1993 and 1994 several research institutes such as DIW, IWH, IAB, IfW and ZEW observed convergence between East German neighboured NUTS-3 regions or counties convergence at all. For this topic refer to (DIW et al., 2002), p. 19.

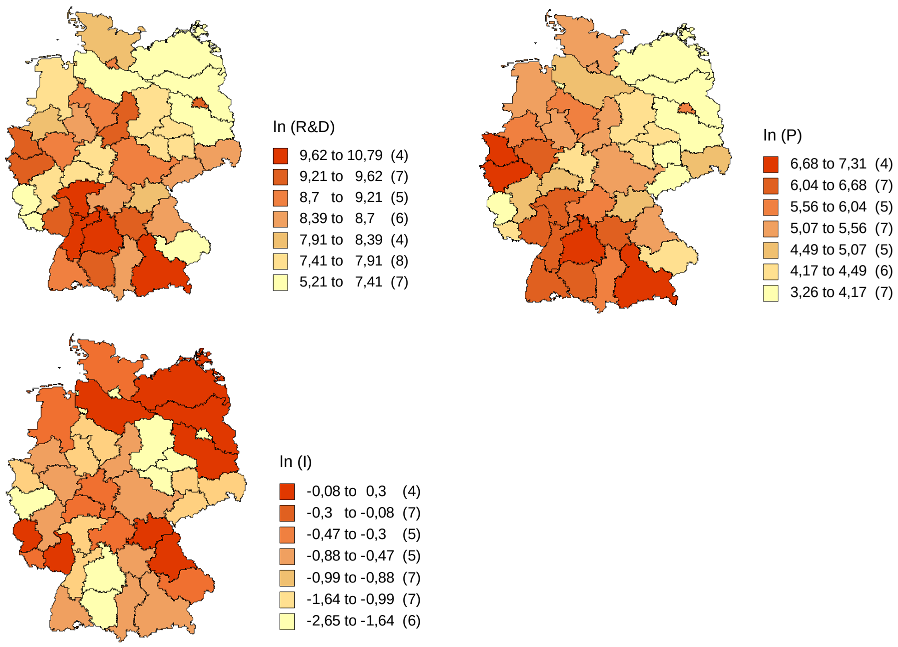

Figure 4.9: Spatial distribution of V (II)

gions. Especially, East German regions have a priori a leading role regarding human capital endowment. Some regions such as Thüringen and Sachsen for instance have a considerably higher human capital endowment than some West German regions, such as Saarland.[59] Particularly, for the case of human capital, East German regions are often confronted with a serious labour market mis match problem.[60] This mis match problem stems particularly from the R&D performance differential between West and East German companies. Defaulted or to a less extent undertaken private financed industrial research can be seen without any doubt as the central weakness of East Germany's research environment: only 4.4% of German R&D expenditures are alloted to East German regions, although it has to be mentioned that governmental financed R&D research tries to fill the gap. Regarding their total R&D reserach expenditures with 360 EUR per head East German regions clearly lie behind their West German neighbours with 659 EUR per head.[61] This impression is amplified if one refers to figure 4.9. Patents are often interpreted as a pre-stage for new products or new production methods. It is an important and often used indicator measuring innovation potential of regions. If we look at the upper right sub map of figure4.9 we observe a remarkable under performance of most East German regions regarding their patent activity. But this is also true for some North-West German regions. Only some regions

[59]Refer to an recent published study of (Hiero, 2008).
[60]Refer to (Kotschatzky et al., 2006), p. 15.
[61]Refer to (Pasternack, 2007).

in Sachsen and Thüringen perform better than the average of East German regions. As expected, we can conclude that this result correlates with the regional distribution of R&D expenditures and with the effective use of human capital in the production process. If we refer to infrastructure which is approximated by highway density, then it is striking at first sight that East German regions perform better than West German regions. But this is sophism. (Seidel, 2000) found that assets of East German roads, including high ways, have only reached 49.3% of West German niveau in 2000. If we look at the relationship between road density and road assets in East German regions, this discrepancy is much more dramatic: East German assets reached only 25% of West German assets. That makes clear, that East Germany has still a backlog with respect to infrastructural endowment.

After describing the data we are now ready to compute the Moran $\mathcal{I}$ coefficient for each variable to get an impression of what degree r regarding spatial knowledge spillover exert an significant influence. Thus, the degree r should yield a proxy for spatial distance with respect of knowledge spillovers.

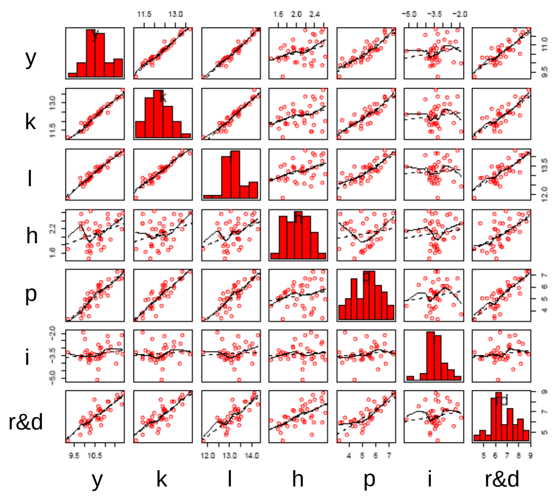

Figure 4.10: Scatter plot of variables used in the analysis

The computation of Moran‘s $\mathcal{I}$ for $v = \{y, h, p, r\&d, i\}$ is done with a program written in `R`, version 2.6.2.[62]. After completing the computation with `R`, figure 4.11 and 4.12 gives a graphical interpretation of spatial dependence between the v and its spatial lagged counterpart $W^{+}v$. Note that the variables are mean standardized, as mentioned before. Thus, besides a regression line the standardization allows us to plot one and two standard deviations areas. The interpretation of figure 4.11 and figure 4.12 is as follows: every subgraph is divided into four areas: the first area is located

[62]The source code is available on request.

in North-East direction, the second in North-West direction, the third in South-West direction and the last in South-East. The first area contains positive standard deviation from a region i and its corresponding neighbour j. On contrary the third area contains negative standard deviation from region i and its corresponding neighbour j. All other areas contain couples of negative standard deviation of region i and positive standard deviation of region j and vice versa. Thus, we have a positive spatial correlation if regions are located in the first and in the third area. Otherwise we have a negative spatial correlation. With other words: If the slope in the scatter plot is negative that means that we have a sort of checkerboard pattern or a sort of spatial competition in which high standard deviation regions are clustered with low standard deviation regions. Alternatively, if the slope is positive, we find the contrary.

If we now have a look at figure 4.11 we see first that positive spatial correlation is significant on a 5% significance level for the output y and for the patents p. Despite the fact that $r\&d$, human capital h and infrastructre i exhibit positive spatial correlation as expected, the Moran‘s $\mathcal{I}$ is not significant on a 10% significance niveau for $r = 1$. Next, the degree of spillover is boosted to $r = 2$ and again the Moran‘s $\mathcal{I}$ coefficient for each variable is computed.

On the next step we take the weighting scheme to the power of two and additionally compute the Moran‘s $\mathcal{I}$ for every variable. The result of this computation can be found in figure 4.12. The interpretation is equal to the preceding analysis.

If we look at the sub pictures of figure 4.12 we find that only the spatial correlation of patents p is significant on a 10% significance niveau. All other variables do not exhibit significant spatial correlation. Therefore, we have to conclude that knowledge spillovers, proxied by p, h and $r\&d$ are limited regarding space and in consequence more or less local and restricted to the nearest neighbours. Hence, we should acknowledge first order and second order degree of knwoledge spillover in the regression analysis. Additionally, we see some evidence from figures 4.11 and 4.12 that spatial outliers exits[63], which implies that spatial heterogeneity matters.

4.4 Spatial model estimation

In this section a spatial model estimation strategy is introduced, which is an expansion of the proposed strategy by (Florax et al., 2003). Before introducing the new estimation method, the classic method of (Florax et al., 2003) for cross section analysis is briefly sketched. First, one has to start by estimating an initial model $y = X\beta + \epsilon$. Second,

[63]Outliers are defined as data points which are situated outside the 2σ area.

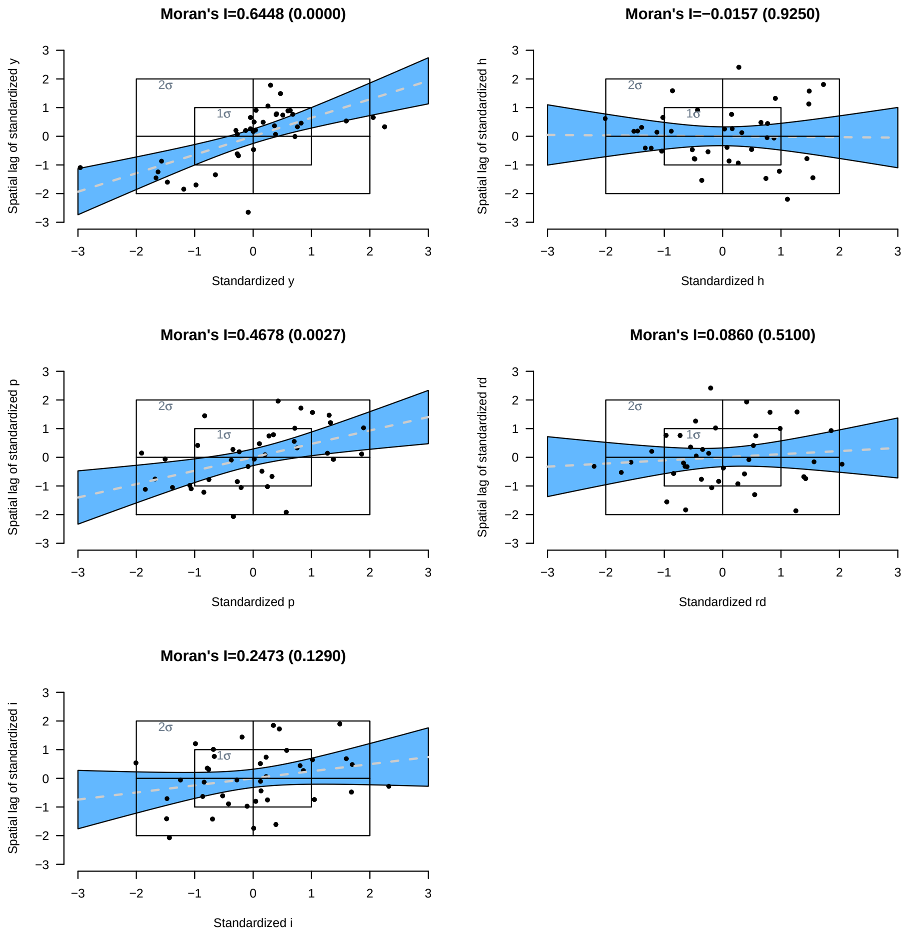

Figure 4.11: Computation of Moran's $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 1$

on the basis of the estimated model, Lagrange Multiplier tests are used to test for for spatial lag or spatial error model. If the null hypothesis is rejected, than spatial dependence matters and an appropriate spatial error or spatial lag model should be estimated. If we further acknowledge higher order spatial effects, the test statistic under the null hypothesis $H := \rho^r = 0, \forall r$ for $\mathcal{LM}_{\rho^r}$ can be written in the following way $r = \{1, ..., R\}$:

$$\mathcal{LM}_{\rho^r} = \frac{\left(\frac{e'W^{+r}e}{s^2}\right)^2}{T}, \tag{4.36}$$

with T as the trace of $(W^{+r\prime} + W^{+r})W^{+r}$, $e = My$ the residuals of regression, $M = I - X(X'X)^{-1}X'$ as the projection matrix and $s^2 = \frac{e'e}{N}$ as the estimated variance of the error term and N the number of observations. On contrary, the test statistic for

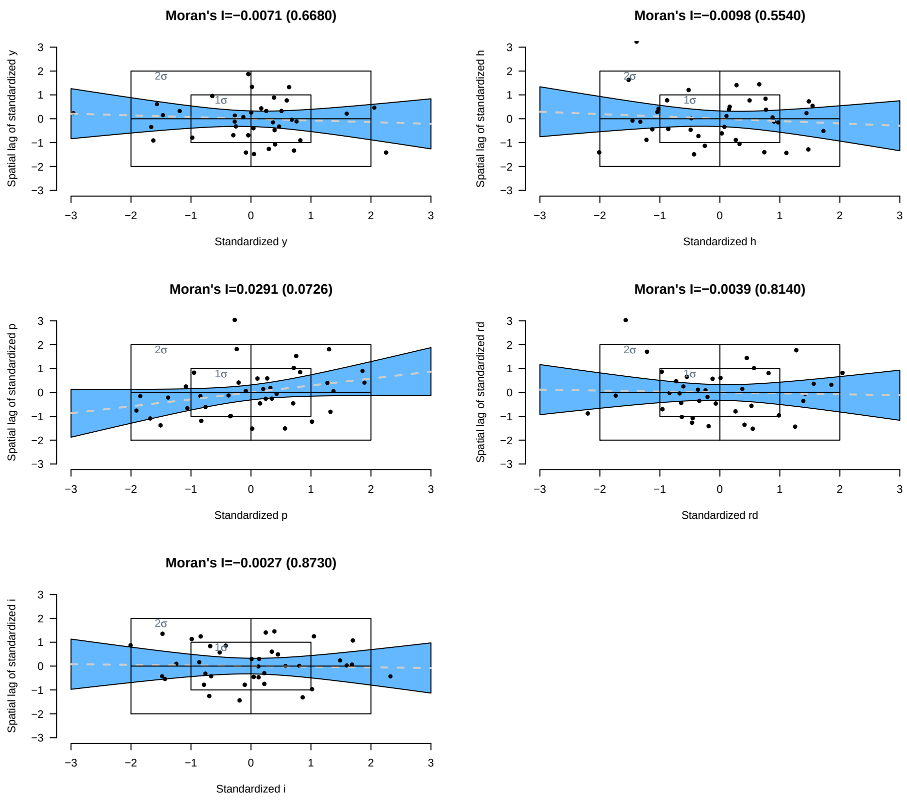

Figure 4.12: Computation of Moran‘s $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 2$

$\mathcal{LM}_{\lambda^r}$ is, given $r = \{1, ..., R\}$, under $H := \lambda^r = 0$, $\forall r$ can be written as:

$$\mathcal{LM}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}y}{s^2}\right)^2}{NJ}, \tag{4.37}$$

with $J = \frac{1}{Ns^2}\left[(W^{+r}Xb^{+++})'M(W^{+r}Xb^{+++}) + Ts^2\right]$ and b^{+++} as the OLS estimator of model 4.25.

Third, if for $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ each the null hypothesis cannot be rejected, then the initial model should be used. Otherwise one should compare both test statistics. If they are both significant, one has to compute additionally the robust versions of $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ to come to a final decision. If only one test is significant, then one has to adopt the initial model with respect to the significant test statistic.

The robust variant of $\mathcal{LM}_{\rho^r}$ read as:

$$\mathcal{L}\tilde{\mathcal{M}}_{\rho^r} = \frac{\left(e'W^{+r}y - \frac{e'W^{+r}e}{s^2}\right)^2}{NJ - T}, \tag{4.38}$$

For the robust variant of $\mathcal{LM}_{\lambda^r}$ we can write:

$$\tilde{\mathcal{LM}}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}e}{s^2} - T(NJ)^{-1}\frac{e'W^{+r}y}{s^2}\right)^2}{T[1-T(NJ)]^{-1}}. \tag{4.39}$$

If $\tilde{\mathcal{LM}}_{\rho^r} > \tilde{\mathcal{LM}}_{\lambda^r}$, then one should decide to estimate a spatial lag model otherwise if $\tilde{\mathcal{LM}}_{\rho^r} < \tilde{\mathcal{LM}}_{\lambda^r}$ then one should refer to a spatial error model. Given, only $\mathcal{LM}_{\rho^r}$ is significant but $\mathcal{LM}_{\epsilon}$ is not, then one should use a spatial lag model, otherwise, if $\mathcal{LM}_{\lambda^r}$ is significant, then a spatial error model should be chosen. Further, it should be kept in mind, that experimental based simulations by (Anselin and Florax, 1995b) and (Anselin et al., 1996) found evidence, that robust counterparts of the $\mathcal{LM}$-tests have more power in pointing out the appropriate alternative than the non robust $\mathcal{LM}$ versions. But as shown by (Florax et al., 2003), the classical top down approach, that means relying on the non robust $\mathcal{LM}$ test, outperforms the robust strategy in means of performance and accuracy. Thus, the same authors emphasise, that one should use the classic approach when testing for spatial effects. It should be further noted that, although this strategy is not theoretically justified yet, it is the only systematic approach of model selection in literature and used in empirical studies.[64]

The estimation strategy proposed by authors such as(Anselin, 2005) has three main drawbacks: first, the strategy lacks regarding their underlying tests hypothesis. For both tests, the LM_{ρ^r} and $\mathcal{LM}_{\epsilon}$ or in their robust form $\tilde{\mathcal{LM}}_{\rho^r}$ and $\tilde{\mathcal{LM}}_{\epsilon}$ the null hypothesis is either $H_0 := \rho^r = 0$ for $\mathcal{LM}_{\rho^r}$ or $\tilde{\mathcal{LM}}_{\rho^r}$ and $H_0 := \lambda^r = 0$ for $\mathcal{LM}_{\lambda^r}$ or $\tilde{\mathcal{LM}}_{\lambda^r}$. The null hypothesis $H_0 := \lambda^r = 0$ is realized in presence of ρ^r for the spatial error and $H_0 := \rho^r = 0$ in presence of λ^r for the spatial lag model. Although, robust $\mathcal{LM}$ tests are available, only one test is available, to compare the two models directly. This test, developed by (Mur, 1999) and (Trivezg, 2004) allows us to differentiate between spatial lag and spatial error models. But a drawback of the test proposed by (Trivezg, 2004) is, that it is only applicable for small samples, because it requires the computation of Eigenvalues and Eigenvectors of the underlying spatial weight matrix, which is cumbersome or even not possible for large data sets as noted by (Kelejian and Prucha, 1998).

Second, the strategy is exclusive in the way, that this strategy does not allow for a ARMA(p,q) model specification, which is as mentioned above, a combination of spatial lag and spatial error model. There is no reason, why one should exclude this combination ex ante. This could create a serious problem, because even if λ^r differs significant from zero but the robust $\mathcal{LM}_{\rho^r}$ test, which exceeds the value of the robust

[64]Refer to (Kim et al., 2003) for instance.

$\mathcal{LM}_{\lambda^r}$ statistic, suggests to model a spatial lag model, we should choose, going in line with (Florax et al., 2003), a spatial lag model. It is obvious, that there is an inherent potential of misspecification using the strategy proposed by (Florax et al., 2003).[65]

Third, both tests, if robust or not do not sufficiently control for heterogeneity of the error term nor do they cover the aspect of outliers. In other words, this methods neglect spatial heterogeneity entirely. Fortunately, spatial heterogeneity can be elegantly considered in an Bayesian approach.

Until today, Bayesian model selection criteria are seldom used in empirical applications. This might be due to three reasons: first, normally, spatial Bayesian model techniques are not included in standard econometricians tools, such as *EViews*. Second, these methods require extended programming techniques. In addition, their use for large sample applications is problematic, because then one is often confronted with numerical problems, especially in calculating the determinant of spatial weight matrix[66]. Third, Bayesian methods are often rejected or disregarded by the class of frequentest or "main stream" econometricians, mainly because of the Bayesian assumption that the vector of coefficients is treated as random, whereas the frequentest treat the vector of coefficients estimate as random.[67]

In this application, both views should be acknowledged, the frequentest based Maximum-Likelihood estimation techniques and Bayesian methods. It should be clear that both methods exhibit advantages and disadvantages, but to acknowledge them within the interpretation the strategy should improve the strategy of (Anselin, 2005), because of the above mentioned advantages of the Bayesian methods, especially their heteroscedastic formulation. The strategy can be formulated as follow:

1. *First, estimate the initial model via OLS.*
2. *Use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting potential spatial dependence. If the proposed tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 1. Otherwise, proceed with step 3.*
3. *If the null hypothesis of no spatial correlation is rejected, then expand the model estimated in step 1 by adding spatial counterparts of the independent variables. Perform an OLS estimation of this model.*
4. *Given the model setup in step 3, use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting po-*

[65]For example, assume $\tilde{\mathcal{LM}}_{\rho^r}$ statistic takes the significant value x and $\tilde{\mathcal{LM}}_{\lambda^r}$ statistic takes the significant value $x+\epsilon$, with a very small but positive value $\epsilon > 0$. In this case we conclude to use the spatial error model, because $\tilde{\mathcal{LM}}_{\lambda^r} > \tilde{\mathcal{LM}}_{\rho^r}$.

[66]To avoid this problem either rely on Bayesian methods or use the Monte Carlo based method proposed by (Barry and Kelley, 1999).

[67]See (Koop, 2003) for an excellent introduction to Bayesian Econometrics.

tential spatial dependence. If tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 3. Otherwise, proceed with step 5.

5. *Expand the model of step 3 with spatial error and spatial lag components. Again, Perform an OLS estimation of this model.*
6. *Use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting potential spatial dependence. If the tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 5. Otherwise, proceed with step 7.*
7. *Estimate a general spatial model (SAC) and separate spatial lag (SAR) and spatial error models (SEM) with MLE. OLS would yield in this case inconsistent parameter estimates even if spatial homogeneity is assumed.*
8. *Use the $\mathcal{LM}$ power comparison mentioned by (Florax et al., 2003) to select the optimal model from the set of models estimated in step 7. Note, this model assumes spatial homogeneity.*
9. *Given the optimal model found with step 8, estimate the Bayesian counterpart of the optimal model selected in step 8 to control for spatial heterogeneity. If both models exhibits similar results and spatial heterogeneity is rejected, then take the optimal model found in step 8 as optimal. Otherwise, if spatial heterogeneity matters, take the Bayesian model as the optimal one.*

It is worth to mention, that Moran's $\mathcal{I}$ is valid, as long as heteroscedasticity is not spatial correlated. This is a very new insight, but until today no appropriate method is developed to test for spatial correlated heteroscedasticity. There is only one test proposed by (Kelejian and Robinson, 2004), which cover the aspect of spatial correlated heteroscedasticity, but it is only valid for large samples and small samples properties are not known.

4.4.1 Initial model estimation

Let us start with the first step of the laid out strategy. First, we estimate the initial model with ordinary least square procedure.[68] The initial model, based on a per head Cobb Douglas production technique, with $ln\left(\frac{Y}{L}\right)$ as the dependent variable, can be written in log-log form as follows:

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + d\gamma + \kappa \quad (4.40)$$

[68] All estimations have been performed with *Matlab* on the basis of the package provided by LeSage with some adoptions. $\mathcal{LM}$ program for spatial lags as other programs are available on request. If appropriate, results have been checked with `R 2.6.2` and `EViews 5.0`.

or in a more compact manner as

$$y = X\beta^X + d\gamma + \kappa \tag{4.41}$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$ and $X = [1, k, l, h, p, i]$ with $\kappa \sim (0, \sigma^2\Omega)$, $\sigma^2\Omega \neq \sigma^2 I$, $\Omega = diag(v_1, ..., v_N)$ and d as West-East dummy. Two remarks regarding the specification of equation 4.40 or equation 4.41: First, as usual, the coefficient vector β^X contains constant production elasticities of the respective values stacked in X. Because we estimate a production technique per capita, the depended variable is $y = ln\left(\frac{Y}{L}\right)$. Thus the elasticity of production for labour l in this context is defined as $\beta^l + 1$. Therefore, we expect a negative sign of β^l . Second please note, that the inclusion of both *R&D* expenditures and P leads to a serious endogenity problem, because patents are produced with *R&D* expenditures or $P = u(R\&D)$ with $u(\cdot)$ as continuous function. It is worth to mention that patents generally outperforms *R&D* expenditures regarding their interpretation as a quality measure of innovativeness.[69]

In table 4.6 one can find four different specifications. For every specification $\mathcal{LM}$ tests have been conducted, both for spatial lag and spatial error. Additionally, the test statistics for first order and second order spatial influence have been computed.[70] Further, Moran‘s $\mathcal{I}$ test has been performed, also for first and second order spatial influence.

Column (1) of table 4.6 reports a simple estimation of y on k and l and a West-East dummy d. The values of the elasticity of production for capital and labour indicate the expected positive sign and have the expected dimension.[71] and have the correct dimension regarding their influence on per capita production. Furthermore, the dummy is positive as expected and highly significant which indicates that West German regions are more productive on average than East German regions. As we can see from column (1) of 4.6, both Moran‘s $\mathcal{I}$ tests cannot reject the null hypothesis of no spatial correlation. Also the $\mathcal{LM}$ lag for $r = 1$ and $r = 2$ are not significant. This is again the case for the $\mathcal{LM}$ error test for $r = 1$. For $r = 2$ the $\mathcal{LM}$ error test of no spatial correlation under the null hypothesis can be rejected at a 5% significance level.

Although, we find a contradiction regarding the evaluation of Moran‘s $\mathcal{I}$ for $r = 2$ and the $\mathcal{LM}$ error test for $r = 2$ with respect to spatial influence we should expand the estimation and include the knowledge variables human capital h and patents p. Further infrastructure i as additional regressor has been included. The estimation

[69] Refer (Lechevalier et al., 2007) for instance.

[70] For example $\mathcal{LM}^2_\lambda$ stands for a test of no spatial correlation up to order $r = 2$ for spatial error component.

[71] The value for the elasticity of production for labour is 1-0.19=0.81.

results of this expanded specification can be found in column (2) of table 4.6. For all three additional included coefficient regressors we should expect a positive sign. This is true for the estimated coefficients of human capital and infrastructure, but not for patents, which is contra intuitive at first glance. But looking at significance we find, that patents are not significant, not even at a 10% significance level. This is also true for infrastructure which is not significant at a 10% significance level. Additionally, looking again on the coeffiecent for patents the influence of own patents on own labour productivity is at least zero. Refering to the test statistics, it should be noted, that the $\mathcal{LM}$ test for spatial lag is significant at a 5% significance level. Moran's $\mathcal{I}$ for $r = 1$ suggests, that a spatial error model should be estimated which is underpined by the siginifcant $\mathcal{LM}$ test for the spatial error component for $r = 2$.

Given our estimation strategy, we should expand our model by exogenous spatial lagged variables. The advantage of this formulation is straightforward: the estimators of this estimation are unbiased using OLS.[72] Keeping in mind our results obtained from picture 4.11 and 4.12 we include first order spatial lags of human capital $\ln(H^{+1})$, of patents $\ln(P^{+1})$ and of infrastructure $\ln(I^{+1})$ and in addition the second order lag of patents $\ln(P^{+2})$. Stacking this values in $\tilde{X}^1 = [h^{+1}, p^{+1}, i^{+1}]$ and $\tilde{X}^2 = [p^{+2}]$ defining $\tilde{X}^{++} := [\tilde{X}^1, \tilde{X}^2]$ and letting $\beta^{++} = [\beta^{\tilde{X}^1}, \beta^{\tilde{X}^2}]'$ with $\beta^{\tilde{X}^1} = [\beta^{h^{+1}}, \beta^{p^{+1}}, \beta^{i^{+1}}]$ and $\beta^{\tilde{X}^2} = [\beta^{p^{+1}}]$, this leads to the following expansion of equation 4.40:

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \tag{4.42}$$

$$+ \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + \ln(P^{+2})\beta^{P^{+2}} + d\gamma + \kappa,$$

or again in compact notation:

$$y = X\beta^X + \tilde{X}^{++}\beta^{++} + d\gamma + \kappa \tag{4.43}$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$, d as West-East dummy and $X = [1, k, l, h, p, i]$ with $\kappa \sim (0, \sigma^2 I)$.

The estimation results for 4.43 can be found in table 4.6 in column (3). Once again, we would expect positive effects from neighbouring regions. But with the exception of patents, we find negative signs of coefficients for neighbouring human capital and neighbouring infrastructure. Over the more the latter two coefficients are highly non significant. The negative second order spillover coefficient of patents is highly insignif-

[72] Again, refer to appendix 1 for more details.

icant, too. As the coefficient for the own patents, this second order coefficient of neighbouring patents is close to zero. But what can we see is, that the first order neighbouring patent activity has a significant positive effect on own productivity. If we look at our test statistics in column (3) we find that the $\mathcal{LM}$ test for spatial lag is, on contrary to column (2), not significant anymore. This could be due to the inclusion of the spatial lagged patent activity. Furthermore, the second order $\mathcal{LM}$ error test is still significant at a 10% significance level, whereas the first order $\mathcal{LM}$ error test is now significant at a 5% significance level. Also the first order Moran‘s $\mathcal{I}$ test is significant at a 5% significance level. This lead us to conclude that a first order spatial error model should be modeled, because of the fact that $\mathcal{LM}_{\lambda^1} > \mathcal{LM}_{\lambda^2}$. The last column of table 4.6 shows the same regression as in column (3) but with the exclusion of the highly non significant spatial second order patent activity. If we compare column (3) and column (4) we can assert, that the exclusion of spatial second order patent activity does not change the sign and significance of the regression. Therefore, we should proceed with the specification which can be found in column (4) in 4.6.

In summary, we can conclude from 4.6 that spatial processes can be detected in the data. In consequence, we have to acknowledge them in our regression equation and in an adequate estimation procedure. From column (4) in table 4.6 we further know, that spatial dependence in the error term should be acknowledged. What we do not know up to this stage is, if spatial heterogeneity matters. This topic is treated in the next section.

dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$				
independent variables $\mathbf{x} \in \mathbf{X}$	OLS	OLS	OLS	OLS
Column	(1)	(2)	(3)	(4))
Constant	10.50177	10.72613	10.39325	10.38857
	(0.0000)$^{\diamond}$	(0.0000)	(0.0000)	(0.0000)
ln(K)	0.251596	0.237242	0.282934	0.282665
	(0.0124)	(0.0139)	(0.0109)	(0.0088)
ln(L)	-0.193904	-0.219739	-0.243889	-0.243392
	(0.0483)	(0.0264)	(0.0186)	(0.0143)
ln(I)	—	0.014814	0.016332	0.016466
	(—)	(0.5218)	(0.4653)	(0.4673)
ln(H)	—	0.149377	0.161219	0.161679
	(—)	(0.0290)	(0.0076)	(0.0046)
ln(P)	—	-0.002382	-0.023189	-0.023487
	(—)	(0.9373)	(0.5216)	(0.5011)
$\ln(H^{+1})$	—	—	-0.066795	-0.066647
	(—)	(—)	(0.6357)	(0.6286)
$\ln(P^{+1})$	—	—	0.054656	0.054706
	(—)	(—)	(0.0453)	(0.0391)
$\ln(I^{+1})$	—	—	-0.019147	-0.019317
	(—)	(—)	(0.8033)	(0.8022)
$\ln(P^{+2})$	—	—	-0.000197	—
	(—)	(—)	(0.9726)	(—)
d	0.218824	0.273835	0.226299	0.226651
	(0.0000)	(0.0002)	(0.0061)	(0.0056)
Moran-$\mathcal{I}_1$	0.96	2.34	3.19	3.29
	(0.2506)	(0.0253)	(0.0024)	(0.0018)
Moran-$\mathcal{I}_2$	-0.11	0.26	0.21	0.27
	(0.3967)	(0.3860)	(3879)	(0.3850)
$\mathcal{LM}_{\lambda^1}$	0.21	2.42	4.90	4.89
	(0.6483)	(0.1201)	(0.0268)	(0.0270)
$\mathcal{LM}_{\lambda^2}$	5.25	3.74	3.52	3.53
	(0.0219)	(0.0532)	(0.0601)	(0.0601)
$\mathcal{LM}_{\rho^1}$	1.21	5.58	1.82	1.81
	(0.2800)	(0.0184)	(0.1775)	(0.1782)
$\mathcal{LM}_{\rho^2}$	1.82	0.11	0.97	0.01
	(0.1774)	(0.7350)	(0.3236)	(0.9202)
Observations	39	39	39	39
adjusted R^2	0.69	0.74	0.75	0.76

$^{\diamond}$White heteroscedasticity-consistent p-values in ().

Table 4.6: Results of OLS estimation for German NUTS-2 regions

4.4.2 Expansion of the initial model

From the discussion before we know that we have to expand our regression equation in 4.6 by an spatial lagged error term. Therefore, we have to reformulate our regression model 4.42 or 4.43 as a spatial error model (SEM). This is done with equation 4.44:

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \tag{4.44}$$

$$+ \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + d\gamma + \epsilon,$$

with $\epsilon = \lambda_1 W^{+1}\epsilon + \kappa$ or again in compact notation:

$$y = X^{+++}\beta^{+++} + d\gamma + \tilde{\epsilon}\Lambda + \kappa, \tag{4.45}$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$, $X = [1, k, l, h, p, i]$, $\Lambda = [\lambda^1]$, $W^{++} = [W^{+1}]$, $X^{+++} = [X, \tilde{X}]$, $\beta^{+++} = [\beta^X, \beta^{++}]$, with $\kappa \sim (0, \sigma^2 I)$ and d as West-East dummy.

Model 4.45 should be estimated via two different ways:

- *The first approach is to estimate this model with the assumption of $\sigma^2\Omega = \sigma^2 I$, implying spatial homogeneity, which is a common assumption in the relevant studies in this subject*[73]*. As mentioned above, model 4.44 should be estimated via ML.*
- *The second approach is to estimate this model with the assumption of $\sigma^2\Omega \neq \sigma^2 I$, implying spatial heterogeneity with a Bayesian approach which is laid out latter.*

If we go back to the first approach, first we have to set up our Likelihood function. This is:

$$\mathcal{L} = \frac{\left|\tilde{N}\right|}{(2\pi\sigma^2)^{\frac{N}{2}}} exp \left\{ \frac{1}{2\sigma^2}(y - X^{+++}\beta^{+++})'\Theta^{-1}(y - X^{+++}\beta^{+++}) \right\}, \tag{4.46}$$

with $\Theta^{-1} = \tilde{N}'\tilde{N}$ and $|\Theta|^{\frac{1}{2}} = |\tilde{N}|$ and N the numbers of observations.

The corresponding log-likelihood for 4.46 is

[73]Refer for instance to (Olejnik, 2008) or (Santolini, 2008).

$$\ln \mathcal{L} = -\frac{N}{2}\ln 2\pi - \frac{N}{2}(\sigma^2) + \ln|\tilde{N}| - \frac{1}{2}\xi'\xi, \tag{4.47}$$

with $\tilde{N} = (I - \lambda^1 W^{+1})$ and $\xi = \tilde{N}(y - X^{+++}\beta^{+++})$. This expression 4.47 can be written in concentrated form as

$$\ln \mathcal{L}_c \propto \ln|\tilde{N}| - \frac{N}{2}\tilde{\xi}'\tilde{\xi}, \tag{4.48}$$

with $\tilde{\xi} = \frac{1}{\sigma}\tilde{N}(y - X^{+++}\tilde{\beta}^{+++}_{ML})$. The obtained Maximum-Likelihood based estimators can be written as

$$\tilde{\beta}^{+++}_{ML} = (X^{+++\prime}\tilde{N}'\tilde{N}X^{+++})^{-1}X^{+++\prime}\tilde{N}'\tilde{N}y \tag{4.49}$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{N}(\tilde{\xi}'\tilde{\xi}), \tag{4.50}$$

obtained from maximizing 4.47. As we can see, equation 4.48 is highly non linear in the parameter λ^1. Because both β^{+++} and κ are a function of λ we should use an iterative method to estimate λ^1. An approach is to first, estimate β^{+++} via OLS, then find with the associated estimated residuals a value of λ^1 which maximizes the concentrated likelihood function 4.48, third update the OLS values of β^{+++}. With the new updated values of β^{+++} then estimate new λ^1, based on the updated estimated residuals. Convergence is achieved, if values for both residuals and for β^{+++} do not change anymore from one to the next iteration step, which means the difference between $\beta^{+++}_t - \beta^{+++}_{t-1} < \vartheta$ for a small value of ϑ near zero.[74]

It is worth to note, that refering on Maximum-Likelihood, we have to impose a restriction on the parameter λ^1. Referring to (Anselin and Florax, 1995a), p. 34, this parameter takes on feasible parameter values in the range of:

$$\frac{1}{\tilde{\lambda}^1_{min}} < \lambda^1 < \frac{1}{\tilde{\lambda}^1_{max}}. \tag{4.51}$$

$\tilde{\lambda}^1_{min}$ is the minimum Eigenvalue of the matrix W^{+r}, whereas $\tilde{\lambda}^1_{max}$ represents the maximum Eigenvalue of W^{+r}. This suggest a constrained Maximum-Likelihood maximization. If W^{+r} is row standardized, as it should be, then of course $\lambda^1_{max} = 1$. Please

[74]In this application ϑ is set to $\vartheta =$ `1e-8`. Further t is set to a maximum value of `500`.

note, that this procedure could become extremely laborouis with respect to computational issues. More precisely, the computational costs increase with the dimension of the weighting scheme matrix W^{+r}. Alternatively, one can set ex ante values for λ, such as $\lambda^1 \in (0,1)$ which implies only positive spatial error dependence. In this work ex ante values for λ^1 ranging from $\lambda^1 \in (-1,1)$ have been imposed, although a direct computation via Eigenvalues would be possible.

The second approach dealing with the estimation of model 4.45 is to refer on a Bayesian approach but with the additional assumption of spatial heterogeneity, which means that $\sigma^2\Omega \neq \sigma^2 I$. If the model yields the same results and spatial heterogeneity is insignificant, we can conclude, that spatial heterogeneity can be ignored, otherwise, there is at least little evidence that spatial heterogeneity a justified assumption and we have to control for it.

Based on the likelihood function expressed by equation 4.46 a spatial Bayesian heteroscedastic model is set up. The core of Bayesian econometrics is the Theorem of (Bayes, 1763) which is needed in this context for parameter estimation. Assume for a moment that θ is a vector of unknown parameters which should be estimated. Before any data are observed, we have beliefs and some uncertainty with respect to our vector of parameter θ. These beliefs are called "a priori" probabilities which are fully represented by the probability function $p(\theta)$. The entire probability model itself is totally defined by the likelihood $p(y\,|\theta)$. $p(y\,|\theta)$ can be described as the core of Bayesian econometrics, because it contains the entire set of information from the data. Given, we have observed y, then we should update our beliefs regarding θ. By using the theorem of Bayes we obtain the so called "a posteriori" distribution of θ, given y, which is

$$p(\theta|y) = \frac{p(y\,|\theta)p(\theta)}{p(y)}, \tag{4.52}$$

with $p(y) = (y|\theta)p(\theta)$, defined by the law of total probability. Because $p(y)$ do not contain any information regarding θ and, over the more, we only interesting in θ, we can ignore $p(y)$. Thus the "a posteriori" probability is proportional to the likelihood times the "a priori" probability:

$$p(\theta|y) \propto p(y\,|\theta)p(\theta). \tag{4.53}$$

Although the dimensionality of $p(\theta|y)$ depends on the number of unknown parameters, we can often focus on individual parameters such as $\theta_1 \in \theta$ by numerically or analytically integrating out other components[75]. For instance we can write:

$$p(\theta_1|y) = \int p(\theta|y)d\theta_2 d\theta_3 ... \tag{4.54}$$

The entire information needed for inference about θ_1 is contained in the marginal distribution of θ_1. What we have to do now is to specify our exogenous given priors and the likelihood function.

In this context, we assume normal priors for β^{+++} and a diffuse prior for σ. The relative variance terms $v_i \in \Omega$ are fixed but unknown and therefore we have to estimate them. We have to treat the v_i as informative priors. The distribution of all elements of Ω are assumed to be independently $\frac{\chi^2}{s}$ distributed, with $s \sim \Gamma(a, b)$. As mentioned we are confronted with a degree of freedom problem, if the number of estimated coefficients exceeds the number of observations. Considering the fact, that the χ^2 distribution is a single parameter distribution we are able to compute N additional parameters v_i by adding only one single parameter s to our model. This idea goes back to (Geweke, 1993) who uses this type of prior to model heteroscedasticity and outliers in a linear regression framework. The idea becomes more clear if one knows that the mean of this priors is unity, whereas the variance of this prior is $\frac{s}{2}$. Thus, if s takes a large value, then all terms of Ω tend to unity, yielding a homoscedastic scenario, because σ is weighted equally for every observation, hence we obtain a constant variance over space. An assumption, which is made within the traditional spatial Maximum-Likelihood approach. On contrary, small values of s lead to a skewed distribution. The role of v_i therefore is, as in a traditional GLS approach, to down weight observations with large variances. For this reason, the degrees of freedom s plays a crucial role when robustifying against outliers. For $s \to \infty$ the limiting normal and therefore a homoscedastic "scenario" is realized. One option could be to assign a improper value to s. The other possibility is to use a proper prior for s which is Gamma distributed:

$$s \sim \Gamma(a, b), \tag{4.55}$$

with hyperparameter a and b. It has to point out, that the virtue of the first option is that less draws compared to the second option are required for parameter estimations and moreover convergence is quicker.

If $\Gamma(a = \frac{s}{2}, b = 2)$ this is equivalent to $\chi^2(s)$, hence we obtain a so called mixing

[75] Refer to (Geweke, 1993).

distribution controlled by s. As shown by (Geweke, 1993) we can write

$$\pi\left(\frac{s}{v_i}\right) \sim \text{iid}\, \chi^2(s),\ \forall\, i, \tag{4.56}$$

with $\pi(\cdot)$ denoting the prior from now. This implies, that the normal mixture model with 4.56 is equivalent to a model based on independently distributed Student-t values with s degrees of freedom, known as the (Theil and Goldberger, 1961) Model. The spatial error parameter is assumed to follow an uniform, but proper distribution with the range $\hat{N}$ as $\pi(\lambda^1) = \frac{1}{\hat{N}} = \frac{1}{\tilde{\lambda}^1_{min} < \lambda^1 < \tilde{\lambda}^1_{max}} \sim \mathcal{U}[-1, 1]$.

Let us summarize our assumptions regarding the priors as follows:

$$\pi(\beta^{+++}) \sim \mathcal{N}(c, T), \tag{4.57}$$

$$\pi(\frac{s}{v_i}) \sim \ \text{iid}\ \frac{\chi^2(s)}{s}, \tag{4.58}$$

$$\pi(\lambda^1) \sim \mathcal{U}[-1, 1]. \tag{4.59}$$

Given the priors defined above, we need the conditional posterior distributions for each parameter $\beta^{+++}, \sigma, \lambda^1, \Omega$ to estimate them. Using the priors, assuming that they are independent from each other, we can define the joint posterior as:

$$p(\beta, \sigma, \lambda^1) = p(\beta)p(\sigma)p(\lambda^1)$$

$$\propto \left|I - \lambda^1 W^{+1}\right| \sigma^{-N} exp\left\{-\frac{1}{2\sigma^2}(\xi'\Omega^{-1}\xi)\right\} \sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}(\beta - c)' T^{-1}(\beta - c)\right\}. \tag{4.60}$$

From 4.60 the conditional distribution of β^{+++} is obtained from the standard non spatial Bayesian GLS approach as:

$$p(\beta^{+++}|\lambda^1, \sigma, \Omega, y) \sim \mathcal{N}[H(X^{+++}\tilde{N}\Omega^{-1}\tilde{N}y + \sigma^2 T^{-1}c, \sigma^2 H)], \tag{4.61}$$

with $H = (X^{+++\prime}\tilde{N}\Omega^{-1}\tilde{N}X^{+++} + T^{-1})^{-1}$, $\tilde{N} = (I^1_\lambda W^{+1})$, mean c and the corresponding variance covariance matrix T.

The conditional distribution of σ is

$$p(\sigma|\lambda^1, \Omega, \beta^{+++}, y) \propto \sigma^{-(N+1)} exp\left\{\frac{1}{2\sigma^2}\xi'\Omega^{-1}\xi\right\}. \tag{4.62}$$

Next the conditional distribution of every element v_i of Ω is considered. (Geweke, 1993) shows, that the conditional distribution for $v_i \in \Omega$ represents a χ^2 distribution with $s+1$ degrees of freedom:

$$p\left(\left[\frac{(\sigma^{-2}e_i^2 + s)}{v_i}\right] |\beta^{+++}, \lambda^1, v_{-i}, \lambda^1\right) \sim \chi^2(s+1), \tag{4.63}$$

with $v_{-i} = \{v_1, ...v_{i-1}, v_{i+1}, ..., v_N\}$.

Now consider the conditional distribution for the parameter σ assuming that we already know the parameters, given we know β^{+++}, λ^1 and Ω. This distribution would be:

$$p\left[\sum_{i=1}^{N}\frac{e_i^2}{v_i}/\sigma^2|\beta^{+++}, \lambda^1, \Omega\right] \sim \chi^2(N). \tag{4.64}$$

With 4.64 we adjust estimated residuals e_i with estimated weights or relative variance terms v_i. This approach corresponds to the simple weighted least square procedure (WLS) known from basic econometricians toolbox.

Finally, the conditional posterior of λ^1 is calculated as follows:

$$p(\lambda^1|\sigma, \Omega, \beta^{+++}, y) \propto |\tilde{N}| exp\left\{\frac{1}{2\sigma^2}\xi'\Omega^{-1}\xi\right\}. \tag{4.65}$$

With exception of 4.65, all other posterior distributions are standard and therefore a Markov Chain Monte Carlo method (MCMC) can be applied to estimate parameters $\beta^{+++}, \lambda^1, \sigma^2, \Omega$. Usually, a Gibbs sampling approach, which is based on the conditional posterior densities is used.

We wish to make several draws to generate a large sample from which we can approximate the posterior distributions of our parameters. Unfortunately, we cannot approximate a posterior distribution for expression 4.65, because this type of distribution do not correspond to any so called standard class of probability densities. For this reason, Gibbs sampling cannot be readily used. Fortunately, a method called "Metropolis-

Hasting" sampling which is an additional sequence in Gibbs sampling procedure[76], allows us to approximate the posterior distribution for λ^1.[77] The only problem one has to solve is to find a suitable proposal density. (LeSage, 2000) suggests to assume a normal or Student t-distribution. Because of the fact, that λ^1 has to be handled as a restricted parameter, which is situated between minus one and one, the sampler rejects values outside the interval $(-1, 1)$ from the sample. This is called "rejection sampling".[78]

The "Metropolis-Within-Gibbs" sampling algorithm can be expressed as follows:

1. *Set t=0.*
2. *Define a starting vector $S_{t=0}$ which contains the initial parameter of interest: $S_0 = [\beta_0^{+++}, \sigma_0^2, v_{i0}, \lambda_i^1]$.*
3. *Compute the mean and variance of β^{+++} using 4.61 conditional on all other initial values stacked in S_0.*
4. *Use the computed mean and variance of β^{+++} do draw from a multivariate normal distribution a normal random vector β_1^{+++}.*
5. *Calculate 4.64 refering on β_1^{+++} from step 4 and use this expression in combination with $\chi^2(N)$ random draw to determine σ_1^2 for $i = \{1, 2, ..., N\}$.*
6. *Use β_1^{+++} and σ_1^2 to calculate 4.63 and use this value together with a N-dimensional vector of $\chi^2(s+1)$ random draws to determine $v_i \in \Omega$ for $i = \{1, 2, ..., N\}$.*
7. *Use metropolis within Gibbs sampling to calculate λ^1 using values $v_i \in \Omega$ for $i = \{1, 2, ..., N\}$, β_1^{+++} and σ_1^2.*
8. *Set t=t+1.*

The question which remains is, how to select the correct Bayesian model. It is sometimes the case that several competing models M_u with $u = \{1, 2, ..., U\}$ exist. Then usually posterior probabilities are computed which should give advice, which model is the correct model in terms of probability. The posterior probability p_u^{pos} for model u is given by[79]:

[76] Because of this reason, the method is also called "Metropolis-Within-Gibbs".
[77] Refer to (Gelman et al., 1995).
[78] Refer to (Gelfand et al., 1990).
[79] Please refer to (Hepple, 2004), p. 105.

$$p_u^{pos} \equiv p(M_u|y) = \frac{p(y|M_u)}{\sum_{u=1}^{U} p(y|M_u)}. \tag{4.66}$$

Bayesian model averaging suggests to weight all possible Bayesian models M_u with $u = \{1, 2, ..., U\}$ with their corresponding posterior probabilities. In terms of probability this means:

$$p(y^*|y) = \sum_{u=1}^{U} p(y^*|y, M_u)p(M_u|y), \tag{4.67}$$

with $p(y^*|y)$ as the posterior, $p(M_u|y)$ as the posterior model probability and $p(y^*|y, M_u)$ as the likelihood function of model M_u. The reason why model averaging should be used is quite simple. The traditional approach is to choose the single best model based on calculating posterior model probabilities with 4.66 for every model of interest.[80] But one has to remember that this rather excluding approach could be lead to wrong decisions, because a researcher has to decide on the basis of model probabilities what is the "good model" and what is the "not so good model" from a sometimes large set of models. Additionally, only referring to the "good model"' ignores model uncertainty. In this study, relying on model probabilities is not a good idea, because "posterior model probabilities cannot be meaningful calculated with improper non informative priors,"[81] which are not common for all models. Therefore we refer to the MCMC literature to compute a posteriori model probabilities. This so called MC^3 approach, introduced by (Madigan and York, 1995) is based on a stochastic Markov Chain process which moves through the model space and samples those regions which have a high superior model support. Thus this approach is very efficient because not the entire model space is of interest.[82]

Knowing these facts, we are now able to interpret our estimation results for both approaches, the Maximum-Likelihood and the Bayesian approach. The results for the first approach can be found in column (1) and (2) of table 4.8. The first regression is a mixture model of spatial lag and spatial error model, the so called spatial ARMA model, which is in this case labeled as SEC(r,r) to avoid confusion with respect to time

[80]A large bulk of literature on Bayesian model averaging (BMA) over alternative linear regression models containing differing explanatory variables exists. For instance refer to (Raferty et al., 1997), (Fernandez et al., 2001b) and (Fernandez et al., 2001a). The MC^3 approach, is set forth for in (Madigan and York, 1995) for the SAR and SEM models.

[81](Koop, 2003), p. 268.

[82]Refer to (LeSage and Parent, 2007) for an excellent contribution to this topic.

Bayesin model (4)[SEM(1)]	Model (4,1)	Model(4,2)	Model (4,3)	Model(4,4)
Runs	10,000	10,000	100,000	100,000
Informative Priors	No	Yes	No	Yes
p_u^{pos}	0.2770	0.2509	0.2374	0.2374

Table 4.7: MC^3 a posteriori model probabilities p_u^{pos} for variants of model (4)[SEM(1)]

series context.[83] This regression is done to corroborate our model selection on inductive statistics, done in the forgoing chapter. After estimation of all possible combinations of first order and second order spatial models[84], we have chosen the SEC(1,1) model as the appropriate model on basis of the value of the log-likelihood. Leaving out the insignificant parameter ρ^1, estimating a pure spatial error model (column (2)) and comparing this with column (1) we can see, that only minor changes of coefficient values result. This is an indicator, that the spatial lag does not provide any further information for our model. Thus, it is justified, to model a spatial SEM(1) model, which is printed in column (2), because the spatial error coefficient λ^1 is highly significant. Comparing the SEM(1) model with the fourth column of 4.6 we can find moreover, that the coefficient for $\ln(I^{+1})$ is not positive, but again highly non significant. All other coefficient have, compared to (4) in table 4.6, roughly the same dimension, the same sign and the same level of significance.

Additionally, the results for the second approach, an estimation of the Bayesian counterpart of equation 4.45 can be found in column (3) of table 4.8. Before discussing the results, we first should get an intuition of what is behind the Bayesian estimation approach.

To obtain estimates from our Bayesian approach we have to simulate draws. To ensure stability of simulated results, one should do a simulation on non informative priors and on informative priors, for which starting values are obtained from a corresponding Maximum-Likelihood estimation with different draws. For this reason, two Bayesian estimations, one with 10,000 draws and one with 100,000 draws, each with informative and non informative priors have been conducted. At all we get 4 models, for each number of draws one should estimate a model with informative and non informative priors.[85] The model probabilities p_u^{pos} for the relevant models can be found in table 4.7.

Calculating this probabilities and comparing them with each other, we find, that the

[83]See appendix 3 for a deviation of the log-likelihood of the spatial ARMA model.

[84]See appendix 4 for a summary.

[85]Because of the fact, that the initial model estimation results on which Bayesian model specification is based are drawn in column (4) of 4.8, we label variants of the Bayesian model as model (4,1), (4,2), (4,3) and (4,4).

first model $(4, 1)$ has slightly a higher probability to be the correct model.

Furthermore, MCMC-convergence checks the four relevant models have been performed.[86] to ensure convergence of the sampler. If the means and variances for the posterior estimates are similar from all runs, convergence seems ensured at all. The convergence tests for all regressions show, that convergence of the sampler is guaranteed for all simulations. Therefore, we rely on model $(4, 1)$ because it request fewer draws. The estimation results for this model can be found in 4.8 in column (3).

If we now turn back to 4.8 and compare the heteroscedastic Bayesian counterpart in column (3) with the homoscedastic Maximum-Likelihood based estimation in column (2) then we can easily see, that estimation results do not differ dramatically. Picture 4.13 and picture 4.14 confirm this result. Again, the coefficient of $\ln(I^{+1})$ is positive but not significant. On contrary, the heteroscedastic Bayesian approach estimates a lower value for the spatial lag component λ^1, as the homoscedastic Maximum-Likelihood does. But again, the parameter range for λ^1 is comparable between the two approaches and both coefficient values are highly significant on a 1% significance level.

The last point we have to tackle is to ask, whether the spatial Bayesian estimation provides us with some evidence of spatial heterogeneity. Picture 4.15 shows a plot of the mean of the v_i draws which should serve as an estimate of these relative variance terms. We can see that one outlier is identified, irrespectively what model we choose. If spatial homogeneity is observed, all elements of Ω should realize the value one. Obviously, this is not the case for all four Bayesian models, as we can see from figure 4.15. From this point of view, we should conclude, that spatial heterogeneity matters, although Maximum-Likelihood and Bayesian estimates correspond each other with respect to parameter estimates. Therefore, we should choose the Bayesian model represented column (3) in table 4.8 as the optimal one, which delivers efficient parameter estimates.

[86]Please refer to appendix 5 for a short description of convergence criteria and appendix 6 for convergence diagnostic of all selected models.

dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$			
independent variables $\mathbf{x} \in \mathbf{X}$	ML	ML	Bayes
Preferred Model	(4)[SAC(1,1)]	(4)[SEM(1)]	(4)[SEM(1)]
Column	(1)	(2)	(3)
Constant	8.531315	10.09034	10.07207
	(0.0034)$^\diamond$	(0.0000)	(0.0000)
ln(K)	0.306738	0.303988	0.294401
	(0.0000)	(0.0000)	(0.0019)
ln(L)	-0.232205	-0.235542	-0.231121
	(0.0004)	(0.0003)	(0.012526)
ln(I)	0.009437	0.011475	0.006971
	(0.6514)	(0.5828)	(0.3904)
ln(H)	0.196265	0.183675	0.187209
	(0.0006)	(0.0003)	(0.0024)
ln(P)	-0.050593	-0.043691	-0.043108
	(0.0847)	(0.1101)	(0.1145)
$\ln(H^{+1})$	-0.055016	-0.071008	-0.004407
	(0.5594)	(0.4149)	(0.4871)
$\ln(P^{+1})$	0.070208	0.083117	0.062044
	(0.0477)	(0.0008)	(0.0164)
$\ln(I^{+1})$	0.028737	0.035589	0.030584
	(0.6262)	(0.5416)	(0.3233)
d	0.261555	0.252277	0.255723
	(0.0000)	(0.0000)	(0.0005)
ρ^1	0.132883	—	—
	(0.5905)	(—)	(—)
λ^1	0.696998	0.710951	0.561134
	(0.0000)	(0.0000)	(0.0081)
Observations	39	39	39
$\ln(\mathcal{L})$	90.47	67.93	—
adjusted pseudo R^2	0.83	0.83	0.81

Table 4.8: Estimation results for German NUTS-2 regions

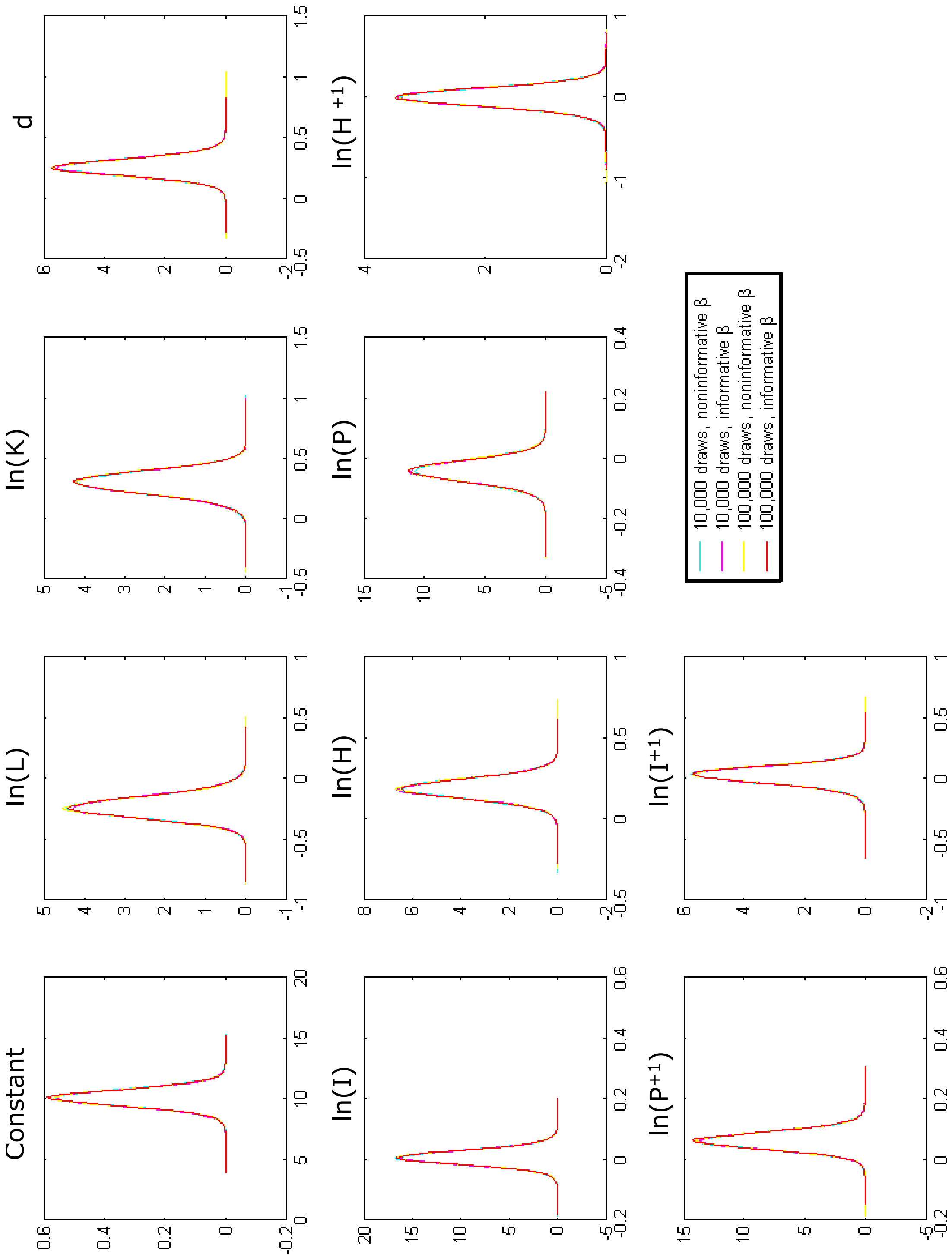

Figure 4.13: Density plots of estimated β^X and β^{++}

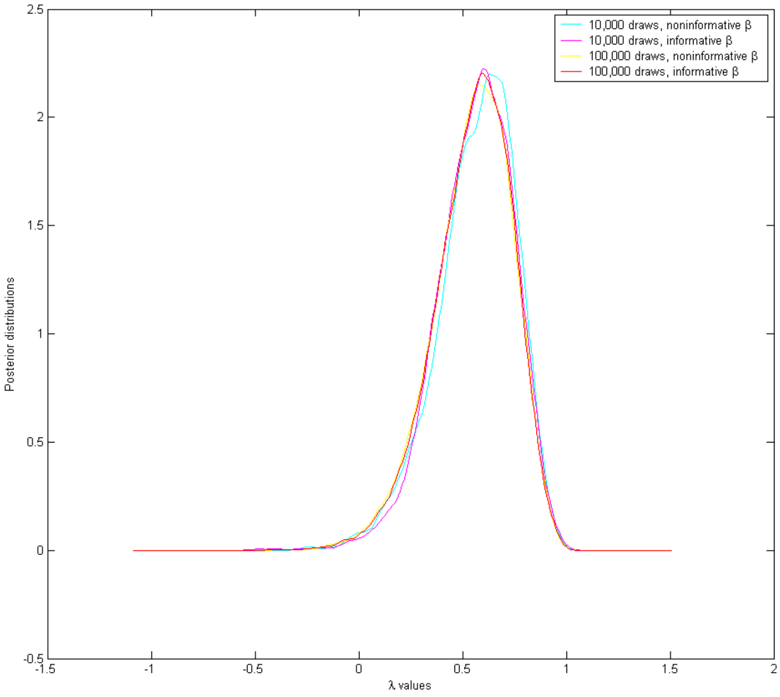

Figure 4.14: Density plots of estimated λ^1

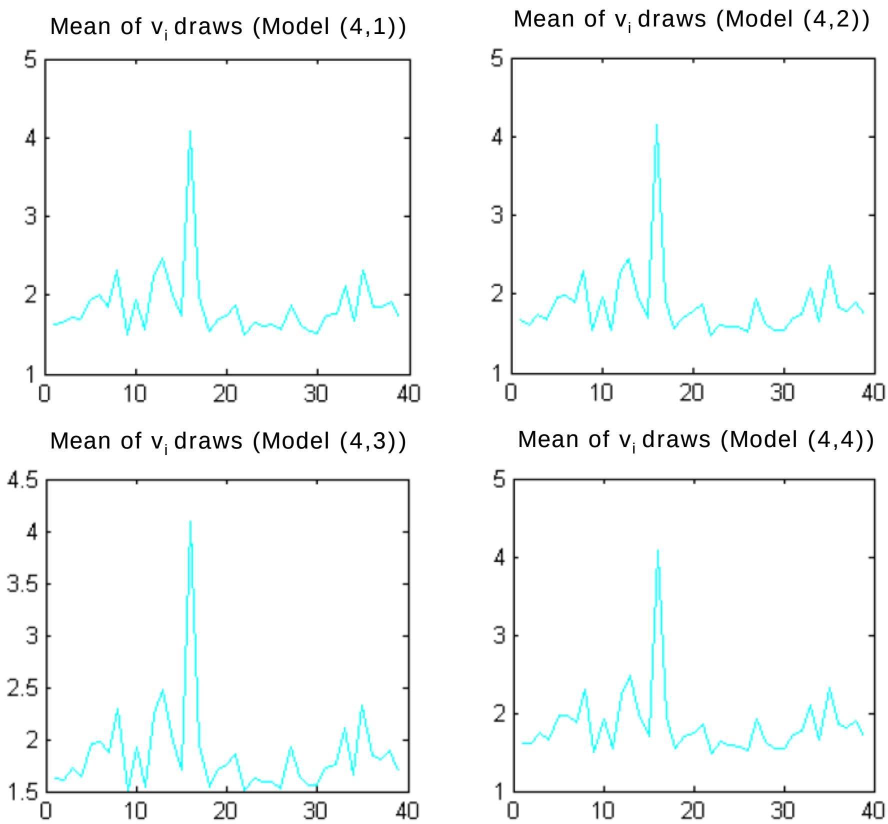

Figure 4.15: Computation of v_i draws of Ω

4.4.3 Interpretation of obtained results

In this section we have tried to find out how regional labour productivity is affected by spatial knowledge processes. We found, that first order neighbouring patent activity influences the regions own labour productivity, while own patent activity does not exhibit a significant influences on own labour productivity. Additionally, most of spatial activity cannot be explained fully by exogenous spatial lagged knowledge. This is the case, because the spatial error term is highly significant, even if one includes spatial lagged counterparts of exogenous variables. Additionally, it was shown with a spatial Bayesian analysis, that spatial heterogeneity is a reasonable assumption and neglecting this issue would lead to inefficient parameter estimates.

The next step is to investigate further the impact of knowledge diffusion on German NUTS-2 regions more systematically. So far, we only have obtained some evidence, that the data generating process can be described also by spatial effects. The next step is, to isolate the spatial neigbbouring influence from the data. In this way it is possible, to distinguish between region specific or home effects and neighbour effects. For instance, regions might have a high labour productivity compared to the average, but this level of labour productivity might be influenced negatively by neighbouring regions et vice versa. The goal is to identify strength and weakness of German NUTS-2 regions and derive implications for regional policy instruments.

4.5 Spatial filtering

In this section we try to isolate spatial spillover effects from region specific labour productivity. In this way it is possible to create a strength and weakness profile of German Nuts-2 regions. Particularly, one should be interested in answering the question which regions have positive effects on neighbouring regions and which regions provide negative effects on neighbouring regions. This has also implications for an appropriate regional policy. In this way we can say that labour productivity is a sum of own labour productivity and spillovers from neighbouring regions which can be either positive or negative. The question is, if the overall effect is positive or negative. We base the spatial filtering procedure on the so far obtained results. Thus we set $r = 1$ and include only patents p and human capital h as exogenous variables in our filter procedure.

4.5.1 Concept of the filtering approach

Spatial filtering is a well established analysis method in spatial econometrics applications. The idea is based on a two step estimation technique. In the first step we have to filter every exogenous variable and in the second step we have to regress the dependend endogenous variables on all spatial filtered exogenous variables.

The starting point of spatial filtering is the Morans‘s $\mathcal{I}$. From equation 4.34 we know that Moran‘s $\mathcal{I}$ for a standardized matrix W^{+1} can be computed as follows:

$$\mathcal{I} = \frac{e'W^{+1}e}{e'e}. \tag{4.68}$$

This equation can be reformulated [87] as

$$\mathcal{I} = \frac{y'C^{+1}y}{e'e}, \tag{4.69}$$

with

$$C^{+1} = \left(I - \frac{\iota\iota'}{N}\right) W^{+1} \left(I - \frac{\iota\iota'}{N}\right), \tag{4.70}$$

[87] Refer to (Griffith, 2000), p. 145.

with ι as a $(N \times 1)$ vector of ones and I as the $(N \times N)$ identity matrix. In this way, the Eigenvectors of C^r load every spatial effect. The first Eigenvalue contains the largest Morans's $\mathcal{I}$ coefficient with a given standardized matrix W^{+r}. The second Eigenvalue contain the value, which leads to the maximal Morans's $\mathcal{I}$ given the second Eigenvalue is not correlated with the first one, which is ensured, because W^{+r} is standardized.

Because of missing degrees of freedom, one cannot use every Eigenvector for spatial filtering [88]. Therefore a rule of thumb for Eigenvector selection is needed. (Griffith, 2003) has proposed to use only those Eigenvectors which fulfill the following condition:

$$\mathcal{I} > 0.25\,\mathcal{I}_{max}. \tag{4.71}$$

Equation 4.71 provides us with an indicator regarding the maximum number of Eigenvectors L which should be included into our regression framework. Based on a top down procedure, one can eliminate all Eigenvectors which do not provide a substantial potential of explanation. Given we have identified the relevant Eigenvectors, we can proceed with the filtering scheme. On the first step we filter the vector of independent variables X by running the following regression:

$$x_k = \gamma_0 + \sum_{l=1}^{L} \gamma_l \hat{v}_l + \epsilon_k, \tag{4.72}$$

with $\epsilon \sim (0, \sigma^2\Omega)$, v_l the l^{th} Eigenvector and x_k the k^{th} exogenous variable.

It is clear that the estimated residual vector $\hat{\epsilon}_k$ contains the spatial filtered counterpart of the not filtered variable x_k. The second step is to regress y on spatial filtered variables and on Eigenvectors v_l. In this equation every variable is spatial filtered and therefore OLS estimation is unbiased. The corresponding regression on the second step can be written as:

$$y = \gamma_0 + \sum_{l=1}^{L} \gamma_l \hat{v}_l + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k + \kappa, \tag{4.73}$$

with $\kappa \sim (0, \sigma^2\Omega)$, v_l the l^{th} Eigenvector and x_k the k^{th} exogenous variable. Of course equation 4.73 can be consistently estimated with OLS.

[88]Refer to (Griffith, 2003), p. 107.

4.5.2 Eigenvector computation

From matrix C of equation 4.70 one can derive the Eigenvectors and compute Moran's-$\mathcal{I}$ with 4.34. This can be done using Matlab for instance. For every Eigenvector v_l the corresponding Moran $\mathcal{I}$ coefficient was computed. As one can see from picture 4.16 only 10 of 39 Eigenvalues meet 4.71. The second Eigenvector leads to $\mathcal{I}_{max}$ which takes the value $\mathcal{I}_{max} = 0.97437$.

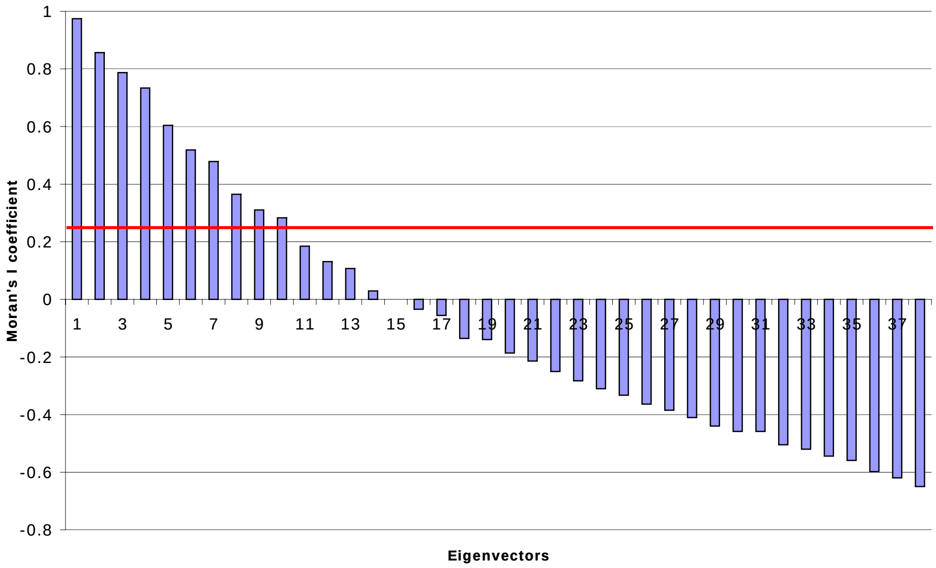

Figure 4.16: Eigenvector selection

4.5.3 Spatial filtering estimation

First we estimated separately 4.34 for k, l, h and p.[89] The results of these regressions, corresponding to equation 4.72 can be found in table 4.9.

As we can see from figure 4.17, we cannot observe a clear spatial pattern, represented by the Eigenvectors v_6 and v_7, labeled as (a) and (b). On contrary, in figure 4.18 the Eigenvector v_1, labeled as (c), the Eigenvector v_2, labeled as (d) and the Eigenvector v_3, labeled as (e) show a clear spatial pattern. The first Eigenvector v_1, labeled as (c) is declining from North to South, the second Eigenvector, the Eigenvector v_2, labeled as (d) exhibit a significant declining West-East pattern, whereas the third Eigenvector v_3, labeled as (e) is affected by low values in North-West and South-East of Germany.

[89]Based on Wald-tests, we should not include i and p both. As argued above again, a Wald-test based on ML estimation ignoring spatial dependence in the data is not valid. Because of the fact that we want to include possible knowledge spillover variables, we only eliminated i. From table 4.8 we see, that neither i nor the spatial lagged counterpart of i are significant, whereas the spatial lagged counterpart of p is significant.

dependend variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$				
estimation method	OLS	OLS	OLS	OLS
dependent variables $\mathbf{x} \in \mathbf{X}$	l	k	h	p
Constant	13.21070	12.30990	2.090592	5.293690
	(0.0000)$^{\diamond}$	(0.0000)	(0.0000)	(0.0000)
v_1	—	—	—	-3.049842
	(—)	(—)	(—)	(0.0019)
v_3	—	—	0.623305	—
	(—)	(—)	(0.0755)	(—)
v_6	-1.259430	-1.229270	—	-1.816913
	(0.0084)	(0.0424)	(—)	(0.0287)
v_7	0.950438	—	—	1.500619
	(0.0176)	(—)	(—)	(0.0961)
Observations	39	39	39	39
adjustedR^2	0.16	0.08	0.07	0.26

Table 4.9: Spatial filtering of exogenous variables X

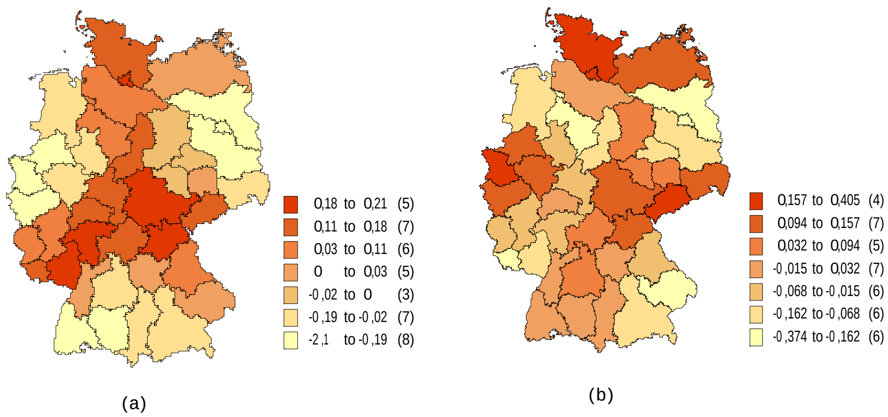

Figure 4.17: Graphical representation of Eigenvectors (I)

After filtering the exogenous variables, the next step is estimating 4.73. Therefore, a stepwise estimating procedure of labour productivity on Eigenvectors and spatial filtered variables is employed. In the regression context, no dummy variable for West-East differences is included, because the dummy would filter spatial information potential and could lead to a biased regression in this context. The results of this estimation can be found in table 4.10. We find, that the first three Eigenvectors, which cover spatial effects, determine a considerable amount of labour productivity. This leads us to conclude, that labour productivity of a given region is not only determined by its own economic potential, but also by neighbouring labour productivity. This implies, that network effects play an important role and should be considered within the embodiment of regional policy. We can therefore conclude, that patents p and human capital h are mainly affected by spatial effects. The latter is only partial true for capital k and

dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$	
estimation method	OLS
Constant	11.21121
	(0.0000)$^{\diamond}$
$\hat{\epsilon}_l$	-0.358058
	(0.0031)
$\hat{\epsilon}_k$	0.261464
	(0.0183)
$\hat{\epsilon}_h$	0.136156
	(0.0747)
$\hat{\epsilon}_p$	0.075354
	(0.1114)
v_1	-0.421669
	(0.0002)
v_2	0.211905
	(0.0323)
v_3	-0.305651
	(0.0225)
Observations	39
adjustedR^2	0.66

Table 4.10: Spatial filtering of labour productivity y

labour l.

As we can see from 4.10, the results for the constant estimated labour elasticity l, the constant estimated capital elasticity k, the constant estimated human capital elasticity h and the constant estimated patent elasticity p have all positive signs and have been, with respect to their dimension correct estimated. With the exception of p, all estimated coefficients are significant on a 5% or 10% level.

Now we are prepared to decompose labour productivity in home effects and neighbour effects. The residual of this simple decomposition cannot be returned neither to home effects, nor to neighbour effects and therefore they are treated as not systematic. Noting the fact, that both, Eigenvectors and spatial filtered variables exhibit a mean of zero, we can conclude that the constant term contains the mean of labour productivity. If we subtract the mean $\bar{y}$ from equation 4.73 we obtain:

$$\breve{y} \equiv y - \bar{y} = \gamma_0 + \sum_{l=1}^{L} \gamma_l v_l + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k - \bar{y} + \kappa, \tag{4.74}$$

with $\kappa \sim (0, \sigma^2\Omega)$, v_l the l-th Eigenvector and x_k the $k-th$ exogenous variable. The term $\gamma_0 + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k - \bar{y}$ can be defined as the own region effect, whereas the term $\sum_{l=1}^{L} \gamma_l v_l$ represents the neighbour effects. The term κ represents the unsystematic component. Because of the fact, that all effects are centered around zero, we can

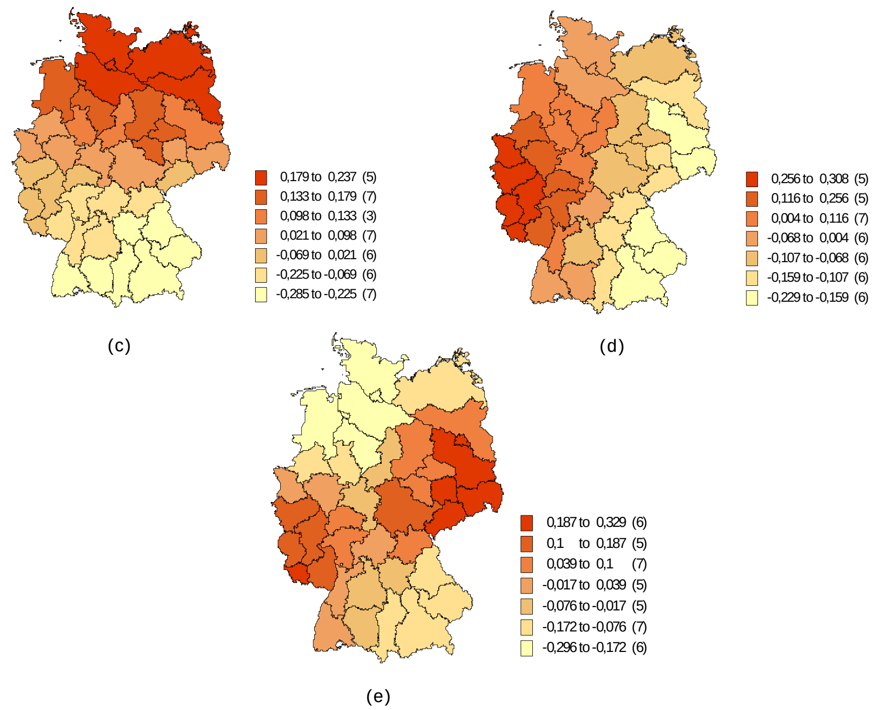

Figure 4.18: Graphical representation of Eigenvectors (II)

interpret equation 4.74 as a deviation from the mean specification. For instance, if $\breve{y} > 0$, which means that a region exhibits a superior labour productivity, this can be due to home effects or due to neighbouring effects. Even if a region has a superior home effect, a negative neighbour effect could lead to a negative overall effect regarding labour productivity, et vice versa.[90]

4.5.4 Interpretation of simulation results

The next two figures in 4.19give an impression of the results of labour productivity simulation, based on equation 4.74. First, we should investigate own regions effects regarding labour productivity, which are separated from regions neighbour effects. The labour productivity effects are deviations from the mean which is, as mentioned above, centered around zero. As we can see with respect to the own region effect (RE) in the left hand map, especially some East German regions, such as "Süd Brandenburg", "Sachsen" and "Berlin" would exhibit a relative high labour productivity if we only apply for own region effects. But with respect to the overall effect (OE), which is

[90]From equation 4.74 we see, that an inclusion of a dummy variable as done before, would bias within the regression context. Even more, a spatial filtering of a dummy is by definition not plausible. Besides that and give, we include the dummy we cannot rule out that these variable also contains spatial information.

plotted in the right hand map, some negative influence from neighbouring regions leads to a reduction of labour productivity in those regions. On the other side, the region "Oberfranken" benefits for the most part from neighbouring effects.

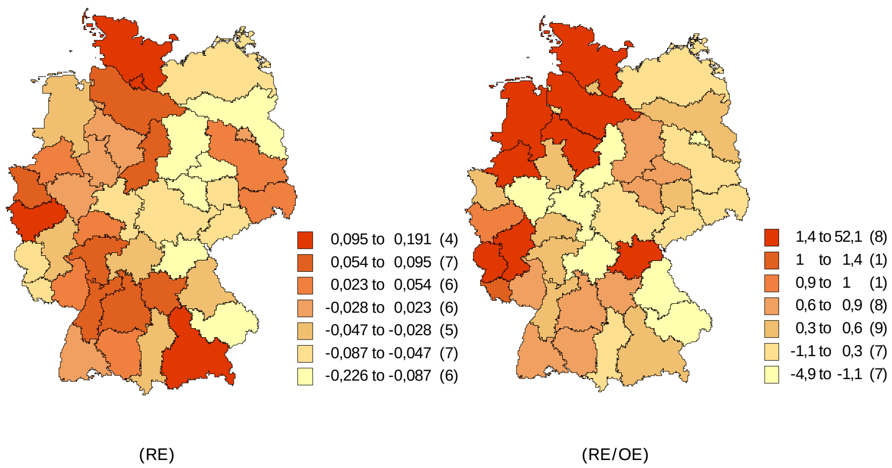

Figure 4.19: Absolute and relative regional effects

Now we turn our attention to the neighbouring effects. These are visualized in figure 4.20. First, we find in the map on the left hand side a rather impressive confirmation that especially South German regions and with some cut backs also West German regions, settled in the "Rhein-Main -Gebiet" and the "Ruhrgebiet", are the source of knowledge spillovers. On contrary, we find maximum negative neighbour influence throughout East German regions. With respect to the overall effect (OE), we find in the map on the right hand side some dramatic changes. The effects for South German regions, some regions of the "Rhein-Main-Area" and some regions of the "Ruhrgebiet" are rather low. Thus, only a little fraction of the superior labour productivity of these regions are due to neighbouring effects.

Finally, we can categorize the regions in a strength-weakness profile, both for the own region and neighbouring region effects.

Regions which are settled in the top right corner of figure 4.21 can be characterized by a superior labour productivity. For those regions, positive or negative neighbouring effects play only a minor role. In these regions, with the exception of "Hamburg", which is top leader with respect to own and overall effects and "Schleswig-Holstein", you can find mainly South German regions, such as "Oberbayern", "Stuttgart", "Tübingen" etc. and West German regions, such as "Düsseldorf", "Köln" etc. situated in this area. These findings supports the findings of (Eckey et al., 2007) for German labour market regions.

Regions which can be found in the down right corner of figure 4.21 exhibit a positive

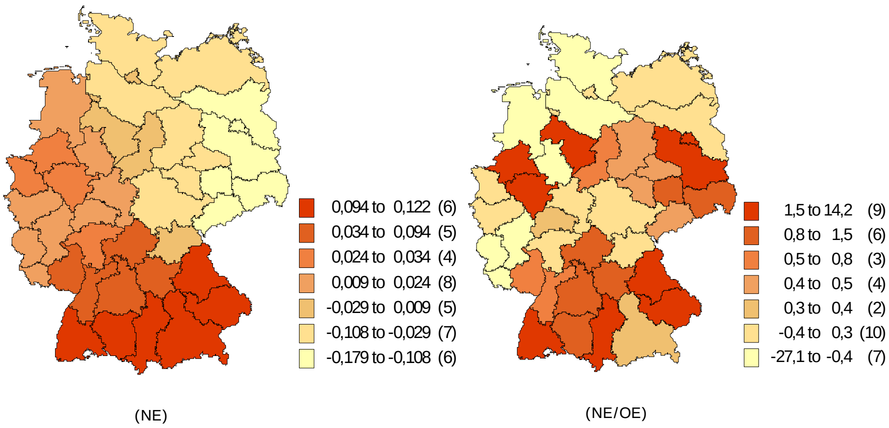

Figure 4.20: Absolute and relative neighbour effects

over all effect, because of the positive neighbouring effects. But, without these effects, a negative overall effect would occur. In this area you can find mainly West German regions, which profit from spillover regions, situated in the upper right regions. This is especially true for some Bavarian regions, such as "Schwaben", "Oberpfalz", "Niederbayern" who profit mainly from "Oberbayern" spillover centers like "Greater Munich area".

Regions in top left exhibit a negative overall effect, despite the fact, that the home effect is positive. In other words, if negative neighbouring effects did not affect those regions, those regions could be associated with a superior labour productivity. In this region you find primarily East German regions, which are compared to other East German regions are relative prosperous with respect to their economic development. This is especially the case for "Dresden", "Süd-Brandenburg". But also "Berlin" and "Braunschweig" can be found in this area.

Regions in the down left regions can be characterized as regions which require economical and political support and should therefore be in the focus of political debate when talking about the allocation of supranational grants. Neither their own labour productivity is superior, nor they can benefit from positive knowledge spillovers from neighbouring regions. Again, in this region we can find by majority East German regions. But, in contrast to East German regions which are settled in the upper left, those regions suffer from structural weaknesses. This is especially the case for "Nord-Brandenburg" but also for West German regions, such as "Saarland" and "Oberfranken". In this picture we can see that when talking about economic performance of German regions, the entity "Bundesländer" is to crude. For instance, it is not as easy as it seems ex ante to get a correct impression of the economic performance of "Sachsen". For NUTS-2 case Dresden performs rather well, whereas "Chemnitz" and

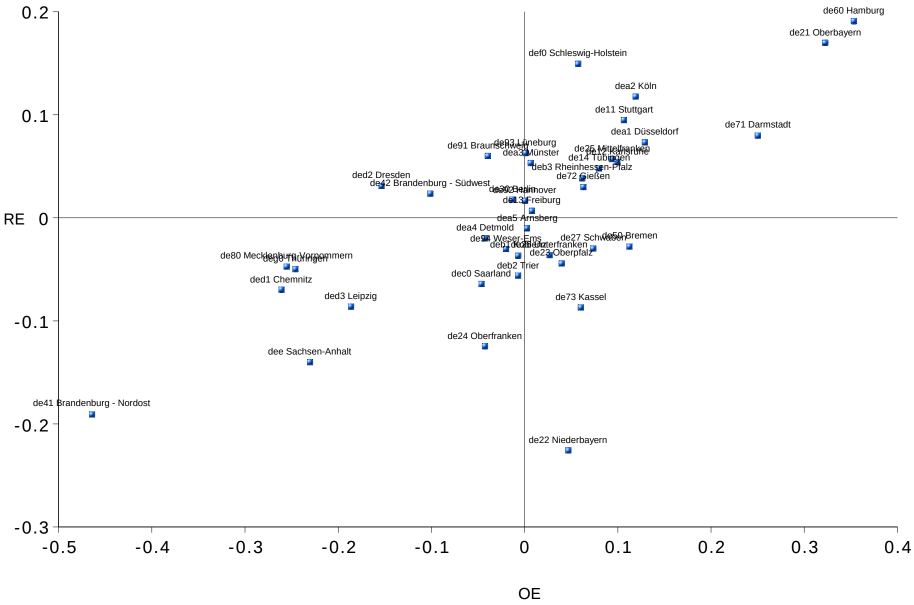

Figure 4.21: Regional and overall effect

"Leipzig" perform bad. This is also true for West German regions, especially "Bayern", which seems to be more heterogeneous than "Baden-Württenberg".

Figure 4.22, which has at the ordinate the neighbour effects and at the abscissa the overall effect has to be interpreted analogously as figure 4.21 and provides an alternative view on the same results obtained before. Again, some regions would exhibit a positive overall effect, unless negative neighbour effects are taken into consideration. Again, this is especially true for "Berlin", "Brandenburg-Südwest" and "Dresden" for instance.

4.6 Policy implications

The purpose of this section is to derive some implications for regional policy, based on simulation results which have been obtained in the last section. The core of regional policy since 1969, coordinated by German administration is the so called "Bund-Länder-Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur"" (GA). Regional policy in the German sense is a cooperation of the German countries and the Federal Republic of Germany which is controlled by Art. 91.a in the German constitution. But (GA) is not only a traditional funding instrument. (GA) is the framework of strategy, regulation and coordination of regional policy, also for EU related founds.

If we have again a look at figure 4.21, and more precisely, if we again take the upper left region. Then from a policy maker's view it should be clear that policy instruments are required which take into account that a comprehensive regional approach goes into

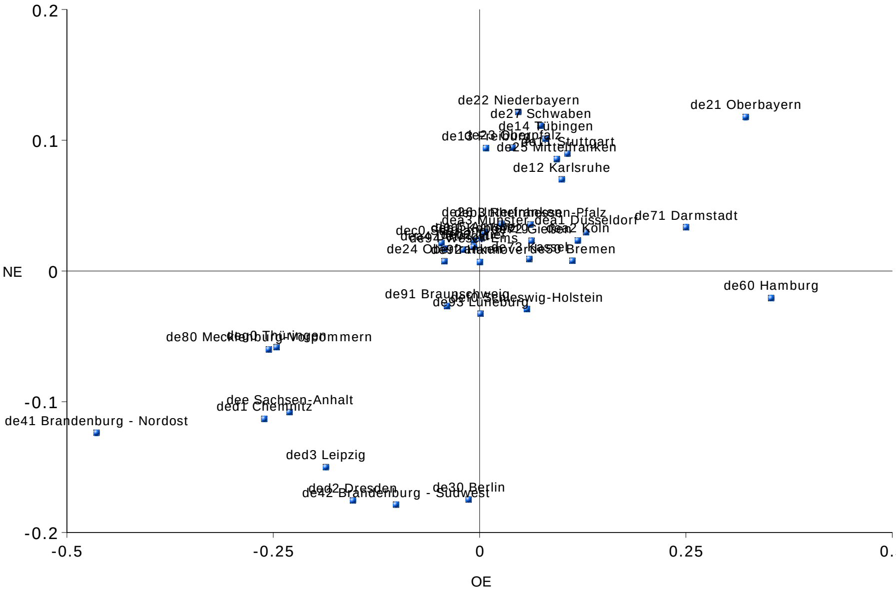

Figure 4.22: Neighbour and overall effect

the wrong direction. For those regions a mixture of traditional structural sponsorship should be supplemented by appropriate public-private-partnerships. Especially for East German regions regional policy has focused instruments which should hinder the Brain drain towards West German regions. Since 2005 the German economic department programme "Kooperationsnetzwerke und Clustermanagement" supports regional and national cooperation between companies, scientific and economy as well as between local administration to strength the network abilities and competitiveness of regions. This seems an appropriate policy instrument for those mentioned regions. For regions in the down left area in figure 4.21, traditional regional policy arrangements seem to be appropriated. Additionally, as mentioned by (Moll, 2000) EU region wide cooperations between country such as Germany and France near the German-French boarder or with Czech Republic to promote former so called borderlands or with Poland to promote close to boarder regions of "Mecklenburg" have been aspired. Thus, EU as German funding instruments includes a prominent regional component which aims to support the creation of regional clusters.

4.7 Conclusion

The aim of this chapter is, to give an alternative insight into the role of spatial dependence, not only in a theoretical spatial hybrid growth model context, but also in an empirical way. The main purpose of the theoretical model is to investigate the role of technological innovations and agglomeration which is a logical combination of the role of growth and innovativeness and the role of growth and agglomeration with respect to space. Space is an economic sphere which has been paid, without any exceptions, rather few attention in literature so far. To combine these aspects spatial knowledge spillovers are necessary. Particularly, within the theoretical model context it is assumed that spatial influence is not constant over space, an assumption which is not considered in the literature so far. After simulating the model, confirmation of the "'folk theorem of spatial economics"' has been found, that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities".

To test the implications of the derived theoretic model, we refer to a cross section production function approach, proposed by (Griliches, 1979), which should measure the effects of innovativeness, measured by knowledge capital, such as human capital, patents or $R\&D$ and spatial spillovers on output. This is done for German NUTS-2 regions. These administration level has been selected due to the fact, that referring on NUTS-3 regions could lead to spatial dependence by "construction" caused by streams of commuters for instance. Spatial econometricians methods have been employed to measure the before mentioned effects. Spatial heterogeneity is mostly neglected in hitherto empirical studies. Thus, employing a new model selection mechanism, which accounts for spatial heterogeneity and which is based both on Maximum-Likelihood and Bayesian methods, one can find that significant spatial knowledge spillovers exist in the data, even though they are small. Especially, patents spillovers have been detected as the driving forces of economic performance. Further, the selected model found that spatial heterogeneity matters. Controlling for spatial heterogeneity is important because neglecting it could lead to insufficient estimates. Until today, the majority of existing studies assume ex ante spatial homogeneity. This could be due to the fact, that Maximum-Likelihood methods are very clumsy for spatial model estimation. Coevally, Bayesian methods are still on the fringes, especially in spatial econometrics, although the conceptual idea of Bayesian methods are more eligible to cover spatial model design than Maximum- Likelihood methods so far. Hence, it can be expected that in the next years some improvements of Maximum-Likelihood methods will be made in terms of efficient spatial model estimation.

Another way to investigate spatial data, is to employ spatial filter methods. This method should be used, if spatial effects should be removed from data. In this context it is obvious to ask the question if regions specific economic strength benefits from economic activity from their neighbours or not. The filtering method is easy to implement and can be conducted with a traditional two step OLS procedure. One of the key findings is, that economic performance differs not primarily between East and West German regions, but is more complex. Especially for East German regions we find, that some well performing regions suffers in great extent from negative neighbour influence. This is also true for some West German regions but plays a minor role. Against this background it is rather logical, that cluster phenomena are suitable for explaining the distribution of economic activity of German NUTS-2 regions over space.

This cluster phenomena can be graphically replicated with a weakness-strength profile. To obtain this, on the basis of the employed filter method a simulation of labour productivity has been conducted. Using the simulated data it is found that especially, South German regions, such as "Bayern" and "Baden-Württemberg" and regions in the "Ruhr-Gebiet" perform well, due to their inherent economic strength. These regions do not rely on positive neighbour effects to beef up their economic performance. Therefore, these regions can be labeled as knowledge generation areas. On contrary, some regions would perform significantly better, if negative spillover from neighbouring regions could be eliminated. This is particularly true for "Brandenburg-Süd" and "Dresden".

What are the political implications? As mentioned above, EU has launched several economic policy programmes to foster regional economic performance. Most of the EU related programmes have recognized the outstanding role of knowledge and economic clusters for regional development. Knowledge spillovers, generated by knowledge generation areas, such as "Munich Greater Area"' should contribute to boost neighbouring regions, which suffer from insufficient knowledge generating potential. Hence, regional politics is on the right track, but should provide further incentives for strengthen regional knowledge networks.

4.8 Appendix

4.8.1 Appendix 1

Proposition: An OLS estimation of a spatial lag model would yield inconsistent and thus biased estimators. An OLS estimation of a spatial error model would yield inefficient but unbiased OLS estimators. An OLS regression of a spatial model with exogenous spatial lagged variables is unbiased but only blue if spatial homogeneity is assumed. □

Let us start with

$$y = \tilde{y}P + X^{+++}\beta^{+++} + \tilde{\epsilon}\Lambda + \kappa, \tag{4.75}$$

or with the familiar notation from expression 4.26:

$$y = \tilde{y}P + X\beta^{X} + \tilde{X}^{++}\beta^{++} + \tilde{\epsilon}\Lambda + \kappa \tag{4.76}$$

with

$X^{+++} = [X, \tilde{X}^{++}]$ and $\beta^{+++} = [\beta^{X}, \beta^{++}]$. Labeling parts of 4.26 with I, II, III, IV, this yields:

$$y = \underbrace{\tilde{y}P}_{(I)} + \underbrace{X\beta^{X}}_{(II)} + \underbrace{\tilde{X}^{++}\beta^{++}}_{(III)} + \underbrace{\tilde{\epsilon}\Lambda}_{(IV)} + \kappa. \tag{4.77}$$

1. Assume, that I=II=IV=0.

   This yields

   $$y = \tilde{X}^{++}\beta^{++} + \kappa, \tag{4.78}$$

   with $\kappa \sim (0, \sigma^2\Omega)$ with $\Omega \neq \sigma^2 I$. From equation 4.78 we can obtain an OLS estimator $b^{++} = (\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}y$. This estimator is unbiased if $E[\tilde{X}^{++\prime}\kappa] = 0$ because:

   $$E[b^{++}] = \beta^{++} + E[(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\kappa] = \beta^{++}. \tag{4.79}$$

The estimated variance covariance matrix of $V[b^{++}]$ is

$$V[b^{++}] = E[(b^{++} - \beta^{++})(b^{++} - \beta^{++})'] = \tag{4.80}$$

$$= E[(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\kappa\kappa'\tilde{X}^{++}(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}], \tag{4.81}$$

or

$$V[b^{++}] = [(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\Sigma\tilde{X}^{++}(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}] \neq \sigma^2(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}, \tag{4.82}$$

with $\Sigma = \sigma^2\Omega$. Thus, OLS is unbiased but only BLUE if $\sigma^2\Omega = \sigma^2 I$, thus if spatial homogeneity is assumed. ■

2. Assume, that III=IV=0.

   This yields

$$y = \tilde{y}P + X\beta^X + \kappa, \tag{4.83}$$

   with $\kappa \sim (0, \sigma^2\Omega)$ with $\Omega \neq \sigma^2 I$. An OLS estimation of 4.83 would yield, under neglecting an element of $\tilde{y}P$:

$$b^X = (X'X)^{-1}X'y. \tag{4.84}$$

   After inserting y this yields the estimator b^X of β^X:

$$b^X = (X'X)^{-1}X'(\rho^r W^{+r}y + X\beta^X + \kappa). \tag{4.85}$$

   Taking the expectation of 4.85 this yields

$$E[b^X] = E[(X'X)^{-1}X'(\rho^r W^{+r}y) + \beta^X] \neq \beta^X. \tag{4.86}$$

   Thus, the bias can be expressed as

$$E[b^X - \beta^X] = E[(X'X)^{-1}X'(\rho^r W^{+r}y)] = \rho^r\beta^l. \tag{4.87}$$

   The expression can be interpreted as ρ^r times the regression of X against $(\rho^r W^{+r}y)$ with the corresponding β^l which is equal to the expected value of the regression $(W^{+r}y)$ on X. Hence, OLS is biased if only one component of $\tilde{y}P$ is neglected in the regression. ■

3. Assume, that II=III=IV=0

   In this case we obtain

$$y = \tilde{y}P + \kappa, \tag{4.88}$$

   with $\kappa \sim (0, \sigma^2\Omega)$ with $\Omega \neq \sigma^2 I$. An OLS-estimator of one element ρ^r of P would yield:

$$\hat{\rho^r} = [(W^{+r}y)'(W^{+r}y)]^{-1}(W^{+r}y)'y, \tag{4.89}$$

   or inserting 4.88 in 4.89

$$\hat{\rho^r} = \rho^r + [(W^{+r}y)'(W^{+r}y)]^{-1}(W^{+r}y)'\kappa. \tag{4.90}$$

   The estimator $\hat{\rho}$ is not consistent because

$$\hat{\rho} \xrightarrow{p} \rho + \text{plim}\left(\frac{1}{N}(W^{+r}y)'W^{+r}y\right)^{-1} \text{plim}\left(\frac{1}{N}(W^{+r})'\kappa\right),$$

$$\hat{\rho} \xrightarrow{p} \rho + S^{-1}_{(W^{+r}y)'W^{+r}y}\,\text{plim}\left(\frac{1}{N}\kappa' W^{+r}(I - \rho^r W^{+r})^{-1}\kappa\right). \tag{4.91}$$

   The expression plim $\left(\frac{1}{N}(W^{+r}y)'W^{+r}y\right)$ converges to a regular and finite matrix $S_{(W^{+r}y)'W^{+r}y}$. The second term plim $\left(\frac{1}{N}(W^{+r})'\kappa\right)$ however converges to an expression which is quadratic in the errors, unless $\rho^r = 0$. Hence, estimating a spatial lag parameter ρ^r via OLS is biased and inconsistent. ■

4. Assume, that I=III=0

   Now it results

$$y = X\beta^X + \epsilon, \tag{4.92}$$

   with $\epsilon = \epsilon\Lambda + \kappa$ and $\kappa \sim (0, \Omega)$ with $\Omega = \sigma^2 I$. An OLS estimation of β^X would be unbiased, because

$$E[b^X - \beta^X] = E[(X'X)^{-1}X'\kappa] = 0, \tag{4.93}$$

   but β^X is not efficient because for a given λ^r from Λ we get for the estimated variance covariance matrix

$$V[b^{++}] = [(X'X)^{-1}X'\epsilon\epsilon' X(X'X)^{-1}X], \tag{4.94}$$

which is

$$V[b^{++}] = \sigma^2[(X'X)^{-1}X'[(I-\lambda^r W^{+r})'(I-\lambda^r W^{+r})]^{-1}X(X'X)^{-1}X]. \tag{4.95}$$

In consequence, OLS estimator of b^X is unbiased but inefficient for a given λ^r from Λ.
■

4.8.2 Appendix 2

Using PIM, the capital stock K_t can be computed as

$$K_t = \iota_0 I_t + \iota_1 I_{t-1} + ... + \iota_T I_{t-T}, \tag{4.96}$$

with I_t as investment in new capital K_t. It is common to set $\iota_0 = 1$ and $\iota_t = (1-\delta)^t$ for $t > 0$. Using a Koyck transformation we get from equation 4.96

$$K_t = I_t + (1-\delta)K_{t-1}, \tag{4.97}$$

with $\delta = \frac{\iota_{T-1}-\iota_t}{\iota_{T-1}}$. To obtain expression 4.30 we assume that investment I_t in stock of capital K_t is growing from $t=0$ with constant rate ζ. Therefore we can write:

$$I_t = (1+\zeta)I_{t-1} = (1+\zeta)(1+\zeta)I_{t-2} = ... = (1+\zeta)^{\infty^+} I_{t-\infty^+}. \tag{4.98}$$

Further it is assumed that devaluation of capital K_t follows a geometric series:

$$K_t = I_t + (1-\delta)I_{t-1} + (1-\delta)^2 I_{t-2} + ... + (1-\delta)^{\infty^+} I_{t-\infty^+}. \tag{4.99}$$

Using 4.98 and 4.99 leads to

$$K_t = I_t + \left(\frac{1-\delta}{1+\zeta}\right) I_t + \left(\frac{1-\delta}{1+\zeta}\right)^2 I_t + ... + \left(\frac{1-\delta}{1+\zeta}\right)^{\infty^+} I_t = I_t \sum_{\kappa=0}^{\infty^+} \left(\frac{1-\delta}{1+\zeta}\right)^{\kappa}. \tag{4.100}$$

Rearranging equation 4.100 by writing:

$$\left(1 - \left(\frac{1-\delta}{1+\zeta}\right)\right) K_t = I_t \left(1 - \left(\frac{1-\delta}{1+\zeta}\right)^{\kappa+1}\right) \tag{4.101}$$

and letting $\kappa \to \infty$ leads to

$$K_t = I_t \frac{1}{\left(1 - \left(\frac{1-\delta}{1+\zeta}\right)\right)}, \tag{4.102}$$

because of $\left(\frac{1-\delta}{1+\zeta}\right) < 1$. Noting, that $I_{t+1} = I_t(1+\zeta)$ yields

$$K_t = \frac{I_{t+1}}{1+\zeta} \frac{1}{\left(1 - \left(\frac{1-\delta}{1+\zeta}\right)\right)}. \tag{4.103}$$

4.8.3 Appendix 3

The log-likelihood function for the spatial variant of a ARMA(1,1), SAC(1,1) can be derived as follows.[91]

$$y = \rho^1 W^{+1} y + X^{+++} \beta^{+++} + \epsilon, \tag{4.104}$$

with

$$\epsilon = \lambda^1 W^{+1} \epsilon + \kappa, \tag{4.105}$$

with $\kappa \sim (0, \sigma^2 I)$ can be written as

$$\xi = \frac{1}{\sigma}(I - \lambda^1 W^{+1})^{-1}[(I - \rho^1 W^{+1})y - X^{+++}\beta^{+++}] \tag{4.106}$$

with $\xi \sim N(0, I)$. The corresponding determinant of the Jacobian $\mathcal{J} \equiv det\frac{\partial \xi}{\partial y}$ can be rewritten as

$$\mathcal{J} \equiv det\frac{\partial \xi}{\partial y} = |\frac{1}{\sigma}[I - \lambda^1 W^{+1}]\,||\,[I - \rho^1 W^{+1}]|. \tag{4.107}$$

Employing the fact that $\xi \sim N(0, I)$ we can write the log-likelihood for the joint distribution as

[91] The proof is based on (Anselin, 1988), p. 74 with some minor adjustments.

$$\ln \mathcal{L} = -\frac{N}{2} \ln 2\pi - \frac{N}{2}(\sigma^2) + \ln |\tilde{N}| + \ln |[I - \rho^1 W^{+1}]| - \frac{1}{2}\xi'\xi, \tag{4.108}$$

If $\rho^1 = 0$ then the log-likelihood 4.47 results.

4.8.4 Appendix 4

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$				
Independent variables $\mathbf{x} \in \mathbf{X}$				
Preferred model	Number of parameters	$\ln(\mathcal{L})$	$\hat{\rho}$	$\hat{\lambda}$
Model 4 [SAC(1,1)]	12	90.47	0.133	$0.697^{\natural}$
Model 4 [SAC(1,2)]	12	88.55	0.284	-0.989
Model 4 [SAC(2,1)]	12	90.28	-0.000	$0.722^{\natural}$
Model 4 [SAC(2,2)]	12	87.74	-0.000	-0.987
Model 4 [SEM(1)]	11	67.93	—	$0.711^{\natural}$

$^{\diamond}$ Selected model. † indicates 10% significance. ‡ indicates 5% significance. $^{\natural}$ indicates 1% significance

Table 4.11: Comparison of selected models

4.8.5 Appendix 5

In the relevant literature, there are some convergence checks for convergence of MCMC based samplers for linear models. In this section there is given a short motivation of some convergence checks instruments. All below mentioned diagnostic tools are implemented in the Matlab function "coda".

4.8.5.1 Autocorrelation estimates

From time series it is known that if ρ is a stationary correlated process, then $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho_i$ is a consistent estimate of $E(\rho)$. Therefore it is allowed to simulate some correlated draws from our posterior distribution to get a hint how many draws we need for uncorrelated draws for our Gibbs sampler. A high degree of correlation should cause someone to carry out more draws which should result in a sample which allows to draw correct posterior estimates.

4.8.5.2 Raftery-Lewis diagnostics

(Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) have suggested a set of diagnostic tools which they have first implemented in `FORTRAN` named "Gibbsit". This function was converted in `Matlab` and called "raftery". (Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) have focused on the quantiles of the marginal posterior. The diagnostic itself is based on the properties of a two state Markov-Chain, because for a given quantile the chain is dichotomized using a binary time series that is unity, if $\rho_i \leq q_{quant}$ and zero otherwise, where q_{quant} denotes the quantile which has to be chosen from the researcher ex ante. For an independent chain, the zeros and ones should be appear randomly. The "coda" function prints the so called thinning-ratio, which is an indicator of autocorrelation in the draws. "Thinning" means, that only every third, fifth,... draw for instance are saved for inference, because the draws from a Markov Chain are not independent. Additionally, the number "burn-in-draws" are reported. The number of "burn-in-draws" are excluded from sampling based on inference. Finally, the I-statistic is reported which is the ratio of the number of total draws and the minimum number of draws to ensure an i.i.d. chain, represented by the draws. (Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) indicate that values larger than 5 exhibit convergence problems of the sampler and therefore, more draws should be carried out.

4.8.5.3 Geweke diagnostics

The `Matlab` function "coda" additionally estimates the numerical standard errors and relative numerical standard errors based on the work of (Geweke, 1992). The code can be found at http://www.biz.uiowa.edu/cbes/code.htm, which is based on BACC. The BACC code itself as `Matlab`, `R` and `S-Plus` routines can be found at
http://www2.cirano.qc.ca/ bacc/bacc2003/index.html. This diagnostics are based on elements of spectral analysis. From time series analysis we know, that an estimate of variance of ρ is based on $Var[\hat{\rho}_i] = \frac{\Delta(0)}{k}$ with Δ_0 as the spectral density of ρ_i evaluated at ω_0 of $\Delta(\omega)$. The question is, how to approximate $\Delta(\omega)$. For this reason, alternative tapering of the spectral window should be used. Using numerical standard errors and relative numerical i.i.d. standard errors and compare them with numerical standard errors and relative numerical standard errors from the tapered version. If the relative numerical standard error of the tapered version is close to one, then convergence seems to be ensured.

4.8.5.4 Geweke-χ^2 test

Geweke's-χ^2 test is based on the intuition that sufficiently large draws have been taken, estimation based on the draws should rather identical, provided the Markov chain has reached an equilibrium state. This test is a simple comparison of the means for each split of the draws. In this work, the χ^2 test, based on the null hypothesis of equality of the means of splits is carried out for each tapered case.

It should be mentioned that the diagnostic tools introduced here are not foolproof and sometimes MCMC diagnostic tools lead to misleading decisions.[92]

4.8.6 Appendix 6

For appendix 6, please refer to the following pages.

[92]Refer for this topic to (Koop, 2003), p. 66.

Dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.800	0.414	0.205	-0.035	—	—
σ	0.410	0.022	-0.029	0.009	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	18	4736	937	5.054	—
σ	1	18	4736	937	5.054	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.561134	0.187292	0.001922	1.000000	—	—
σ	0.003283	0.001214	0.000012	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.007401	0.067106	0.007782	0.060969	0.008809	0.047582
σ	0.000014	0.767126	0.000014	0.830173	0.000013	0.904161
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.569899	0.002264	0.000000	—	—	—
4% taper	0.565011	0.008253	0.057527	—	—	—
8% taper	0.562932	0.008323	0.055775	—	—	—
15%taper	0.565765	0.008427	0.065577	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003270	0.000015	0.378822	—	—	—
4% taper	0.003267	0.000021	0.592235	—	—	—
8% taper	0.003267	0.000018	0.536563	—	—	—
15%taper	0.003266	0.000017	0.530511	—	—	—

◇ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 4.12: MCMC-convergence summary for model (4,1)

Dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.734	0.245	0.076	0.039	—	—
σ	0.457	0.034	-0.002	0.035	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	19	5047	937	1.000	5.386
σ	1	19	5047	937	1.000	5.386
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.555595	0.183009	0.001878	1.000000	—	—
σ	0.003266	0.001241	0.000013	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.006591	0.81155	0.006416	0.085646	0.005519	0.115728
σ	0.000018	0.528629	0.000015	0.682827	0.000014	0.856455
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.559135	0.002221	0.000089	—	—	—
4% taper	0.560335	0.006356	0.152210	—	—	—
8% taper	0.561610	0.006683	0.162430	—	—	—
15%taper	0.560764	0.006756	0.173295	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003255	0.000015	0.380291	—	—	—
4% taper	0.003257	0.000022	0.582485	—	—	—
8% taper	0.003259	0.000019	0.572221	—	—	—
15%taper	0.003259	0.000018	0.547883	—	—	—

◊ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 4.13: MCMC-convergence summary for model (4,2)

Dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.000	-0.003	0.015	-0.005	—	—
σ	0.146	-0.006	0.000	-0.002	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	2	974	937	1.039	—
σ	1	2	974	937	1.039	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.156684	0.193235	0.001983	1.000000	—	—
σ	0.008386	0.002139	0.000022	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.002051	0.934374	0.001926	1.059081	0.001864	1.131157
σ	0.000024	0.864318	0.000023	0.924805	0.000023	0.941651
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.159254	0.002375	0.694409	—	—	—
4% taper	0.159402	0.002474	0.689738	—	—	—
8% taper	0.159500	0.002392	0.672743	—	—	—
15%taper	0.159553	0.002236	0.648753	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.008387	0.000026	0.285231	—	—	—
4% taper	0.008380	0.000031	0.325237	—	—	—
8% taper	0.008381	0.000031	0.329946	—	—	—
15%taper	0.008380	0.000030	0.303602	—	—	—

◇ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 4.14: MCMC-convergence summary for model (4,3)

Dependent variable **y**: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.677	0.160	0.025	-0.003	—	—
σ	0.440	0.032	0.004	-0.006	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	15	4023	937	4.293	—
σ	2	15	4023	937	4.293	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.541436	0.189626	0.000601	1.000000	—	—
σ	0.003268	0.001236	0.000004	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.001422	0.178616	0.001433	0.176005	0.001312	0.210008
σ	0.000007	0.296403	0.000007	0.305018	0.000007	0.285996
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.541633	0.000722	0.013795	—	—	—
4% taper	0.541560	0.001743	0.299201	—	—	—
8% taper	0.541598	0.001789	0.316022	—	—	—
15%taper	0.541719	0.001697	0.306570	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003267	0.000005	0.018101	—	—	—
4% taper	0.003269	0.000009	0.173843	—	—	—
8% taper	0.003273	0.000008	0.135064	—	—	—
15%taper	0.003275	0.000007	0.077024	—	—	—

⋄ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 4.15: MCMC-convergence summary for model (4,4)

5 Conclusions

The aim of this chapter is to draw some major conclusions from the previous chapters of this thesis. The first section gives a summary of the contents of the thesis. Particularly, it focuses on the results obtained. The second section provides an overview of possible revenues for further research.

5.1 Summary

In the introductory chapter it has been laid out, that knowledge as an input factor of production exhibits a strong influence on economic development. The increasing knowledge intensity in the globalised economy needs to focus on the determinants of the "knowledge based society". Two major determinants on which the "knowledge based society" and its economic analogon the "knowledge based economy" rely, are the creation and the diffusion of knowledge. The main motivation for this thesis stems exactly from the importance of knowledge and "knowledge diffusion" for the "knowledge based economy" and finally for the modern economic theory and empirics. As mentioned in the first chapter of this thesis, knowledge diffusion topics are not only considered as a cornerstone of modern growth literature and of new economic geography but is also important for microeconomic related fields, such as dynamic applications of industrial organization. As mentioned in the introduction, the economic field of "knowledge diffusion" literature is widespread and many applications which cover the topic "knowledge diffusion" can be found in literature which are on the one hand, microeconomic based but on the other hand knowledge diffusion is also a relevant topic in macroeconomics, especially in (regional) growth theory. As a consequence, only some recent topics of "knowledge diffusion" literature have been outlined in this thesis. As mentioned in the introduction of this thesis, particularly knowledge diffusion in the context of dynamic industrial organization and knowledge diffusion in the context of new economic geography are currently discussed in the relevant literature but also generates revenues for further research. That defines the field where the contributions of this thesis set in.

The first two chapters after the introduction are direct applications from dynamic industrial organization. The first chapter deals with the question how knowledge transfer affects knowledge diffusion, whereas the second chapter tackles the relationship between firm size, innovation, market structure and learning.

Knowledge transfer and knowledge diffusion are two sides of one medal. Knowledge transfer is defined as the pure exchange process of knowledge between sender to receiver. Particularly, knowledge networks can be considered as the ideal environment in which sender and receiver of knowledge come together. But as mentioned in this chapter, knowledge transfer is not a sufficient condition for knowledge diffusion. Knowledge diffusion is completed if transferred knowledge can be understood and used by the receiver. Thus, it is worth to integrate both aspects, knowledge transfer and knowledge diffusion in a comprehensive knowledge framework of industrial organization. The so called (Bass, 1969) model which is referred to in this chapter stems originally from product diffusion literature and is very popular in applied diffusion research and some disciplines of business administration such as marketing. The idea of the (Bass, 1969) model is pretty simple. The before mentioned contribution assumes, that two groups of adopters, so called innovators and imitators have to decide when they should adopt a certain technology or product. The adoption decision is influenced by external and internal factors, such as marketing effort and communication between these groups. But this model has some limitations. One major drawback of this model is that (Bass, 1969) does not replicate the behaviour of these subgroups of adopters in a notational form. In the recent years several extensions of the (Bass, 1969) model have been proposed. But these models are all less suitable to cover the aspect of knowledge diffusion and knowledge transfer. Therefore the aim of this chapter was to setup a model which first, integrates innovators and imitators as well as their specific adoption decision of new knowledge. Thereby, so called network effects have been acknowledged. According to the network structure, knowledge transfer is easier or more difficult. If a dense network structure is available, "knowledge transfer" is easier and thus the imitator should adopt faster. On contrary, if networks do not exist, knowledge transfer is excluded and thus adoption takes place later. The latter scenario often leads to the so called "chasm" pattern between early and late adoptions, which is extensively discussed in diffusion related literature. In consequence, network effects should also have an influence on the shape of the adoption curve, which is in the latter case not necessarily unimodal but bimodal for the entire market. The point is, that the introduced model treats the "chasm" pattern as endogenous, not as a given exogenous number. The literature is still silent on this topic and only a few micro based paper take these network effects into account. Additionally, the model was extended towards a stochastic knowledge diffusion model to capture the idea that uncertainty of adoption is a function of time, which means at the beginning and at the end of the diffusion process uncertainty regarding the adoption should tend to zero, while in the middle of the diffusion process uncertainty of adoption is high. Another feature of the proposed model is, that it can be adopted directly to empirical research.

In a simulation study it was shown, that the shape of the adoption pattern depends on, whether knowledge diffusion occurs or not. If knowledge transfer occurs and the stronger network effects are, so called unimodal patterns are more probable, because right before innovators have realized the inflection point, imitators have nearly reached their respective inflection point. On the contrary, the longer the discrepancy between the realization of the in inflection point of innovators and the beginning of imitators adoption is, the less important network effects are, the more probable the so called bimodal adoption phenomena are. Thus "chasm" patterns of adoption curves occur if network effects are from less importance.

As laid out in the end of the second chapter, the advantage of this new model is twofold: from a theoretical point of view, not only the so called unimodal diffusion phenomena in an uncertain environment can be replicated, but also the bimodal diffusion phenomena can occur. From an empirical point of view, the model which incorporates heteroscedastic errors and mean reverting can be theoretically estimated directly with a SUR approach.

Thematically closely related to the second chapter is the third chapter, with the explicit focus laying on firm level size. The exploring of so called feedback processes between innovation, market share and firm size has gained much attention in recent years. For many years, the effects of innovation and firm size and the relationship between market share evolution and innovation have been discussed isolated. As mentioned in this chapter, there subsists a large body of literature covering the relationship between firm size and innovation, which are primarily focused on manufacturing industries. The majority of the relevant literature is heavily empirical based and is somehow ambiguous with respect to the effects of firm size on innovation. Regarding to to firm size there is no clear evidence, whether small or large firms are more innovative.

If we now turn the focus on the effect of market structure on innovation, in principle two different scenarios are imaginable: the first is, that a positive relationship between monopoly power and innovative activity can be assumed, the second is, that innovative activity suffers from monopoly power. The first as well as the second relationship are empirically documented in a voluminous literature. Again, in the relevant literature we find no clear effect of market structure on innovation. Some studies found negative correlations between market structure and innovation whereas other detected an inverted U-pattern between market structure and innovation. The latter reflects the fact, that insufficient market power hinders firms to reduce so called up-front R&D

effort, whereas an increasing market power reduces the incentive to engage in further R&D effort.

Within an evolutionary framework and with the support of replicator dynamics studies it was found, that small firms benefit from decreasing returns to scale, while large firms can use learning-by-doing activities to tighten their market dominance. The latter is based on the assumption of increasing returns to scale, which coincides with the "Schumpeter-Mark-II" hypothesis. As a result, small firms will be swamped out off the market if technological progress is not too fast for any case of returns to scale. On the other side, several authors highlighted the fact, that inter-firm cooperation, based on knowledge sharing which is boosted up itself by learning activities can explain the predominance of small firms in the market. Hence, learning activities and knowledge diffusion play an important role when exploring feedback processes between innovation, market share and firm size. The latter fact cannot be replicated by the model proposed by (Mazzucato, 2000).

Therefore, the aim of this study is to work out the reciprocal dependences of firm size, innovation, market structure, knowledge diffusion and learning. Learning effects are covered by psychologically motivated learning curve aspects and are endogenously driven by agent specific characteristics, such as talent and a historically given stock of knowledge. This model constitutes an extension of the work of (Mazzucato, 2000) that it explicitly introduces a channel of knowledge diffusion, which is endogenously determined by learning activities, which are themselves endogenously influenced. To integrate both aims, the so called replicator dynamics approach is disposed. The employed tool stems from evolutionary economics and is based on Darwin´s principle of natural selection. Particularly, on the basis of simulation experiments it will be investigated how learning and knowledge diffusion affect market structure. Further, with this model it can be proofed whether learning activities need a dilution of one of the stylized facts regarding firm size dynamics which states, that early stages of an industry life cycle are characterized by instable market patterns. This is the first time that a replicator dynamics approach is combined with psychological motivated learning curves, which seems appropriate to cover knowledge based learning effects.

On the basis of a conducted simulation study it is shown that for any degree of technological progress, small firms still remain in the market, also for the case of increasing returns to scale (IRS), where large firms are in advance. Hence, this model is able to replicate the fact, that small firms are more likely to benefit from knowledge networks and thus from spillovers which define a source of innovativeness, from which large firms cannot profit. But integrating (dynamic) learning effects leads to a damping of market fluctuations.

In contrast to the second and the third chapter which fit closer to the dynamic industrial organization literature, the fourth chapter deals with the spatial dimension of knowledge diffusion, which is quite popular. Nowadays, space as an economic number has gained much attention, after it has been neglected or even ignored for a long time. Since the rising popularity of the new trade theory, which is basically designed to explain trade patterns between countries or regions the question arises, what is the role of space within an economic system. The new economic geography, which goes a step further than the new trade theory, mainly focuses on the explanation of trade patterns and on the location decisions of firms. Both aspects can only be answered superficially without explicitly acknowledging the economic role of space.

Until today, new economic geography applications, which cover knowledge diffusion topics are mainly empirically orientated and suffer from theoretical justification. We have a relative precise understanding about the grasp of knowledge diffusion but this aspect is not treated in regional growth literature. Thus, the first aim of the fourth chapter was to integrate the so called "folk theorem of spatial economics", which states that increasing returns to scale are essential for understanding the geographical distributions of economic activity, in a hybrid two sectoral regional growth model with an explicit focus on different grasps of knowledge diffusion. Although the model setup is deliberately kept simple, it becomes complex if integrating the aspect of knowledge diffusion with an explicit spatial focus. Knowledge diffusion exhibits feedback loops and this has to be acknowledged further in the model setup. Consequently, a solution by hand cannot be derived anymore and numerically simulation methods have to be employed. With the focus on space, it was decided to refer to cellular automaton programming technique because it is ideally designed to cover spatial phenomena. After performing several simulation studies for different grasps of knowledge diffusion as main results it can be concluded that first, for increasing returns to scale the model exhibits large spatial concentration and further exhibits an uneven distribution of per capita production. This is only partly the case of constant returns to scale but entirely not for decreasing returns to scale. As a main result, the model seems to justify the "folk theorem of spatial economics".

As mentioned above, the majority of economic applications treating knowledge diffusion topics is empirically based. In recent years, a new discipline, so called spatial econometrics, which can be labeled as a pendant to spatial economics, has emerged and has been established rather broadly in econometric society. Of course, this discipline is developing with respect to new estimation methods. Particularly, it seems that Bayesian methods are very attractive because of their inherent conception of allowing for priors. Moreover, Bayesian methods do a great job within spatial econometrics, particularly when talking about spatial heterogeneity, which means that variances of observations are not constant over space, or outliers exist. The classic or frequentist methods today are not able to deal sufficiently with the phenomena of spatial heterogeneity. From this point it is rather astonishing, that most of spatial econometrics applications are still based only on the frequentist methods, despite their inherent limitations.

The second aim of this chapter is to find out on the basis of a conducted cross section analysis, to what extent knowledge spillovers do contribute to regional growth. For this purpose it has been referred to German NUTS-2 regions. Data for human capital and patents are used to proxy regional specific knowledge stock. Further it has been referred on data for labour and regional specific capital stock to explain regional specific labour productivity. For this purpose spillover variables have to be employed to account for spatial knowledge spillover. It was shown that knowledge spillovers are more local than global and hence only first order neighbour effects have been included in the regression context. The estimation has been first performed with the frequentist methods with the assumption of spatial homogeneity and further it was conducted with a Bayesian approach which explicitly controlls for spatial heterogeneity. As a main result, labour productivity can be described at best with a spatial moving average process. Further it has been shown, that spatial heterogeneity matters and ignoring them would lead to insufficient parameter estimates. Additionally it results, that neighbourhood‘s first order patent activity has a significant influence on own labour productivity.

Until now, we only know, that spatial phenomena matter. But we do not know, to what extent regions benefit or suffer from spatial neighbouring effects. Because of this fact, as a next step a filtering procedure was employed to remove spatial effects from regional per capita production. For this reason first regional explanatory variables have been filtered and second, regions specific labour productivity has been further corrected for spatial neighbouring effects. The aim of this filtering procedure is to create a strength weakness profile in which German NUTS-2 regions are embedded. As a main conclusion it has been shown, that mainly East-German but also some West-German regions suffer more than benefit from spatial neighbouring effects. Particularly, some South-German regions such as "Oberbayern" or "Stuttgart" do not rely on spatial effects from the neighbourhood, because these regions benefit mostly from their own innovative potential. Therefore, these regions can be described as knowledge creation centers. Hence, we can see that German NUTS-2 regions are very heterogeneous with respect to their economic performance and only focusing on East-West differentials seems to be hasty. From this background, it seems plausible that economic policy regarding to knowledge diffusion should have not only a national but also a regional focus.

The following section will allude some revenues for further research, which are based on the previous chapters of this thesis.

5.2 Prospects for future research

As mentioned before in this section, the thesis contributes to some specific topics of knowledge diffusion. Taking the obtained results into consideration, there are several avenues for further research. In this section some ideas for further research are given, which could extent the level of knowledge derived from this thesis.

The model proposed in chapter two is very strict with respect to the assumption that the market saturation level is exogenous and constant over time. Therefore, an attempt to extent this model could be to endogenize the market saturation level. Second, from a technical point of view, mean reverting is assumed to be the same over the entire population. Thus, a further source of heterogeneity could be introduced in the model by assuming different values for ζ. Third, after examining the large and small sample properties via a Monte-Carlos-Simulation of the derived model the forecasting ability should be of major interest.

One interesting extension of the derived model in chapter three would be to integrate location decision of firms in this model setup. With this extension it will be possible to cover explicit cluster effects of firms and further controll for large and small firms. Also the derived results of chapter three should be tested empirically on the basis of an industry level dataset. Particularly, it can be tested to what extent firm productivity growth can contribute to explain aggregate productivity over all industries. Further one can ask the question, to what extent do location decisions change the results. Hence, one possible research question could be whether a significant relationship of location decision and industry specific productivity growth can be detected.

The proposed hybrid regional growth model can be expanded in several ways. On principle, every semi-endogenous or endogenous growth model could be embedded into a cellular automaton frame to controll explicitly for spatial dependence. With respect to the proposed model, a next step could be trying to embed this model in a general equilibrium environment. Further a convergence study could be conducted on the proposed model to answer the question whether regions exhibits convergence. This is an important question with respect to the mentioned aim of regional coherence formulated within the Lisbon strategy supported by the EU.

With respect to the proposed spatial econometric model there are many aspects which could be included. First of all, the cross-section analysis should be expanded by a spatial panel-data analysis, based on GMM to obtain a deeper insight into spatial knowledge spillover and its effects on regional economic performance. It has to be mentioned that spatial panel-data methods which were limited to a balanced panel application are recently expanded towards an unbalanced panel approach. Although there has been made much progress in the development of spatial panel-data during the last four years, literature is still silent to the question, whether GMM based fixed effects or random effects estimation is still valid even if spatially lagged endogenous variables are taken into account. From a non-spatial-panel-data perspective this is definitely not the case. Another innovative idea for further research relates to the assumption that spatial heteroscedasticity is itself spatially correlated. Particularly for this interesting topic, until today the relevant literature has not found an appropriate answer. This is also true for spatial unit roots, whose existence is mostly excluded ex ante. This procedure is from an econometricians point of view not convincing and therefore future research should be focused on the development of analytical instruments coping with particularly spurious spatial phenomena.

Bibliography

Abernathy, W. J. and Wayne, K. (1974). *Limits to the learning curve.* In: Harvard Business Review, 52, pp. 109-120.

Abramowitz, M. and Stegun, C. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* New York.

Acs, Z. and Audretsch, D. B. (1987). *Innovation, Market Structure, and Firm Size.* In: The Review of Economics and Statistics, 69, pp. 567-574.

Aghion, P. and Howitt, P. (1992). *A Model of Growth through Creative Destruction.* In: Econometrica, 60, pp. 323-351.

Aghion, P. and Howitt, P. (1998). *Endogenous growth theory.* Cambridge.

Almus, M. and Czarnitzki, D. (2003). *The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany.* In: Journal of Business and Economic Statistics, 21, pp. 226-236.

Anderson, J. R. and Schooler, L. J. (1991). *Reflections of the environment in memory.* In: Psychological Science, 2, pp. 396-408.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht.

Anselin, L. (1992). *SpaceStat tutorial: A workbook for using SpaceStat in the analysis of spatial data.* Morganton, West Virginia: Regional Research Institute, West Virginia University.

Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook.* Can be downloaded from www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf.

Anselin, L., Bera, A., Florax, R. J. G. M., and Yoon, M. (1996). *Simple diagnostic tests for spatial dependence.* In: Regional Science and Urban Economics, 26, pp. 77-104.

Anselin, L. and Florax, R. J. G. M. (1995a). *New Directions in Spatial Econometrics: Introduction.* In: Anselin, L., Florax, R. J. G. M. (eds.). Advances in Spatial Science, pp. 3-16, Berlin.

Anselin, L. and Florax, R. J. G. M. (1995b). *Small sample properties of tests for spatial dependence in regression models: Some further results.* In: Anselin, L., Florax, R. J. G. M. (eds.). New Directions in Spatial Econometrics, pp. 21-74, Berlin.

Anselin, L. and Rey, S. (1991). *The performance of tests for spatial dependence in linear regression models.* In: Geographical Analysis, 23, pp. 110-131.

Anselin, L., Varga, A., and Acs, Z. (1997). *Local Geographic Spillovers between University Research and High-Technology Innovations.* In: Journal of Urban Economics, 42, pp. 422-448.

Anselin, L., Varga, A., and Acs, Z. J. (2000). *Geographical Spillovers and University Research: A Spatial Econometric Perspective.* In: Growth and Change, 31, pp. 501-515.

Arrow, K. J. (1962). *The Economic Implications of Learning by Doing.* In: Review of Economic Studies, 29, pp. 155-173.

Arthur, W. B. (1989). *Competing technologies, increasing returns, and lock-in by historical events.* In: The Economic Journal, 99, pp. 116-131.

Aschauer, D. (1989). *Aschauer, D., 1989. Is public expenditure productive?* In: Journal of Monetary Economics, 23, pp. 177-200.

Asheim, B. (1996). *Industrial districts as learning regions: a condition for prosperity?* In: European Planning Studies, 4, pp. 379-400.

Audretsch, D. and Mahmood, T. (1994). *The Knowledge Production Function and R&D Spillovers.* Wissenschaftszentrum Berlin für Sozialforschung. Discussion Paper FS IV 94-6.

Audretsch, D. B. (1995). *Innovation and Industry Evolution.* Cambridge.

Audretsch, D. B. (1998). *Agglomeration and the Location of Innovative Activity.* In: Oxford Review of Economic Policy, 14, pp. 18-29.

Audretsch, D. B. and Feldman, M. P. (1996). *Innovative Clusters and the Industry Life-cycle.* In: The Review of Industrial Organization, 11, pp. 253-273.

Backus, D. K., Kehoe, R. J., and Kehoe, T. J. (1992). *In Search of Scale Effects in Trade and Growth.* In: Journal of Economic Theory, 58, pp. 377-409.

Badinger, H. and Tondl, G. (2002). *Trade, Human Capital and Innovation: The Engines of European Regional Growth in the 1990s.* in: Fingleton, B. (Hrsg.), European Regional Growth, Heidelberg.

Baldwin, R. E. and Martin, P. (2003). *Agglomeration and Regional Growth.* CEPR Discussion Paper No. 3960.

Barro, R. J. and Sala-I-Martin, X. (1995). *Economic Growth.* New-York.

Barro, R. J. and Sala-I-Martin, X. (1997). *Technological diffusion, convergence, and growth.* In: Journal of Economic Growth, 2, pp. 126.

Barry, R. P. and Kelley, R. P. (1999). *Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices.* In: Linear Algebra and its Applications, 289, pp. 41-54.

Bass, F. M. (1969). *A new product growth for model consumer durables.* In: Management Science, 15, pp. 215-227.

Bass, F. M. (1993). *The Future of Research in Marketing: Marketing Science.* In: Journal of Marketing Research, 30, pp. 1-6.

Bass, F. M. (1995). *Empirical Generalizations and Marketing Science: A Personal View.* In: Marketing Science, 14, G6-G19.

Bayes, T. (1763). *An Essay towards solving a Problem in the Doctrine of Chances.* In: Philosophical Transactions of the Royal Society of London, 53, pp. 370-418.

Bellmann, L. and Brussig, M. (1998). *Ausmaß und Ursache der Produktivitätslücke von ost- und westdeutschen Betrieben. Eine Analyse auf der Grundlage des Betriebspanels.* In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 31, pp. 648-660.

Bertschek, I. and Entorf, H. (1996). *On Nonparametric Estimation of the Schumpeterian Link between Innovation and Firm Size: Evidence from Belgium, France and Germany.* In: Empirical Economics, 21, pp. 401-426.

Best, M. H. (2001). *The New Competitive Advantage: The Renewal of American Industry.* Oxford.

Blundell, R., Griffith, R., and Reeenen, J. V. (1995). *Dynamic Count Data Models of Technological Innovation.* In: Economic Journal, 105, pp. 333-344.

Bohl, M. T. (1998). *Konvergenz westdeutscher Regionen? Neue empirische Ergebnisse auf der Basis von Panle-Einheitswurzeltests.* In: Konjunkturpolitik, 44, pp. 82-99.

Boswijk, H. P. and Franses, P. H. (2005). *On the Econometrics of the Bass Diffusion Model.* In: Journal of Business & Economic Statistics, 23, pp. 255-268.

Boswijk, P., Fok, D., and Franses, P. H. (2006). *A New Multivariate Product Growth Model.* In: Tinbergen Institute Discussion Paper No. 06-027/4.

Bottazzi, L. and Peri, G. (2003). *Innovation and spillovers in regions: evidence from European patent data.* In: European Economic Review, 47, pp. 687-710.

Brakman, S., Garretsen, H., and Marrewijk, C. (2001). *An Introduction to Geographical Economics.* Cambridge.

Bröcker, J. (2002). *Schlussfolgerungen aus der Theorie endogenen Wachstums für eine ausgleichende Regionalpolitik*. In: Raumforschung und Raumordnung, 3-4, pp. 185-194.

Brendstetter, L. G. (2001). *Are knowledge spillovers are international of intranational in scope? Microeconometric evidence from the U.S. and Japan.* In: Journal of International Economics, 53, pp. 53-79.

Brezis, E. S. and Krugman, P. (1993). *Technology and the Life-Cycle of Cities.* NBER Working Paper, No. 4561.

Campagni, R. (1991). *Local milieu, uncertainty and innovation networks: Towards a new dynamic theory of economic space.* In: Campagni, R. (ed.). Innovation Networks: Spatial Perspectives, pp. 121-142. London.

Cantner, U. and Hanusch, H. (1998). *Industrie-Evolution.* Volkswirtschaftliche Disskusionsreihe 177, Universität Augsburg.

Chand, S. and Sethi, S. P. (1990). *A dynamic lot sizig model with learning in setups.* In: Operation Research, 38, pp. 644-655.

Chatterjee, R. and Eliashberg, J. (1990). *The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach.* In: Management Science, 36, pp. 1057-1079.

Christensen, C. (1997). *The Innovator´s Dilemma: When New Technologies Cause Great Firms to Fail.* Boston.

Clark, K. B. and Fujimoto, T. (1991). *Product Development Performance. Strategy, Organization, and Management in the World Auto Industry.* Boston.

Coe, D. T. and Helpman, E. (1995). *International R&D Spillovers.* In: European Economic Review, 39, pp. 859-887.

Cohen, W. and Klepper, S. (1996). *A Reprise of Size and R&D.* In: The Economic Journal, 106, pp. 925-951.

Cohen, W. and Levinthal, D. A. (1990). *Absorptive Capacity: A New Perspective on Learning and Innovation.* In: Adminsitrative Science Quarterley, 35, pp. 128-152.

Cohen, W. and Levinthal, R. (1989). *Empirical Studies of Innovation and Market Structure.* In: Schmalensee, R., Willig, R. Handbook of Industrial Organization, 2, pp. 1059-1107.

Comanor, W. S. (1967). *Market structure, product differentiation, and industrial research.* In: The Quarterly Journal of Economics 81, pp. 639-657.

Dasgupta, P. and Heal, G. (1979.). *Economic Theory and Exhaustible Resources.* Cambridge.

D'Aspremont, C. and Jacquemin, A. (1988). *Cooperative and noncooperative R&D in duopoly with spillovers*. In: American Economic Review, 78, pp. 1133-1137.

David, P. A. (1985). *The economics of QWERTY*. In: American Economic Review, 75, pp. 332-337.

Dinopoulos, E. and Thompson, P. . (1998). *Schumpeterian growth without scale effects*. In: Journal of Economic Growth, 3, pp. 313-335.

DIW, IAB, IfW, IWH, and ZEW (2002). *Fortschritte beim Aufbau Ost. Fortschrittsbericht wirtschaftswissenschafticher Forschungsinstitute über die wirtschaftliche Entwicklung in Ostdeutschland*. DIW, IAB, IfW, IWH, ZEW.

Dollar, D. (1986). *Technological innovation, capital mobility, and the product cycle in North-South trade*. In: The American Economic Review, 76, pp. 177-190.

Dosi, G. (1982). *Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change*. In: Research Policy, 2, pp. 147-182.

Dosi, G. (1984). *Technical Change and Industrial Transformation*. London.

Dosi, G. and Orsenigo, L. (1985). *Order and change: an exploration of markets, institutions and technology in industrial dynamics*. SPRU Discussion Paper 22, Brighton.

Dutton, J. M. and Thomas, A. (1984). *Treating progress functions as a managerial opportunity*. In: Academy of Management Review, 9, pp. 235-247.

Easingwood, C. J., Mahajan, V., and Muller, E. (1983). *A non-uniform influence innovation diffusion model of new product acceptance*. In: Marketing Science, 2, pp. 273-295.

Eckey, H. F., Kosfeld, R., and Türck, M. (2007). *Regionale Entwicklung mit und ohne räumliche Spillover-Effekte*. In: Jahrbuch für Regionalwissenschaft, 27, pp. 23-42.

Eicher, T. S. and Turnovsky, S. J. (1999). *Convergence in a Two-Sector Nonscale Growth Model*. In: Journal of Economic Growth, 4, pp. 413-428.

Entorf, H. (1988). *Die endogene Innovation. Eine mikro-empirische Analyse von Produktphasen als Innovationsindikatoren*. In: Jahrbücher für Nationalökonomie und Statistik, 204, pp. 175-189.

EU (2003). *The European R3L Learning Region Network Services*. European Union.

EU (2004). *Education and training 2010 diverse systems, shared goals - The education and training contribution to the Lisbon strategy*. European Union.

Fernandez, C., Ley, E., and Steel, M. (2001a). *Benchmark priors for Bayesian model averaging.* In: Journal of Econometrics, 100, pp. 381-427.

Fernandez, C., Ley, E., and Steel, M. (2001b). *Model uncertainty in cross-country growth regressions.* In: Journal of Applied Econometrics, 16, pp. 563-576.

Fisher, F. M. and Temin, P. (1973). *Returns to scale in research and development: what does the Schumpeterian hypothesis imply?* In: Journal of Political Economy, 81, pp. 56-70.

Fisher, R. A. (1930). *The genetical theory of natural selection.* Oxford.

Flaig, G. and Stadler, M. (1994). *Success breeds success. The dynamics of the innovation process.* In: Empirical Economics, 19, pp. 55-68.

Florax, R. J. G. M., Folmer, H., and Rey, S. J. (2003). *Specification searches in spatial econometrics: the relevance of Hendry's methodology.* In: Regional Science and Urban Economics, 33, pp. 557-579.

Florida, R. (1995). *Toward the learning region.* In: Futures, 27, pp. 527-536.

Florida, R. (2002). *The Economic Geography of Talent.* In: Annals of the Association of American Geographers, 92, pp. 743-755.

Fok, D. and Franses, P. H. (2007). *Modeling the diffusion os scientific publications.* In: Journal of Econometrics, 139, pp. 376-390.

Forni, M. and Paba, S. (2001). *Knowledge Spillovers and the Growth of Local Industries.* in: CEPR Discussion Papers No. 58, London.

Frauenhofer (2000). *Regionale Verteilung von Innovations- und Technologiepotentialen in Deutschland und Europa.* Fraunhofer-Institut für Systemtechnik und Innovationsforschung, Karlsruhe.

Frenkel, M. and Hemmer, R. (1999). *Grundlagen der Wachstumstheorie.* München.

Frey, P. W. and Sears, R. J. (1978). *Models of conditioning incorporating the Rescorla-Wagner association axiom, a dynamic attention process, and a catastrophe rule.* In: Psychological Review, 85, pp. 321-340.

Frisch, A. J. (1993). *Unternehmensgröße und Innovation.* Frankfurt am Main.

Fujita, M. and Thisse, J. (2002). *Economics of Agglomeration.* Cambridge.

Funke, M. and Niebuhr, A. (2000). *Spatial R&D Spillovers and Economic Growth Evidence from West Germany.* HWWA Discussion Paper Nr. 98, Hamburg.

Gamerman, D. and Migón, H. (1991). *Forecasting the Number of AIDS Cases in Brazil.* In: The Statistician, 40, pp. 427-442.

Gatignon, H. and Robertson, T. S. (1985). *A propostitional inventory for new diffusion research.* In: Journal of Consumer Research, 11, pp. 849-867.

Gelfand, A. E., Racine-Poon, A., and Smith, A. F. M. (1990). *Illustration of Bayesian inference in normal data models using Gibbs sampling.* In: Journal of the American Statistical Association, 85, pp. 972-985.

Gelman, A., Carlin, B., J., Stern, H. S., and Rubin, D. (1995). *Bayesian Data Analaysis.* London.

Geroski, P. and Mazzucato, M. (2002). *Learning and the sources of corporate growth.* In: Industrial and Corporate Change, 11, pp. 623-644.

Geweke, J. F. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.* In: Berger, J. O., Bernardo, J. M. , Dawid, A. P., Smith, A. F .M. (eds.). Bayesian Statistics, 4, pp. 169-194.

Geweke, J. F. (1993). *Bayesian treatment of the independent Student-t linear model.* In: Journal of Applied Econometrics, 8, pp. 19-40.

Gladwell, M. (2000). *The Tipping Point. How Little things Can Make a Big Difference.* Little Brown and Company, London.

Glass, A. (1997). *Product cycles and market penetration.* In: International Economic Review, 38, pp. 865-891.

Goldenberg, J., Libai, B., Muller, E., and Peres, R. (2006). *Blazing saddles: the early and mainstream markets in the high-tech product life cycle.* In: Israel Economic Review, 4, pp. 85-108.

Gordon, H. S. (1954). *The economic theory of a common-property.* In: Journal of Political Economy, 62, pp. 124-142.

Greif, S. (1998). *Patentatlas Deutschland.* München.

Griffith, D. A. (1996). *Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data.* In: Canadian Geographer, 40, pp. 351-367.

Griffith, D. A. (2000). *A Linear Regression Solution to the Spatial Autocorrelation Problem.* In: Journal of Geographical Systems, 2, pp. 141-156.

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering.* Berlin.

Griliches, Z. (1957). *Hybrid corn: An exploration in the economics of technological change.* In: Econometrica, 25, pp. 501-522.

Griliches, Z. (1979). *Issues in Assessing the Contribution of Research and Development to Productivity Growth.* In: The Bell Journal of Economics, 10, pp. 92-116.

Grossman, G. M. and Helpman, E. (1991a). *Innovation and Growth in the Global Economy.* Cambridge.

Grossman, G. M. and Helpman, E. (1991b). *Quality ladders and product cycles.* In: The Quarterly Journal of Economics, 106, pp. 557-586.

Guckenheimer, J. and Holmes, P. (1983). *Nonlinear Osciallations, Dynamical Systems, and Bifurcations of Vector Fields.* in: Springer, Vol. 21/4, pp. 25-45.

Hansen, M. T. (1999). *The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits.* In: Administrative Science Quarterly, 44, pp. 82-111.

Hardy, G. H. (1908). *Mendelian proportions in a mixed population.* In: Science, 28, pp. 49-50.

Hargadon, A. (1996). *Diffusion of innovations: The current model and directions for future research.* In: Dorf, R.C. (ed.). The Handbook of Technology Management, 3, pp. 20-27. Boca Raton.

Hayek, F. A. (1945). *The Use of Knowledge in Society.* In: American Economic Review, 35, pp. 519-530.

Hayes, R., K. C. (1986). *Why some factories are more productive than others.* In: Harvard Business Review, 64, pp. 66-73.

Helpman, E. and Krugman, P. R. (1985). *Market Structure and Foreign Trade.* Cambridge.

Henderson, R. and Cockburn, I. (1996). *Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery.* In: The RAND Journal of Economics, 27, pp. 32-59.

Hendry, D. F. (1979). *Predictive failure and econometric modelling in macroe- conomics: The transactions demand for money.* In: P. Ormerod (ed.). Economic Modelling. London.

Hendry, D. F. (2006). *A comment on specification searches in spatial econometrics: The relevance of Hendrys methodology.* In: Regional Science and Urban Economics, 36, pp. 309-312.

Hepple, L. W. (2004). *Bayesian Model Choice in spatial econometrics.* In: Lesage, J. P., Pace, R. K. (eds.). Advances of Econometrics, 18, pp. 101-126.

Hiero (2008). *Wirtschaftliche Zukunftsfelder in Ostdeutschland - Kurzfassung der Studie* -. Hanseatic Institute for Entrepreneurship and Regional Development an der Universität Rostock.

Hill, S., Provost, F., and Volinsky, C. (2006). *Network-Based Marketing: Identifying Likely Adopters via Consumer Networks.* In: Statistical Science, 21, pp. 256-276.

Hotelling, H. (1929). *Stability in competition.* In: The Economic Journal, 39, pp. 41-57.

Hull, C. L. (1943). *Principles of behavior, an introduction to behavior theory.* New York.

Hymer, S. and Pashigan, P. (1962). *Turnover of firms as a measure of market behavior.* In: Review of Economics and Statistics, 44, pp. 82-87.

Jacobs, J. (1970). *The Economy of Cities.* New York und Toronto.

Jaffe, A. B. and Trajtenberg, M. (1996). *Flows of knowledge from universities and federal labs: Modeling the flowof patent citations over time and across institutional and geographic boundaries.* In: NBER Working Papers, 5712.

Jaffe, A. B. and Trajtenberg, M. (1998). *International knowledge flows: Evidence from patent citations.* In: NBER Working Papers, 6507.

Jaffe, A. B., M. T. and Fogarty, M. S. (2000). *Knowledge spillovers and patent citations: Evidence from a survey of inventors.* In: American Economic Review, 90, pp. 215-218.

Jaffe, Adam B., T. M. and Henderson, R. (1993). *Geographic localisation of knowledge spillovers as evidenced by patent citations.* In: Quarterly Journal of Economics, 108, pp. 577-598.

Jeuland, A. (1981). *Parsimonious models of diffusion of innovation: Part a, derivations and comparison.* Working paper, Graduate School of Business, University of Chicago.

Jones, C. (1995a). *R&D Based Models of Economic Growth.* In: Journal of Political Economics, 103, pp. 759-784.

Jones, C. (1995b). *Time Series Tests of Endogenous Growth Models.* In: Quarterly Journal of Economics, 110, pp. 495-525.

Jones, C. (1998). *Population and Ideas: A Theory of Endogenous Growth.* Stanford University mimeo.

Kahnert, R. (1998). *Wirtschaftsentwicklung, Sub- und Desuburbanisierung.* In: Informationen zur Raumentwicklung, 24, pp. 509-520.

Kaldor, N. and Mirrlees, J. A. (1962). *A New Model of Economic Growth.* In: Review of Economic Studies, 29, pp. 25-43.

Kalish, S. (1985). *A new product adoption model with price, advertising and uncertainty.* In: Management Science, 31, pp. 1569-1585.

Kalish, S. and Sen, S. K. (1986). *Diffusion models and the marketing miex for single products.* In: Mahajan, V., Wind, Y. (eds.). Innovation Diffusion Models of New Product Acceptance, V, pp. 87-116. Cambridge.

Kaltenbrunner, R. (2003). *Scholle und Rand. Wohnen und Suburbanisierung-ein kaum steuerbarer Zusammenhang.* In: Raumordnung und Raumforschung, 61, pp. 319-333.

Kamien, M., Muller, E., and Zang, I. (1992). *Research joint ventures and R&D cartels.* In: American Economic Review, 82, pp. 1133-1137.

Kamien, M. I. and Schwartz, N. L. (1975). *Market Structure and Innovation: A Survey.* In: Journal of Economic Literature, 13, pp. 1-37.

Keilbach, M. (2000). *Spatial Knowledge Spillovers and the Dynamic of Agglomeration and Regional Growth.* Heidelberg.

Kelejian, H. H. and Prucha, I. R. (1998). *A generalized spatial two stage least square procedure for estimating a spatial autoregressive model with autoregressive disturbances.* In: Journal of Real Estate Finance and Economics, 17, pp. 99-121.

Kelejian, H. H. and Robinson, d. P. (2004). *The influence of spatially correlated heteroskedasticity on tests for spatial correlation.* In: Anselin, L., Florax, R. J. G. M., Rey, S. J. (eds.). Advances in Spatial Econometrics, pp. 79-98. Berlin.

Keller, W. (2001). *Geographic Localization and International Technology Diffusion.* In: American Economic Review, 92, pp. 120-142.

Kühn, M. (2001). *Regionalisierung der Städte. Eine Analyse von Stadt-Umland-Diskursen räumlicher Forschung und Planung.* In: Raumordnung und Raumforschung, 59, pp. 402-411.

Kim, C. W., Phipps, T. T., and Anselin, L. (2003). *Measure the benefits of air quality improvement: A spatial hedonic approach.* In: Journal of Enviornmental Economics and Management, 45, pp. 24-39.

Klein, B. H. (1977). *Dynamic Economics.* Camebridge, M. A.

Klepper, S. (1996). *Entry, exit, growth and innovation in the product life cycle.* In: American Economic Review, 86, pp. 562-583.

Klodt, H. (2000). *Industrial Policy and the East German Productivity Puzzle.* In: German Economic Review, 1, pp. 315-333.

Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations.* Applications of Mathematics 23, Berlin.

Klotz, S. (1996). *Ökonometrische Modelle mit raumstruktureller Autokorrelation: eine kurze Einführung*. Mimeo: Univeristät Konstanz. Can be downloaded from www.uni-konstanz.de/FuF/wiwi/pohlmei/www.html.

Kluyver, C. A. (1977). *Innovation and industrial product life cycles*. In: Californian Management Review, 20, pp. 21-33.

Kmenta, J. (1971). *Elements of Econometrics*. New York: Macmillan.

Koop, G. (2003). *Bayesian Econometrics*. Wiley.

Kortum, S. . (1997). *Research, patenting, and technological change*. In: Econometrica, 65, pp. 1389-1419.

Kotschatzky, K., Lo, V., and Stahlecker, T. (2006). *Innovationsbedinungen und Innovationspotenziale in Ostdeutschland - Exemplarische Analyse von drei Grenzregionen*. Frauenhofer-Insitut für System- und Innovationsforschung ISI.

Kraft, K. (1989). *Market Structure, Firm Characteristics and Innovative Activity*. In: The Journal of Industrial Economics, 37, pp. 329-336.

Kremer, M. (1993). *Population growth and technological change: One million B.C. to 1990*. In: Quarterly Journal of Economics, 108, pp. 681-716.

Krugman, P. (1979). *A model of innovation, technology transfer, and the world distribution of income*. In: Journal of Political Economy, 87, pp. 253-266.

Krugman, P. (1991). *Increasing Returns and Economic Geography*. In: Journal of Political Economy, 99, pp. 483-499.

Krugman, P. (1995). *Development, Geography and Economic Theory*. Cambridge.

Krugman, P. (1998a). *Space: The Final Frontier*. In: Journal of Economic Perspectives, 12, pp. 161-174.

Krugman, P. (1998b). *Whats New About New Economic Geography?* In: Oxford Review of Economic Policy, 14, pp. 7-17.

Kumar, N. and Saqib, M. (1996). *Firm size, opportunities for adaption and inhouse R&D activity in developing countries: the case of Indian manufacturing*. In: Research Policy, 25, pp. 713-722.

Kwasnicki, W. (1996). *Knowledge, innovation adn economy: an evolutionary exploration*. Aldershot.

Kwasnicki, W. and Kwasnicka, H. (1992). *Market, innovation, competition: an evolutionary model of industrial dynamics*. In: Journal of Economic Behaviour and Dynamics, 19, pp. 343-368.

Kydland, F. E. and Prescott, E. C. (1982). *Time to build and aggregate fluctuations*. In: Econometrica, 50, pp. 1345-1370.

Lechevalier, S., Ikeda, Y., and Nishimura, J. (2007). *Investigating Collaborat ive R&D Using Patent Data: The Case Study of Robot Technology in Japan*. Discussion Paper Series No. 498, The Institute of Economic Research, Hitot subashi University.

LeSage, J. P. (2000). *Bayesian estimation of limited dependent vaiable spatial autoregressive models*. In: Geographical Analysis 32, pp. 19-35.

LeSage, J. P. and Parent, O. (2007). *Bayesian Model Averaging for Spatial Econometric Models*. In: Geographical Analysis, 39, pp. 241-267.

Levin, R., Cohen, W., and Mowery, D. (1985). *R&D appropriability, opportunity, and market structure: new evidence on some Schumpeterian hypotheses*. In: American Economic Review Proceedings, 75, pp. 20-24.

Li, M. Y. (1996). *Dulac Criteria for Autonomous Systems Having an Invariant Affine Manifold*. in: Journal of Mathematical Analysis and Applications, 199, pp. 374-390.

Lichtenberg, F. and van Pottelsberghe de la Potterie, B. (1996). *International R&D Spillovers: A Re-Examination*. NBER Working Paper No. 5668.

Lucas, R. E. (1976). *Econometric policy evaluation: A Critique*. in: Karl Brunner and Allan Meltzer (Hrsg.), The Phillips Curve and Labor Markets. Carnegie Rochester Conference Series 1, pp. 19-46.

Lucas, R. E. (1988). *On the Mechanics of Economic Development*. In: Journal of Monetary Economics, 22, pp. 3-39.

Machlup, F. (1980). *Knowledge, its Creation, Distribution and Economic Significance*. Princeton.

Madigan, D. and York, J. (1995). *Bayesian Graphical Models for Discrete Data*. In: International Statistical Reveiws, 63, pp. 215-232.

Mahajan, V., E., M., and Kerin, R. A. (1984). *Introduction Strategy for New Products with Positive and Negative Word-of-Mouth*. In: Management Science 30, pp. 1389-1404.

Mahajan, V., Muller, E., and Bass, F. M. (1990). *New Product Diffusion Models in Marketing: A Review and Directions for Research*. In: Journal of Marketing, 54, pp. 1-26.

Mahajan, V., Muller, E., and Bass, F. M. (1993). *New Product Diffusion Models*. In: Eliashberg, J., Lilien, G.L. (eds.). Handbooks in Operations Research & Management Science, V, pp. 349-408.

Mahajan, V., Muller, E., and Wind, Y. (2000). *New-Product Diffusion Models*. Norwell.

Mahajan, V. and Peterson, R. A. (1985). *Models for Innovation Diffusion*. Beverly Hills.

Mahajan, V. and Wind, Y. (1986). *Innovation Diffusion Models of New Product Acceptance*. Cambridge.

Malerba, F. and Orsenigo, L. (1993). *Technological regimes and firm behavior*. In: Industrial and Corporate Change 2, pp. 45-72.

Malerba, F., Orsenigo, L., and Peretto, P. (1997). *Persistence of innovative activities, sectoral patterns of innovation and international technological spillovers*. In: International Journal of Industrial Organization, 15, pp. 801-826.

Mansfield, E. (1968). *Industrial Reseach and Technological Innovation-An econometric analysis*. New-York.

Markides, C. (1998). *Strategic innovation in established companies*. In: Sloan Management Review, Spring, pp. 31-42.

Marshall, A. (1920). *Principles of Economics*. 8th edition. London.

Maskell, P. and Malmberg, A. (1999). *Localised learning and industrial competitiveness*. In: Cambridge Journal of Economics, 23, pp. 167-185.

Maurseth, P. B. and Verspagen, B. (2002). *Knowledge spillovers in Europe. A patent citations analysis*. In: Scandinavian Journal of Economics, 104, pp. 531-545.

Mazzucato, M. (2000). *Firm Size, Innovation and Market Structure*. Cheltenham.

Mazzucato, M., S. W. (1998). *Market Share Instability and Stock Price Volatility during the Industry Life - Cycle: The U.S. Automobile Industry*. In: Journal of Evolutionary Economics, 9, pp. 67-96.

McEvily, S. K. and Chakravarthy, B. (2002). *The persistence of knowledge-based advantage: an empirical test for product performance and technological knowledge*. In: Strategic Management Journal, 23, pp. 285 - 305.

Mchich, R., Auger, P., and Poggiale, J.-C. (2007). *Effect of predator density dependent dispersal of prey on stability of a predator-prey system*. In: Mathematical Biosciences 206, pp. 343-356.

Meade, N. (1988). *A Modified Logistic Model Applied to Human Populations.* In: Journal of Royal Statistical Society, 151, pp. 491-498.

Meade, N. and Islam, T. (1995). *Forecasting with Growth Curves: An Empirical Comparison.* In: International Journal of Forecasting, 11, pp. 199-215.

Metcalfe, J. (2002). *Is study time allocated selectively to a region of proximal learning?* In: Journal of Experimental Psychology. General, 132, pp. 530-542.

Metcalfe, S. (1994). *Competition, Fisher's Principle and increasing returns in the selection process.* In: Journal of Evolutionary Economics, 4, pp. 321-346.

Moll, P. (2000). *Probleme und Ansätze zur Raumentwicklung in der europäischen Grenzregion Saarland-Lothringen-Luxemburg-Rheinland-Pfalz-Wallonien.* In: Raumordnung und Raumforschung, 58, pp. 265-275.

Moore, G. (1995). *Inside the Tornado. HarperBusiness.* New-York.

Moore, G. (2002). *Crossing the Chasm.* New-York.

Moran, P. A. P. (1948). *The interpretation of Statistical Maps.* In: Journal of the Royal Statistical Society, 10, pp. 243-251.

Moran, P. A. P. (1950). *Notes on Continuous Stochastic Phenomena.* In: Biometrika, 37, pp. 17-33.

Morgan, K. (1997). *The learning region: institutions, innovation and regional renewal.* In: Regional Studies, 31, pp. 491-503.

Mulligan, C. B. and Sala-I-Martin, X. (1993). *Transitional Dynamics in Two-Sector Models of Endogenous Growth.* In: Quarterly Journal of Economics, 108, pp. 739-773.

Mur, J. (1999). *Testing for spatial autocorrelation: Moving average versus autoregressive processes.* In: Enviornment and Planning, 31, pp. 1371-1382.

Myrdal, G. (1959). *Ökonomische Theorie und unterentwickelte Regionen.* Stuttgart.

Nelson, T. O. and Narens, L. (1990). *Metamemory: A theoretical framework and new findings.* In: Bower, G. H. (ed.). The psychology of learning and motivation, 26, pp. 125-141, New-York.

Nelson, T. O. and Narens, L. (1994). *Why investigate metacognition?* In: Metcalfe, J., Shimamura, A. P. (eds.). Metacognition: Knowing about knowing, pp. 1-25. Cambridge.

Neumann, M., Böbel, I., and Haid, A. (1982). *Innovations and Market Structure in West German Industries.* In: Managerial and Decision Economics, 3, pp. 131-139.

Newell, K. M., Liu, Y. T., and Mayer-Kress, G. (2001). *Time scales in motor learning and development.* In: Psychological Review, 108, pp. 57-82.

Nickell, S. J. (1996). *Competition and Corporate Performance.* In: Journal of Political Economy, 104, pp. 724-46.

Nickell, S. J., Nicolitasas, D., and Dryden, N. (1997). *What makes firms perform well?* In: European Economic Review, 41, pp. 783-796.

Noailly, J. (2008). *Coevolution of economic and ecological systems. An application to agricultural pesticide resistance.* In: Journal of Evolutionary Economics, 18, pp. 1-30.

Noailly, J., van den Bergh, J. C. J. M., and Withagen, C. A. (2003). *Evolution of harvesting strategies: replicator and resource dynamics.* In: Journal of Evolutionary Economics, 13, pp. 183-200.

Nooteboom, B. (1992). *Towards a Dynamic Theory of Transactions.* In: Journal of Evolutionary Economics, 2, pp. 281-299.

Nooteboom, B. (1999). *Inter-firm Alliances: Analysis and Design.* London.

Olejnik, A. (2008). *Using the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU25.* In: Regional Science 87, pp. 371-384.

Ord, K. (1975). *Estimation Methods for Models of Spatial Interaction.* In: Journal of American Statistical Association, 70, pp. 120-126.

Oren, S. S. and Schwartz, R. G. (1988). *Diffusion of new products in risk-sensitive markets.* In: Journal of Forecasting, 7, pp. 273-287.

Parker, P. M. (1994). *Aggregate Diffusion Models in Marketing: A Critical Review.* In: International Journal of Forecasting, 10, pp. 353-380.

Pasternack, P. (2007). *Forschungslandkarte Ostdeutschland.* Sonderband "die hochschule" . Wittenberg.

Peretto, P. (1998). *Technological change and population growth.* In: Journal of Economic Growth, 3, pp. 283-311.

Phillips, A. (1971). *Technology and Market Structure: A Study of the Aircraft Industry.* Lexington.

Pisano, G., Bohmer, R., and Edmondson, A. (2001). *Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery.* In: Management Science, 47, pp. 752768.

Polany, M. (1967). *The Tacit Dimension.* New-York.

Porter, M. E. (2000). *Locations, clusters and company strategy.* In: F. M. P. Clark, G. L., Gertler, M. S. (eds.). The Oxford Handbook of Economic Geography. Oxford.

Raferty, A., Madigan, D., and Hoeting, J. (1997). *Bayesian Model Averaging for Linear Regression Models.* In: Journal of of the American Statistical Association, 92, pp. 179-191.

Raftery, A. E. and Lewis, S. (1992a). *Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo.* In: Statistical Science, 7, pp. 493-497.

Raftery, A. E. and Lewis, S. (1992b). *How many iterations in the Gibbs sampler?* In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. (eds.). Bayesian Statistics, 4, pp. 763-773.

Raftery, A. E. and Lewis, S. (1995). *The number of iterations, convergence diagnostics and generic Metropolis algorithms.* In: Gilks, W. R., Spiegelhalter, D. J., Richardson, S. (eds.). Practical Markov Chain Monte Carlo. London.

Rice, M. L., Wexler, K., and Hershberger, S. (1998). *Tense over time: The longitudiual course of tense aquisition in children with specific language impairment.* In: Journal of Speech, Language, and Hearing Research, 41, pp. 1412-1431.

Rink, D. and Swan, J. (1979). *Product life cycle research: a literature review.* In: Journal of Business Research, 7, pp. 219-42.

Rivera-Batiz, L. A. and Romer, P. M. (1991). *International trade with endogenous technological change.* In: Economic Review, 35, pp. 971-1004.

Rogers, E. M. (1983). *Diffusion of Innovations.* 3rd edt. New-York.

Romer, P. (1986). *Increasing Returns and Long-Run Growth.* In: Journal of Political Economy, 94, pp. 1002-1037.

Romer, P. (1990). *Endogenous Technical Change.* In: Journal of Political Economy, 98, pp. 71-102.

Rosen, E. (2000). *The Anatomy of Buzz.* New York.

Rosenberg, N. (1982). *Inside the Black Box: Technology and Economics.* Cambridge.

Sachverständigenrat (2005). *Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2005) Jahresgutachten 2005/06. Die Chance nutzen Reformen mutig voranbringen.* Wiesbaden.

Santolini, R. (2008). *A spatial cross-sectional analysis of political trends in Italian municipalities.* In: Regional Science, 83, pp. 431-451.

Saviotti, P. and Mani, G. (1995). *Competition, variety and technological evolution: A replicator dynamics model.* In: Journal of Evolutionary Economics, 5, pp. 369-392.

Scherer, F. (1967). *Market Structure And The Employment Of Scientists And Engineers.* In: American Economic Review, 57, pp. 524-31.

Schmalen, H. (1982). *Optimal prica and advertising policy for new products.* In: Journal of Business Research, 10, pp. 17-30.

Schmalwasser, O. and Schidlowski, M. (2006). *Kapitalstockrechnung in Deutschland.* In: Wirtschaft und Statistik, 11, pp. 1107-1123.

Schmookler, J. (1972). *Patents, Invention, and Economic Change.* In: Griliches, Z., Hurwicz, L. (eds.). Data and selected essays, Cambridge.

Schumpeter, J. A. (1912). *Theorie der wirtschaftlichen Entwicklung.* Berlin.

Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy.* New York.

Scotchmer, S. and Thisse, J.-F. (1992). *Space and Competition: A Puzzle.* In: Annals of Regional Science, 26, pp. 269-286.

Segerstrom, P. (1998). *Endogenous Growth Without Scale Effects.* In: American Economic Review, 88, pp. 1290-1310.

Seidel, B., V. D. (2000). *Anlagevermögen der ostdeutschen Länder und Gemeinden - noch erheblicher Nachholbedarf.* In: Wochenbericht DIW, 24, pp. 365-373.

Sinn, H.-W. (2000). *Germanys Economic Unification-An Assessment After Ten Years.* CESifo, Working Paper Series, 247, Munich.

Slywotzky, A. J. and Shaprio, B. P. (1993). *Leveraging to beat the odds: The new marketing mind-set.* In: Harvard Business Review, 71, pp. 97-107.

Smolny, U. (2003). *Smolny, U. (2003), Produktivitätsanpassung in Ostdeutschland. Bestandsaufnahme und Ansatzpunkte einer Erklärung.* In: Jahrbücher für Nationalökonomie und Statistik, 223, pp. 239-254.

Sorenson, O., Rivkin, J. W., and Flemming, L. (2005). *Informational Complexity and the Flow of Knowledge across social boundaries.* In: Discussion Papers in Evolutionary Economic Geography (PEEG), 0511, Utrecht University.

Spence, A. M. (1981). *The Learning Curve and Competition.* In: Bell Journal of Economics, 12, pp. 49-70.

Strogatz, S. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering.* Reading.

Sultan, F., Farley, J. U., and Lehmann, D. R. (1990). *A meta-analysis of applications of diffusion models.* In: Journal of Marketing Research, 27, pp. 70-77.

Tanny, S. M. and Derzko, N. A. (1988). *Innovators and imitators in innovation diffusion modelling*. In: Journal of Forecasting, 7, pp. 225-234.

Tappeiner, G., Hauser, C., and Walde, J. (2008). *Regional knowledge spillover: Fact or artifact?* In: Research Policy, 37, pp. 861-874.

Teece, D. J. (1998). *Research directions for knowledge management*. In: California Management Review, 40, pp. 289-292.

Tellis, G. J. and Crawford, C. M. (1981). *An evolutionary approach to product growth theory*. In: Journal of Marketing, 45, pp. 125-134.

Theil, H. and Goldberger, A. S. (1961). *On pure and mixed statistical estimation in economics*. In: International Economic Review, 2, pp. 65-78.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge.

Tirole, J. (1995). *Industrieökonomik*. München.

Todo, Y. (2001). *Growth, Population, and Knowledge Diffusion*. In: Knowledge, Technology, and Policy, 13, pp. 94-113.

Trivezg, F. J., M. J. (2004). *Some proposals for discriminating between spatial process*. In: Getis A., Mur J. and Zoller H. G. (eds). Spatial Econometrics and Spatial Statistics, pp. 150-175.

Tushman, M. and Nadler, D. (1986). *Organzing for Innovation*. In: California Management Review, 28, pp. 74-92.

Uzawa, H. (1965). *Optimum Technical Change in an Aggregative Model of Economic Growth*. In: International Economic Review, 6, pp. 18-31.

Van den Bulte, C. and Joshi, Y. V. (2007). *New Product Diffusion with Influentials and Imitators*. In: Marketing Science, 26, pp. 400-421.

Van den Bulte, C. and Lilien, G. L. (2001). *Medical Innovation Revisited: Social Contagion Versus Marketing Effort*. In: American Journal of Sociology, 106, pp. 1409-1435.

Van Gert, P. (1991). *A dynamic system model of cognitive and language growth*. In: Psychlological Review, 98, pp. 3-53.

Varga, A. (1998). *Local Academic Knowledge Spillovers and the Concentration of Economic Activity*. in: Regional Research Institute, West Virginia University, Research Paper No. 9803, Morgantown.

Vernon, J. M., G. P. (1974). *Technical Change and Firm Size: The Pharmaceutical Industry*. In: Review of Economics and Statistics, 56, pp. 294-302.

Verspagen, B. (1992a). *Localized technological change, factor substitution and the productivity slowdown.* In: Freeman C., Soete, L (eds). New Explorations in the Economics of Technological Change, pp. 193-221, London.

Verspagen, B. (1992b). *Uneven Growth between Interdependent Economics.* Maastricht.

Wakasugi, R. and Koyata, F. (1997). *Are Japanese Firm Efficient in Product Development.* In: Journal of Product Innovation Management, 14, pp. 383-392.

Wolfram, S. (1994). *Cellular Automata and Complexity.* Reading.

Wright, T. (1936). *Factors Affecting the Cost of Airplanes.* In: Journal of Aeronautical Science, 4, pp. 122-128.

Yelle, L. E. (1979). *The Learning Curve: Historical Review and Comprehensive Survey.* In: Decision Sciences, 10, pp. 302-328.

Yildizoglu, M. (2002). *Competing R&D Strategies in an Evolutionary Industry Model.* In: Computational Economics, 19, pp. 51-65.

Young, A. (1998). *Growth without scale effects.* In: Journal of Political Economy, 106, pp. 41-63.

Young, P. and Ord, J. (1989). *Model Selection and Estimation for Technological Growth Curves.* In: International Journal of Forecasting, 5, pp. 501-513.