Spatial Decisions of Multinational Enterprises and their Effect on Local Firms

by

Dirk Ehnts

Stuttgart-Hohenheim
ISSN 1618-5358
Spatial Decisions of Multinational Enterprises and their Effect on Local Firms

Dirk H. Ehnts*
Department of Economics
University of Oldenburg, Germany
August 2004

Abstract

Since the 1980s empirical research has been conducted on the influence of MNEs on local firms. The spillovers predicted by growth theory models used in the research designs have not been found. The main result is the importance of increased competition for the productivity of local firms. When FDI flows expanded rapidly in the 1990s, it became clear that MNEs play an important role in international technology transfer. However, growth theory models are limited in that market structures and firms cannot be modeled explicitly. The New Economic Geography (NEG) is better equipped to handle these issues. Instead of spillovers it relies on linkages. Therefore, new insights might be gained by basing empirical research on a NEG model and looking for linkages among firms.

*Uhlhornweg A5-0-009, 26129 Oldenburg, Germany, dirk@ehnts.de
1 Introduction

Since the days of Adam Smith economists have sought to explain economic growth and all the questions that come with it: Why are some regions growing faster than others? Why are regions sometimes growing fast, then not growing at all (or even shrinking)? Why do regions at one point ‘outgrow’ their neighbors, later they lag behind only to catch up again?\(^1\)

Growth and underdevelopment has been a big issue before. After World War II economists tackled the problems of what became known as development economics. Rosenstein-Rodan (1943) called for a government-initiated ‘Big Push’ to get countries growing. Hirschman postulated his ideas about unbalanced growth. Capital should be invested in leading sectors, he reckoned, so that growth in these sectors would be beneficial for other sectors of the economy due to input/output relations of firms - ’linkages’. Hirschman was thereby the first to employ the term ’linkage’ in this context.\(^2\) Foreign capital was necessary to achieve this growth, but the resulting higher growth path would be sufficient to pay off the initial debt. Chenery (1960) tried to find patterns of growth while Rostow (1960) described the process of industrialization by countries going through ’stages of growth’. These ideas were not formalized into mathematic models but nevertheless (or maybe therefore?) greatly influenced people working in governments and international organizations. When Latin American nations followed these ideas by borrowing heavily, aided by international organizations like the IMF and the World Bank, they finally achieved the ’Big Push’, but in the wrong direction: indebted and growth less some countries of Latin America tumbled through ’the lost decade’ of the 1980s.

The neo-classical growth theory meanwhile started as a tool to study macro-economic stability in a Keynesian context.\(^3\) It became obvious that the outcome of the Harrod-Domar model (Harrod, 1939 and Domar, 1946) would be a constant rate of growth, which only coincidentally happened to be the growth rate necessary to supply a growing work force with jobs. Solow (1956) predicted that the growth rates of developed and developing countries would tend to converge, given that the available technology would be the same. In 1986 Romer revived the subject with his seminal paper ’Endogenous Technical Change’. He states that not only physical capital can be accumulated, but also other types of reproducible capital, like knowledge. Knowledge can be understood as the ’product’ of a research sector, whose production function shows increasing returns to scale in respect to knowledge. Romer also introduces externalities as knowledge spillovers.

The predictions of the (endogenous) growth theory were the subject of

---

1 See Brakman et al. (2001)
2 The Idea of firms gaining in competitiveness by forming a cluster was first brought up by Weber (1909)
3 van de Klundert and Smulders (2001)
an empirical literature. Blomström and others tried to find the predicted knowledge spillovers. The search was mainly taking place in developing countries such as Mexico, Indonesia and Morocco. Data was adjusted to a fit an AK model. Researchers hoped to find that productivity growth would be significantly higher in industries with a large foreign share, compared to those dominated by national firms. The results were disappointing. Spillovers were not found.

Finally, the New Economic Geography deals (indirectly) with growth. Merging core-periphery spatial patterns with Dixit-Stiglitz (1977) competition and Samuelson’s iceberg costs (1954) proved to be a powerful combination (Krugman, 1991). While this work was written as an extension of the new trade theory, the model can give answers for most of the questions about growth posed at the beginning of this paper.

In the following I will argue that the growth theory is not well suited to model MNEs and their effect on local firms. The main reason is that neither firms nor market structure (and with it the degree of competition) can be explicitly modeled by growth theory. The role of knowledge spillovers as the vehicle of technology transfer differs from what is offered by recent research. NEG models build on local linkages among firms. The different concepts of growth theory and NEG are likely to influence the results of empirical research. While the search for spillovers using a growth theory setup has failed to turn up supporting evidence, it might be useful to try a NEG model instead and look for linkages.

The next section deals with links between FDI and growth via technology transfer. Spillovers and linkages are characterized and categorized. This is followed by a description of the part FDI plays in growth theory. In Section 4 follows an assessment of the empirical research regarding growth through FDI. Section 5 deals with the New Economic Geography more comprehensively so that in section 6 this theory can be confronted with spillover concepts and the empirical results from section 4. Section 7 concludes this paper.

2 FDI - Engine of Growth

The relationship of inward FDI and Growth is a close one. As can be seen in Graph 1, there is a correlation between the two. It seems that whenever growth of the world economy takes a dive, FDI contracts in the following period. Take 1982 or 1991, for instance. Growth slows down to a multi-year-low, with FDI flows shrinking. What does that say about the relationship between growth and FDI? It seems that growth and FDI go
together, with FDI depending on growth and lagging behind a bit. One could follow Schumpeter and argue that a growing economy is producing new/better products and technologies. These can be built elsewhere to make a good profit. One way to reap these profits would be FDI (the other important one being trade).

2.1 Overall effects of FDI

While there is no doubt that 'traditional' direct effects of investment will occur - like rising employment, rising output, rising taxes - doubt has been cast over the net effects. Some argue that the arrival of MNEs harms domestic firms by stealing their market shares and destroying their profitability. So after accounting for these effects, the welfare effects on the host country would remain unclear. The question at hand is: What is the effect of FDI on local firms in the host economy?

![Fig. 1: Growth rates of world FDI flows and GDP, 1980-2002 (per cent)](source: UNCTAD, FDI/TNC database and data from IMF, World Economic Outlook, 2003.

Very similar to this question has been the discussion about gains from trade. Trade and FDI seem to be complementary. Two types of FDI can be identified. Horizontal FDI are investments in firms of the same industry. They form the biggest part of Mergers and Acquisitions. Vertical FDI are investments in vertically linked industries. That may be a greenfield investment or a joint venture. Often the goal is to enhance the supply chain. This inevitably leads to fragmentation of the global economy.

Whatever form FDI takes, it is quite likely that foreign products (which embody foreign technology) are sold on the market of the host economy, hurting local competitors in the process. The profits gained by the MNE

---

6See Schumpeter (1942), Part II, Chapter VII: The Process of Creative Destruction
are often expatriated, to some extent. Trade theory tells us that there are winners and losers in trade. Consumers get new/improved products and therefore are better off, while some firms face tougher competition in their respective markets, seeing profits decline. Still, countries are supposed to gain from trade because of higher overall efficiency in their economies. If there is no big difference between trade and FDI, why then is the discussion of gains from FDI so different from the one we had about trade?

While with trade firms are spatially separated, multinationals arriving in host economies are neighbors with local firms. Therefore, they are part of an agglomeration, given that they are not isolated from the rest of the economy.  

2.2 Linkages and Spillovers

According to the growth theory international knowledge spillovers are a result of international economic integration. Therefore, arriving MNEs will boost the transmission of spillovers to local firms. In other words: The host economy will benefit from inward FDI. Empirical researchers might want to single out spillover channels. Hopefully, data would be available for one or more of the suspected spillover channels.

<table>
<thead>
<tr>
<th>Channel</th>
<th>How does it work?</th>
</tr>
</thead>
<tbody>
<tr>
<td>labor exchange</td>
<td>workers move from MNE to local firms</td>
</tr>
<tr>
<td>learning effect</td>
<td>local firms learn from MNEs (linkage exists)</td>
</tr>
<tr>
<td>demonstration effect</td>
<td>local firms copy MNEs processes</td>
</tr>
<tr>
<td>competition effect</td>
<td>outcome depends on market structure</td>
</tr>
</tbody>
</table>

Table 1: Spillover Channels between local firms and MNE

As it turns out, none of the spillover channels seem to be very measurable. When knowledge is transferred, there is not much ‘noise’ in the real world. Labor exchange, learning and demonstration effects also might take time to happen. A certain time lag would then blur the results, causing a delay between cause and effect. Only the competition effect would be instant, but nonetheless difficult to measure.

The concept of spillovers was developed by Marshall (1890) who sought to explain the concentration of industries in particular locations. He found that an agglomeration of firms is more attractive for workers, as the higher supply of jobs raises the probability of finding a job. In the literature this

---

7Like being located in an export processing zone
8I do not differentiate between ‘Knowledge’ and ‘Technological’ Spillovers, because there is hardly any difference. Take the learning effect: the local firm could gain knowledge or technology. The effect in terms of increased productivity will probably be the same.
9Book IV, Chapter 10, ‘Industrial Organization Continued. The Concentration of Specialized Industries in Particular Localities’
phenomenon is known as labor pooling. Another spillover is the spread of ideas. If a firm invents a new product, neighboring firms might feel inclined develop similar products. In the following table the effects that Marshall described are covered:

<table>
<thead>
<tr>
<th>Spillover</th>
<th>effect for firms</th>
<th>EGT</th>
<th>NEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor Pooling</td>
<td>positive</td>
<td>not featured</td>
<td>not featured</td>
</tr>
<tr>
<td>Price Index</td>
<td>positive</td>
<td>not featured</td>
<td>featured</td>
</tr>
<tr>
<td>Demonstration</td>
<td>unclear</td>
<td>featured</td>
<td>not featured</td>
</tr>
<tr>
<td>Learning</td>
<td>unclear</td>
<td>featured</td>
<td>not featured</td>
</tr>
<tr>
<td>Congestion</td>
<td>negative</td>
<td>not featured</td>
<td>featured</td>
</tr>
<tr>
<td>Competition</td>
<td>unclear</td>
<td>not featured</td>
<td>featured</td>
</tr>
</tbody>
</table>

Table 2: Types of spillovers

The competition effect according to Marshall was a supply-side kind effect: Industries fought for inputs. When inputs were in strong demand, their prices rose. Today the competition effect is understood as the demand-side competition of firms for customers. It depends on market structure - in a competitive environment the entry of a MNE is more likely to increase productivity then in an oligopolistic situation. Marshall’s competition effect is embodied in today’s idea of 'congestion’. Copying or imitating a MNE is causing the demonstration and learning effects. Local firms are likely to benefit while the MNE might give away part of its tacit knowledge. The price index effect consists of sinking input prices as more and more firms are producing at the same location. The argument rests on sinking transport costs and exploiting economies of scale.

Spillovers are a part of the theory of public goods, where they are called externalities. Main criteria for spillovers are non-excludability and non-rivalrous consumption. This is where the differences of linkages and spillovers become apparent: while spillovers are not influenced by the emitter, linkages absolutely are. A linkage arises only if an arriving firm interacts with a local firm. Contracting a local supplier, for instance, represents a linkage. Therefore, constructing a linkage demands action on both sides, MNE and local firm. Linkages consist of contracts, which indicates that transaction costs might play an important role.

There are two types of linkages: forward linkages and backward linkages. A forward linkage is a supply-related effect: The production in sector Y might induce firms in sector Z, which use Y as an input, to expand production. An arriving multinational setting up shop in the Y sector will produce

---

10 Results taken from Ehnits (2002), EGT = Endogenous Growth Theory, NEG = New Economic Geography

11 According to Wang and Blomström (1992) spillovers can be suppressed by firms, and receiving spillovers can induce learning costs on the local firm, too.
either a new variety of good Y or an existing variety, which tends to lower its price. Firms that use sector Y goods as inputs will now find themselves better off: either they pay a lower price for their inputs or they have the chance to incorporate a new variety. This improves their utility if consumers value variety in inputs. The backward-linkage is demand-related, as the addition of a firm of sector Y raises demand of their inputs, say goods of sector X. This might increase output per firm or promote the entry of new firms. Hirschman thought that by increasing demand, economies of scale could be exploited. These ideas were taken up again by the NEG, which I will discuss later on.

A backward linkage might feature a MNE helping its suppliers to raise their productivity or product quality. Lipsey (2002) found that MNEs show domestic firms how to fit into worldwide production networks. While the intended transfer of knowledge of the MNE to a local firm would be a linkage, an unintended transfer would be a spillover. It is important to be clear about the concepts of spillovers and linkages. The empirical literature discussed below is built around the concept of (knowledge or technological) spillovers, while the New Economic Geography emphasizes the role of linkages. When empirical results and theoretical concepts are compared further on, the distinction between spillovers and linkages and its application will be a topic again. Advancing chronologically, I will start with a description of the empirical literature and the theories in the background.

3 FDI in a Growth Theory framework

The most relevant theoretical model, Romer (1986), is the endogenous growth model. Romer used Arrow’s (1962) AK model as a starting point. The AK model in its original form is quite simple. The new feature of the model was that it did not yield diminishing returns. In its simplest form, that would be

\[ Y(t) = AK(t), \]  

(1)

with \( Y \) being output, \( A \) representing the level of technology, \( K \) human capital, and \( t \) for time. Growth in the model is now endogenous because even as technology stagnates (\( \dot{A} > 0 \)) the accumulation of human capital leads to a positive growth rate. To explicitly model knowledge-driven growth, it is necessary to modify the production function so that we get

\[ Y(t) = F(K(t), AL(t)), \]  

(2)

with \( K \) now representing physical capital and \( L \) human capital. So output is a function of physical capital and the product of technological level and human capital. According to Arrow (1962), two assumptions are made.
to introduce knowledge spillovers. First, learning-by-doing works through investment. Empirically, the relationship between patents and investment backs up this first assumption. The second assumption is not so undisputed: Knowledge is a public good. While the first condition of a (pure) public good (non-rivalry) is fulfilled, there are problems with the concept of non-excludability. Firms have incentives to keep their knowledge secret. They might fear that other firms use their technology to enter the market and outbid them. This would be a classic case of free-riding: The second firm does not have to include investment in R&D in their pricing, so they can offer lower prices. If new technology formerly was an entry barrier for competitors, it has now turned into an invitation to enter the market and drive down revenues. The usual way to counter this is to acquire a patent of the technology invented.

Anyway, a model that would allow for knowledge spillovers would replace the level of technology $A$ with the world stock of knowledge, embodied in the world physical capital $K^\omega$, resulting in

$$Y(t) = F(K(t), K^\omega(t)L(t)), \quad (3)$$

An investment of a firm would result in two effects. First, the single firm produces more efficiently because $K$ rises and $K^\omega$ enhances the productivity of human capital. Second, all other firms produce more efficiently, because the enhancement in labor productivity is not restricted to the single firm, but beneficial to all other firms as well. This is the knowledge spillover.

If growth could be described like equation (2), the empirical search for spillovers would be quite simple. One would try to find statistics to represent $Y$, $A$, $L$, and $K$. For instance, Sjöholm (1999) defines $Y$ as value added, $A$ as level of productivity, which is influenced by the level of FDI, $L$ as number of employees and $K$ as total investment. $A$ is then represented in three different cases by (1) capital intensities in an industry, (2) patent fees in an industry, and (3) difference in labor productivity between MNEs and national firms.

To allow for competition spillovers, it is common to use the Herfindahl-Index$^{12}$ and the effective rate of protection (ERP). The ERP is used as a measure of the degree of openness to foreign competition. I would call it an ‘index of uncompetitiveness’. The higher the index is, the more uncompetitive is the industry relative to the world market$^{13}$. Thereby, industries can be separated into groups with different market structures. Markets can be competitive and open (Herfindahl-Index close to 0, ERP close to 0 or even negative), or uncompetitive and closed against foreign competition (Herfindahl-Index close to 1, ERP much higher than 0). This brings us

$^{12}$A measure for competition in an industry. It equals the sum of the squared relative market shares of all companies in a market.

$^{13}$See Corden (1966)
straight to the empirical literature.

4 Empirical findings and their interpretation

4.1 Industry-level data and empirical findings

The empirical search for intra-industry spillovers through FDI began with studies of Australia (Caves, 1974) and Canada (Globerman, 1979), followed by Blomström and Persson (1983). These authors wanted to find out whether differences in technical efficiencies of Mexican plants in part derive from spillover efficiency associated with foreign direct investment\(^{14}\).

Usually, an AK model was estimated and industry-level data used. Then the industries’ productivity and productivity growth were estimated. It was assumed that sectors with a relatively high foreign participation would feature higher productivity and/or productivity growth rates because of ‘productivity spillovers’. Finding higher productivity of local firms would signify an one-time ‘push’ from FDI, while a higher productivity growth rate would point to a long-term effect.

Blomström and Persson (1983) chose Mexico for their research. Although the authors did not explicitly mention it, proof of higher productivity in industries with a high foreign share could have paved the way for a new development strategy. The productivity increase would have been found to be transmitted via spillovers between local and foreign firms. Economic growth or catch-up could then be achieved by attracting MNEs, which today is a standard strategy for many countries\(^{15}\). Certainly this conclusion was not reached because the results did not support it.

The next point concerns the applied empirical setup. Blomström and Persson (1983) state that the competition effect might be the most important spillover channel. This finding was highlighted by following papers. Just like trade, FDI increases competition\(^{16}\). The problem was that many researchers tried to neutralize any effects on competition in their setups. They used the Herfindahl-Index to isolate this effect from their findings, believing that they would get rid of an error. But this error that they separated might have been the most important part of the whole study.

The competition effect is easy to observe because it has a short time lag. Entry of firms instantly raises competition, and sometimes even the possibility of entry has this effect. The other spillover channels would have a time lag that is measured in years. Workers switching from MNEs to local

\(^{14}\)Blomström and Persson, 1983, p.493
\(^{15}\)See Kumar (1998), Chapter 8
\(^{16}\)If it does not result in a monopoly of the MNE, that is.
firms, adaption of local suppliers to MNEs needs, all this takes time. So the standard growth theory setup has a time lag problem regarding spillovers. Balasubramanyam, Salisu and Sapsford (1996), Sjöholm (1999) as well as Kokko, Zejan and Tansini (2001) used similar theoretical foundations trying to shed light on the effect of trade regimes (import/export-oriented, ERP) on the existence of spillovers from MNEs. While evidence for the influence of the trade regime on productivity spillovers was mixed, one fact was undisputed: Trade policy can change the incentives for MNEs. There are assumed to be two types of FDI\textsuperscript{17}. The first type is motivated by supply-conditions, mostly cheap factors of production. Vertical production networks are a prime example: Parts are produced at the cheapest location, then exported to a destination next to a large market, where the parts are assembled and the product is sold. For instance, a big German car maker gets its seat fillings from Brazil. A supplier there uses coconut filling to produce seats. The Brazilian firm’s seat fillings are cheaper than German products. The product is then exported to Germany where it is delivered to the assembly plants.

The second type of FDI is motivated by demand-conditions, mostly proximity to large markets. This might explain why China accounts for such a big part of global FDI inflows (See UNCTAD (2003), p. 8). Restrictions to trade influence the attractiveness of the country for the two types of FDI. While the introduction of trade barriers might even be attracting demand-motivated FDI (by raising the relative costs of trade), supply-motivated FDI will be harmed because of higher exporting costs.

4.2 Panel data on a firm-level and empirical findings

Haddad and Harrison (1993) criticized the use of aggregated data and suggested to look at the firms one by one, thereby using panel data on a firm-level. Total Factor Productivity (TFP) was selected as the main variable in place of labor productivity. Labor productivity in foreign firms was found to be only 70\% compared to local firms (I pick up that point later). Maybe that prompted Haddad and Harrison to switch to TFP instead. A high share of foreign firms in a sector was assumed to be a reason for spillovers. However, no significant spillovers were found. It seems that the MNEs take over considerable market share in the short run. This leads to under-utilization of capacity at domestic-owned plants, which is lowering their TFP. Overall, local firms’ growth rates outgrew those of foreign firms nonetheless, but this was not found to depend on the share of foreign firms in the sector.

The results were confirmed by Aitken and Harrison (1999) for Venezuela. FDI is found to harm the productivity of domestically owned plants. On the other hand, joint ventures are likely to gain from FDI. This depends\textsuperscript{17}I do not stick to the words of the authors, but use a different set of vocabulary. That way things are easier to compare in the next sections.
on the share of foreign investment, though. In foreign-dominated sectors the likeliness of spillovers from FDI to joint ventures increases. Aitken and Harrison also deal with the question of FDI being cause or reason for high productivity. They find that foreign investors choose higher than average productive plants. This productivity bias would account for at least a part of the higher productivity in industries with a high foreign share.

The next question to be examined is that of ‘localized spillovers’. The authors test that hypothesis by adding a regional FDI coefficient to their model. Once more they find little evidence of spillovers from local FDI. Variation of parameters would not change the result. Even when stretching time periods for spillovers to up to eight years, spillovers remain non-existent. Still, long-run effects might not have been captured due to lags in time. Also, Venezuela might not have been the best place to look for spillovers from FDI. Aitken and Harrison reckon that results might be different in more export-oriented economies.

4.3 What can be learned of the empirical research?

After reviewing the literature on spillovers, the first thing to note is that it has proved to be very difficult to measure spillovers. This might not come unexpected. Spillovers are hard to trace, and difficult to isolate from other effects. Further, the statistical methods influenced the results. Goerg and Strobl (2001) find that results depend on the kind of data used. What is clear, however, is the highlighted role of spillovers found in nearly all the research designs.

Only Blomström and Kokko (1998) turn to the role of linkages between MNEs and local firms. Backward linkages are important in that MNEs help their suppliers to reach a certain level of product quality. MNEs can provide training and help, give advise on marketing and use their international connections to find buyers/suppliers, for instance. However, Blomström and Kokko cannot come up with an empirical study to support their claims, basing them on different papers instead. It is surprising that the empirical literature on linkages is very slim, given that the term is *en vogue*.

Throughout all conducted studies foreign firms paid higher wages than firms of the host country. Wages in the MNE’s subsidiaries are up to 50% higher than in local firms\(^{18}\). This holds for developing countries as well as developed countries. The reason of that is still unclear. Lipsey (2002) reckons that MNEs are interested in attracting the best workers and pay a skill-premium. Another finding is the higher productivity level of foreign firms. This goes for TFP as well as labor productivity. Lipsey (2002) finds that the competitive advantage of foreign firms stems from their larger size and, at times, their greater use of purchased inputs. Baldwin, Braconier

\(^{18}\)Lipsey (2002)
and Forslid (2001) went on to conclude that a higher FDI penetration level leads to more rapid growth in that industry.

No new insights have been derived from the technology gap discussion. Sjöholm (1999) found that the higher the technology gap is, the higher spillovers are. This could be translated into: The more one lags behind, the more room there is for catch-up. On the other hand, the existence of extremely large technology gaps could prohibit spillovers. The gap would be too big to close. A large technology gap is the result of under-development in the past. A technology gap can then be understood as the outcome of a set of factors, like a lack of infrastructure, an ill-educated work force, and so on.

Barrell and Pain (1999) find that an MNE’s decision on the location of FDI - and therefore the cost of introducing a new technology - depends on skills and knowledge of the local labor force, among other things (mainly the scope of benefits from agglomeration economies). It would now be difficult to think of a reason why a MNE should invest in that economy and bring some advanced technology with it. The existence of a technology gap is caused by under-development. As far as I see it, there does not seem to be a difference between the concepts of ‘underdevelopment’ and ‘technology gap’.

Summing this section up, the recent empirical literature on technological spillovers does not establish a clear link between FDI inflows and productivity of local firms. Spillovers that flow from MNEs to local firms and raise their productivity have not been proved to exist. The research design has often been like this: if industries with a large foreign share grow faster than the mostly domestic industries, the existence of spillovers would be confirmed. The authors of the literature focus on technological spillovers, while mostly ignoring the concept of linkages. This problem is rooted in the growth theory: spillovers are always understood as knowledge spillovers. Knowledge is a public good, and externalities arise between firms.

The assumption of knowledge being a global public good is a strong one. Often, knowledge is protected by copyright laws, trademarks and other intellectual property rights. A fair amount of knowledge is embodied in machinery, which cannot be assumed to spill over easily. Skills necessary to make good use of technology have to be acquired. Then, the actual process of innovation is very complicated - as it involves social interactions - and is not likely to be explained well by a couple of equations. For instance, the existence of knowledge is not enough to produce an innovation. What is necessary is an economic motive for innovation. Thus, MNEs looking for a way to enter new markets or serve existing markets from a cheaper basis might act as a catalyst. When they demand better product quality from their local suppliers, they deliver a reason to innovate. This might be

\[ \text{von Hayek (1945)} \]
what Hirschman had hoped for, and that is also why many governments in developing countries are trying to attract MNEs.

The focus of growth theory on knowledge is too tight. While knowledge surely is important, knowledge alone is not sufficient for economic growth. To innovate successfully there needs to be an economic motive, like being a monopolist. Market structure influences innovation\textsuperscript{20}. The empirical literature on spillovers from FDI reviewed above claims that competition spillovers are influencing the growth rates of local firms. This essential issue cannot be modeled by growth theory. The aggregated view, while assuring solvable models, blinds the theory with respect to technological change through MNEs.

5 Next, please: New Economic Geography

The fast-growing literature on NEG has already produced some elaborate models including MNEs and FDI. Coming from a different direction than the empirically motivated literature, it is considered necessary to explain the basic workings of the core-periphery (CP) model and to clarify the vocabulary. After that we will take a look at different types of models. One model features spillovers and growth, the other includes endogenous MNEs. The aim of this section is to establish a link between the NEG theory and the empirical literature on spillovers from FDI.

5.1 The core-periphery model

The core-periphery model is the standard NEG model. It includes all the main features and will be discussed shortly\textsuperscript{21}. The CP model is constructed from a mix of ingredients. There is a spatial concept dating back to von Thünen (1826) which allows for multiple regions in space. Transport costs arise when anyone sells a product in a distant region. These transport costs are modeled as iceberg costs. A certain amount of the good ’melts away’ on the way to its destination. The farther a product is transported, the higher are the absolute ’costs’ of transport. In the end, to deliver one unit to a distant location the firm has to send more then one unit, depending on the distance. Someone once thought of iceberg costs as a horse pulling a wagon of grain to a distant market, thereby consuming part of the grain on his way\textsuperscript{22}.

The CP model features two sectors, manufacturing and agriculture. There are two corresponding factors: workers in manufactures and agricultural workers, both located in two symmetric regions. The products of

\textsuperscript{20}The discussion of cheaper drugs for Africa is an example. Health companies claim that resulting diminishing profits would in return lower the incentives to innovate.

\textsuperscript{21}For an extensive discussion, see Brakman et al. (2001)

\textsuperscript{22}See von Thünen (1826), Chapter 4
the manufacturing sector compete with each other in a Dixit-Stiglitz monopolistic competition framework. Consequently, firms diverge only in the variety they produce. The agricultural goods are transported costlessly and sold in a perfectly competitive market. Returns are constant and products homogeneous. Workers in both sectors seek to maximize their utility, which consists of the consumption of manufactures and agriculture. These are bought on the local market. The real wage depends on the nominal wage and the local price index. Workers spend their income locally and are mobile, moving to the region that offers the highest real wage.

Three forces exist that shape the effects of agglomeration. The home market effect is a self-reinforcing agglomeration force. It manifests itself through worker migration and their impact on markets. We assume a symmetric equilibrium in which one worker moves from south to north. Then it gets more expensive for Southern firms to reach this consumer in the north. There is an incentive to follow him and minimize transport costs. Put in another way, the arrival of a new worker/firm in the north enlarges the local market and makes it more attractive.

The home market effect depends on transport costs (the higher they are, the stronger the effect) and fixed costs in production (again: the higher they are, the stronger the effect). The home market effect is often considered to be a "backward linkage". But does this fit well with Hirschman’s original concept of linkages? Obviously, in the NEG model the producers benefit from lower transport costs since the market has moved closer to their location. On the other hand the market effect is a cumulative effect, because it also lowers the consumers’ price index.

Since consumers and producers are identical, it is a little bit tricky to decide if this is a linkage. According to Hirschman, linkages exist between firms or industries. Here, firms benefit from consumers moving to their direction so they can save on transport costs. Strictly speaking, this is not a backward linkage, although consumers and firms are identical in the model. Still, the home market effect is based on a magnification effect that exists because consumers cluster together to form a big market.

Coming back to our example of a worker migrating from south to north, there is a second effect on the local price index. The worker brings with him a variety that had to be transported from the south before. It is locally available cheaper than before and lowers the local price index. This would be a forward linkage, were not consumers instead of firms to benefit from it. In the scenario this implies a rise in real wages in the north, while in the south real wages fall correspondingly. With real wages rising in the north and falling in the south, more workers will be inclined to migrate to the north. Venables (1996) developed a model where firms use other firms

---

23 Without transport costs, there is only one integrated ‘world’ market
outputs as inputs, thereby generating a model with forward linkages à la Hirschman.

The third effect is pushing in the opposite direction. It is called ‘market crowding’ or ‘local competition’ effect. More firms translate directly into more competition. That is a result of the love for variety: Because consumers value variety, each new product cuts into the markets of all the other products. This lowers profits and ultimately translates into lower wages (and less firms). The location is loosing attractiveness in terms of real wages\textsuperscript{25}. There is a tension between locating a firm in the neighborhood of the competitors to form a large market, and locating far from each other to soften competition. This is maybe ‘the’ essential trade-off of the CP-model\textsuperscript{26}.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{cp_model_medium_trade_costs.png}
\caption{The CP model, medium trade costs\textsuperscript{27}}
\end{figure}

Because both agglomeration effects are self-reinforcing, the model’s equilibria need to be checked for stability. The equilibrium is considered stable only if there is no incentive for any worker to switch to the other region (and enjoy a higher real wage there). If the relocation of one worker would

\textsuperscript{25}The extent of the effect depends crucially on the degree of consumer’s love for variety. Because consumers value variety, the price index sinks with every new variety. This may overcompensate the loss in nominal wages and therefore lead to more agglomeration.

\textsuperscript{26}The footloose capital model by Martin and Rogers (1995) was adapted by Baldwin et al. (2003). Although much simpler analytically than the CP model, it shows quite the same characteristics.

\textsuperscript{27}Graph taken from a lecture by Wooton (2001), with permission
improve his real wage, it will often be followed by what is called circular causation: After the worker has relocated, changes in real wages induce more workers to follow. Accordingly, in Figure 2 there are three stable equilibria: the symmetric outcome, where $L_1 = L_2$, and the core-periphery outcomes, where $L_1 = 0$ and $L_2 = 0$. The equilibria to the left and to the right are unstable.

When we take the left equilibrium, for instance, we see that a worker defecting to region 1 (we move to the right) raises the relative wage in region 1 ($\omega_1 > \omega_2$). This will lead to even more workers switching regions until we are at the symmetric equilibrium. At this point any incentives to relocate have ceased. A worker relocating from region 2 to region 1 (we move even farther to the right) will find that after his move is completed, real wage in region 2 is lower than that in region 1. Would we have started with a worker relocating to region 2 instead of region 1, the result would have been a different story. To the left of the left equilibrium the real wage of region 2 is higher than that of region 1. A cumulative process sets in, but this time it goes in the other direction and leads to complete agglomeration! There is a fair amount of fuzzy math in the dynamics of the CP model 28.

5.2 A model with endogenous MNEs

Ekholm and Forslid (1999) extend the CP model by adding MNEs. They find that the core-periphery structure is an outcome of restrictive assumptions. The most significant, Ekholm and Forslid argue, are that MNEs can locate only in one location and that economies of scale exist on a plant level only. In reality firms would realize economies of scale on a firm-level which would then prompt them to operate in more than one region. Two types of multi-region firms would then have to be examined: Horizontal MNEs and vertical MNEs.

Horizontal MNEs are driven by a tension between the costs of an additional plant and the costs of trade. If both of these costs are relatively high, there will be only single-region firms. Only if the setup costs are low enough, a firm will choose to become multinational to save transport costs. The main finding in this model is that in symmetric equilibrium there is no trade. Producing in both regions means that a firm can adjust output to changes in factor prices and demand in either of the locations 29. The entry of single-region firms is not possible due to the MNEs competitive advantage. The competitive advantage stems from the fact that two plants share the costs of the headquarter, which produces under economies of scale.

Vertical MNEs are different in that they consist of one or more plants and exactly one headquarter. Now variable production costs arise only at

---

28 Maybe that is why it looks so realistic.
29 No factor movement is required: a rise in demand stems from the relocation of consumers, which are at the same time making available their labor in that region.
the plant, while fixed costs take the form of headquarter services. These may include back-office facilities, management, marketing, and R&D. A vertical MNE would have its headquarters in the low-wage region, since services are freely tradable. Its plant would then be located in the other region, since otherwise it would not be a MNE. Two effects arise from a model with vertical MNEs: First, the symmetric equilibrium is destabilized at high trade costs, since firms tend to locate their headquarters in the region with the lower wage. Second, the tendency of the model to produce a core-periphery outcome is reduced. It is now less costly for firms to move their plant to the smaller region, because the fixed costs for headquarter services do not change.

A problem arises with the resulting equilibria: Domestic and foreign firms do never coexist. Either there are domestic firms (and multinationals) or foreign firms (and multinationals). At times, the regime of country A switches from foreign firms to domestic firms. This would have catastrophic results in region B. The number of firms would sink dramatically, then price indexes and nominal wages would follow. Serious repercussions could be expected in A, but the link is cut off in the model.

5.3 A model with endogenous growth

The first NEG models featuring spillovers and endogenous growth came from Martin and Ottaviano (1998) and Baldwin, Martin, Ottaviano (2001). While the first paper contains global spillovers, the later work assumes that spillovers are localized. The main idea is the introduction of an innovation sector\(^{30}\), where labor is transformed into capital. Capital is then used as the only factor to cover fixed costs of the manufacturing sector. New capital increases the number of varieties, thus lowering revenue for all existing firms due to the Dixit-Stiglitz monopolistic competition framework. Due to Learning-by-doing the cost of innovation decreases in time. Going into details here would be interesting, but that certainly lies beyond the scope of this paper. As a compromise, the logic of the model will be outlined in the following passage.

Baldwin, Martin, and Ottaviano (hereafter BMO) consider a standard NEG model which features a third capital-constructing I-sector. This sector produces capital, which is needed for production of the manufactures sector. In the I-sector, one unit of capital is needed to cover fixed costs, while marginal costs are covered by labor. In this model, units of capital equal regional capital stocks, I-sector production and number of varieties. Capital is depreciating over time. The spatial effect of knowledge spillovers is expressed through \(\lambda\), which can take a value between ’local spillovers’ and ’knowledge as a public good’. These spillovers cause an inter-temporal ver-

\(^{30}\)This is a standard feature in many endogenous growth models
tical linkage, because workers benefit from the location of past production. Being more productive, they make innovation more attractive in their region. At that point the circular causality is completed. This is described in the literature as the emergence of growth poles and sinks. While the region in the growth sink stagnates, the growth pole attracts all new varieties.

An interesting result that is derived from the BMO model is that geography affects growth. When trade costs are high, the regions are integrated neither in industry nor in innovation. As the break point is reached, a CP outcome arises. When innovation is located in a single region, local spillovers in that sector raise productivity. The equilibrium growth path is rising with it so that in the end economic integration has a positive influence on growth. The benefits of stronger growth are different in the regions. The core region definitely enjoys a rise in welfare, while the results for the periphery region are not so straightforward. Much depends on $\mu$, the expenditures share of manufacturing goods. The higher $\mu$ is in the periphery region, the more do its people profit from a lower price of manufactures. Therefore a sufficiently high enough $\mu$ can lift the periphery’s welfare over its pre-CP level. Here, both regions gain from a CP outcome\textsuperscript{31}.

6 Perfect Match?

Now that some NEG models have been characterized, they will be confronted with the empirical results of section 4. Hopefully, the results of the empirical literature are to some extent reflected in the NEG models. As theory and reality are put side by side, the existing models will be evaluated.

6.1 The CP model

The basic CP model and the empirical spillover literature do not mix very well. The main issue is that FDI is modeled through migration of workers. This is highly unrealistic in an international context and can thus only be understood as a metaphor. International labor mobility is low, and it does not seem to change soon - not even in the European Union\textsuperscript{32}. If labor migration is a metaphor for shifts in demand, then it can be questioned if it is modeled well. Why should such a shift in demand start a cumulative process? Is there a reason why a shift in demand can only be modeled by a simultaneously shift in supply\textsuperscript{33}?

What the CP model basically does say about FDI looks utterly pessimistic: If we have symmetric regions, inequality will arise when transport

\textsuperscript{31}For a more detailed description of the model’s workings see Chapter 7 ‘Global and Local Spillover Models’ in Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud (2003)

\textsuperscript{32}Puhani (1999)

\textsuperscript{33}This is due to the fact that each worker represents a firm. That way demand and supply always ‘go together’.
costs go down. The first region to attract an additional worker will come out to be the core region. During the restructuring process, there will be one-way FDI flows into that region. That would be the interpretation of the migration of workers, fleeing the sinking ship of the periphery region, attracted by higher real wages in the core region. The mechanism does not fit well with empirical results in general. For instance, empirical findings tell us that similar regions feature relatively large FDI flows between them (take France and Germany, for instance). The FDI flows in the CP model’s symmetric equilibrium are exactly zero\textsuperscript{34}.

Also, differences in technology are difficult to model because the monopolistic competition setup does not easily allow for different techniques\textsuperscript{35}. Differences in technology are an important issue for FDI-related decisions. Productivity to a certain extent depends on technology. Differences in firm technology would also allow for an equilibrium with domestic and foreign firms. If technologies are identical, then between domestic or foreign firms there will be one cheaper alternative. In the end, all firms will choose to be of that kind. Since there is only one input factor, this problem will persist.

6.2 The MNE model

Ekholm and Forslid (1999) tried to shed light on the role of MNEs in a CP model. They introduce horizontal and vertical MNEs. Horizontal MNEs face the task of minimizing transport costs, choosing to export from their home plant or to build a second plant in the foreign economy. The choice here is made to minimize costs. Vertical MNEs can choose different locations for headquarter and plant. Headquarter services are traded freely in the firm, therefore they will be produced wherever nominal wages are lowest. There is a cost of splitting headquarters from the plant.

While the CP model was not build to feature FDI and its examination with respect to it could have been a little unfair, the model of Ekholm and Forslid features FDI explicitly. The construction of a plant in a foreign country is nothing but FDI, although the authors did not emphasize it. Maybe they avoided that vocabulary because the model is not well suited to discuss FDI. However, the theme of cost-cutting MNEs fits very well with the assumed role of MNEs in technology transfer.

Nevertheless, the CP model’s weaknesses are not erased. Still, labor is the mechanism through which FDI ‘works’. Fixed costs of a second plant are ‘worked out’ in labor units. A firm that economizes on transport costs by relocating its plant to the foreign region moves its workers there. This is

\textsuperscript{34}In equilibrium wages are equalized, therefore worker migration - which equals FDI in this case - has ceased.

\textsuperscript{35}We already have one degree of freedom in the CP model, which makes finding solutions very difficult. Adding firms that realize a point from a technology function will make work harder still.
modeled as a short term decision. The fixed costs are paid for by the end of the period, then the firm chooses locations again! This does not fit well with empirical results, which rank FDI as a long-term decision that is only reversed with a huge monetary loss.

Also, the modeling of headquarter services seems a little bit awkward. Firms can place their headquarters freely, because its services are traded freely within the firm. The authors then assume that the headquarter goes to the region with the lowest nominal wages. If this overcompensates the splitting costs, headquarter and plant will be separated. The question here is: Why should headquarter jobs go to the region with the lowest nominal wages? This breaks with the structure of the rest of the model. Workers are always attracted by high (real) wages, they form big markets that firms themselves are attracted to. Why should firms think that they could offer workers low-paid jobs?

Despite those shortcomings the Ekholm/Forslid model seems to be better equipped to model international growth processes. MNEs arise and with them FDI is brought into the setup. For horizontal MNEs, the construction of a second plant in a foreign region would be a FDI flow. Spillovers arise from the influence of the MNE on existing linkages. This looks just like the scenario from the empirical search for spillovers from FDI.

6.3 The endogenous growth model

The endogenous growth model by Baldwin, Martin and Ottaviano (hereafter BMO) is a modification of the CP model. The mobile factor is not labor, but capital. This is definitely a better choice to model FDI. Capital flows arise endogenously, resulting from actors optimizing their income over time. Contrary to that Lipsey (2002) denies the nature of FDI being capital movement: FDI would be about ownership, not location. Skills and knowledge would be the content of FDI rather than capital. The character of capital flows in the BMO model might be more of a portfolio investment-type than of a typical FDI-type. Another difference from the CP model is that, because cost and demand linkages are cut, all self-reinforcing effects are gone. However, linkages were found to be important spillover channels. Also, empirical studies proved the importance of a big markets\textsuperscript{36}.

The BMO model itself seems to work very well, and most of its results seem reasonable. The question of symmetry or CP outcome depends on the localization of spillovers. If knowledge is not a public good but a local spillover, the innovative sector will end up in one region. The core will attract more firms because of the market effect, so a CP outcome is quite likely. The empirical literature affirmatively finds that spillovers are local\textsuperscript{37}. The BMO model seems to use a compatible concept regarding spillovers.

\textsuperscript{36}See Davis and Weinstein (2003)
\textsuperscript{37}See Audretsch (2003)
What is also interesting is the emergence of growth sinks and growth poles. Lagging regions seem to stall for a long time, while the advanced region keeps on growing constantly. This is a prediction that was examined empirically. Fischer (2003) found that in the last 20 years developing countries outgrew industrialized countries, if one focuses on population and not countries. However, there is not much empirical support for the BMO model’s prediction that capital flows from the periphery region sustain growth in the advanced region.

7 Conclusion

After reviewing growth theory and NEG models in the previous sections, it seems that NEG models are better equipped to feature interactions of MNEs and local firms. MNEs and market structure can be modeled explicitly, as well as spillovers or linkages. These seem to be the main ingredients when modeling technology transfer through MNEs. Supporting this view, the empirical literature found the role of MNEs in technology transfer to be important. The growth theory with its aggregated views of firms and markets is not able to incorporate these ingredients. Empirical studies based on a growth theory framework did not confirm its predictions. This may be due to a flawed research design, so that it does not necessarily mean that local firms do not benefit from arriving MNEs.

The NEG suggests a different approach. The role of linkages is emphasized, market access and transport costs play an important part in economic integration. MNEs can be modeled explicitly. The behavior of MNEs in the model of Ekholm and Forslid could be a starting point from which to develop a more comprehensive model of MNEs in a globalization-scenario NEG model. It could turn out to be an improvement if a vertical structure is expressed through firms in a vertical production network instead of firms splitting themselves up. That assumption is somewhat unrealistic, as recent relocations of outsourced headquarter services, like accounting, from developed countries to developing countries tend to show. It might be a good idea to extend the model by a capital good that embodies technology. That way, the transfer of technology and its spillovers can be modeled explicitly.

This leads us to another problem. In the NEG models all firms use the same technology. Therefore it is not possible to let domestic and multinational firms co-exist. A model with different technologies might bring relief. When firms are allowed to choose between different technologies, some of them might want to go multinational while others prefer to stay domestic.

---

38 See Keller (2001)
39 See Venables (1996) might make a good point to start from.
40 See "The great hollowing-out myth", Economist Feb 21th: p. 41-43
Decisions could resemble the OLI-Paradigm\textsuperscript{41} in that firms choose to export or move to the other region (FDI), depending on questions of ownership.

A workhorse for incorporating capital goods might be the BMO model featuring endogenous growth. It is suited very well to incorporate capital flows like FDI. Modeling FDI through labor migration, even if it is taken metaphorically, might lead to dangerous misunderstandings about trade. The biggest drawback is that labor migration always creates a looser and a winner. The core region has a higher real wage and usually it is very unlikely that it will break up again. Also, labor migration across national borders is a phenomenon not to be expected on a large scale.

It seems unlikely that one model can explain the whole world (trade) satisfactorily. Some assumptions have to be made, and then modeling a NEG framework with MNEs and technological spillovers may turn out to be like chasing a mole: of all the mole-hills, one can cover only two at a time. Still, it should be worth to consider which holes there are and where to look for new insights.

\textsuperscript{41}Dunning (1977)
Bibliography


Brakman, Steven, Garretsen, Harry, and van Marrewijk, Charles (2001), 'An Introduction to Geographical Economics: Trade, Location and Growth', Cambridge


Christaller, Walter (1933), 'Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen', Jena


Dunning, John H. (1977), 'Trade, Location of Economic Activity, and the Multinational Enterprise: A Search for an Eclectic Approach', in *The International Allocation Of Economic Activity* by Ohlin, Bertil, Hesselborn, Per-Ove, and Wijkman, Per (eds), Chapter 12: 395-418


Globerman, Steven (1979), 'Foreign direct investment and 'spillover' effi-
ciency benefits in Canadian manufacturing industries’, Canadian Journal of Economics, 12, 1: 42-56

Haddad, Mona and Harrison, Anne, (1993), 'Are there positive spillovers from direct foreign investment?', Journal of Development Economics, 42: 51-74


Hirschman, Albert O. (1958), 'The strategy of economic development’, New Haven


Kumar, N., ed. (1998), 'Globalization, foreign direct investment and technology transfers : impacts on and prospects for developing countries’, London


Lösch, August (1940), 'Die räumliche Ordnung der Wirtschaft : eine Untersuchung über Standort, Wirtschaftsgebiete und internationalen Handel’, Jena


Puhani, Patrick (1999), 'Labour mobility - an adjustment mechanism in Euroland? Empirical evidence for Western Germany, France and Italy', *IZA Discussion Paper* 34


Rosenstein-Rodan, Paul N. (1943), 'Problems of industrialization of Eastern and South-Eastern Europe', *Economic Journal*, 53: 202-211


Schumpeter, Joseph A. (1942), 'Capitalism, Socialism, and Democracy', New York


van de Klundert, Theo, and Smulders, Sjak (2001), 'Growth Theory in Historical Perspective', Cheltenham


von Thünen, Johann Heinrich (1826), 'Der isolirte [isolierte] Staat in Beziehung
auf Landwirtschaft und Nationalökonomie: oder Untersuchungen über den Einfluß, den die Getreidepreise, der Reichtum des Bodens und die Abgaben auf den Ackerbau ausüben’, Hamburg


Wang, Jian-Ye, and Blomström, Magnus (1992), 'Foreign investment and technology transfer', *European Economic Review*, 36: 137-155

Weber, Alfred (1909), 'Über den Standort der Industrien', Tübingen

Wooton, Ian (2001), 'International Trade', *Presentation, held at the University of Göttingen, May 2002*