universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

"Evidence on trade-related R&D spillovers from the North to the South"

Verfasser

Wendelin Moritz

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien, im August 2009

Studienkennzahl lt. Studienblatt:	A 140
Studienrichtung:	Diplomstudium Volkswirtschaft
Betreuer:	Dr. Neil Foster

Contents

1 Preface....4

2 Introduction....5

3 Literature Review....7
- 3.1 Endogenous Growth Models....7
- 3.2 R&D spillovers....8
- 3.3 Empirical research on international R&D spillovers....9
  - 3.3.1 The starting point – CH (1995)....9
  - 3.3.2 Critics on CH's foreign knowledge stock – LP (1998)....11
  - 3.3.3 Using randomly matched trade partners – Keller (1998)....12
  - 3.3.4 The North-South dimension – Coe/Helpman/Hoffmaister (1997)....12
  - 3.3.5 Focus on capital goods trade – Xu/Wang (1999)....13
  - 3.3.6 Interpretations of different weighting schemes – FFG (2002)....13
  - 3.3.7 Dealing with “Indirect” Spillovers – Lumenga-Neso et al. (2001)....14
  - 3.3.8 Including different control variables....15
    - 3.3.8.1 Human Capital....15
    - 3.3.8.2 Spillover control variables....17
    - 3.3.8.3 Relative Backwardness....18
    - 3.3.8.4 Institutional variables....19
    - 3.3.8.5 Geographic distance....20

4 Model specification and variable description....21

5 Data description....28
- 5.1 GDP....28
- 5.2 Foreign knowledge stocks....30
- 5.3 Human capital....32
- 5.4 openness....35

6 Econometric preliminaries 36
- 6.1 Panel data estimation....36
- 6.2 Unit root tests and cointegration tests in panel data....38

6.2.1 What is a unit root, what is cointegration?.......................................38
6.2.2 Testing for unit roots in panel data..39
6.2.2.1 Levin/Lin/Chu (2002)..40
6.2.2.2 Im/Pesaran/Shin (2003)..41
6.2.2.3 Pesaran/Smith/Yamagata's CADF-test (2007).........................41
6.2.3 Testing for cointegration in panel data..42
6.2.3.1 Using OLS and modifications..42
6.2.3.2 Westerlund's cointegration test (2007)....................................43

7 Empirical Results..45
7.1 Results of Unit Root Tests..45
7.2 Results of Westerlund's test for cointegration..47
7.3 Results of a DOLS estimation..49
7.3.1 Step 1 – simple regression on ln(LP), openness, spillover control...50
7.3.2 Step 2 – simple regression on ln(CH), openness, spillover control. .52
7.3.3 Step 3 – exploring the impact of human capital 55
7.3.4 Step 4 – introducing "relative backwardness".................................58
7.3.5 Step 5 – the role of patent protection..61
7.3.6 Step 6 – the ease of doing business and regional dummies 64
7.3.7 Overview over all interaction-term-specifications............................67
7.3.8 Detailed discussion of the preferred specification (regression 19)…68

8 Conclusion..72

List of Literature...73

Appendix A – "Perpetual Inventory Method" 77
Appendix B – Abstract in English...78
Appendix C – Abstract in German..79
Appendix D – curriculum vitae in German..80

1 Preface

Theories of economic growth have always had some sort of undefined attractiveness for me as they try to model the economic evolution of the aggregate society and therefore provide many important policy consequences. I consider the results of neoclassical theory to be a great contribution to the understanding of growth, yet since my first lesson in growth theory I have always been a little unsatisfied with the assumption of technological progress being exogenously determined. Therefore the role that R&D plays when it comes to explaining economic growth was soon the focus of my interest. Besides the impact of a country's R&D performance on its own level of economic output, it has especially been the idea that a country's R&D performance affects another country's GDP that fascinated me. The rough choice of which issue to deal with in my thesis was thus not hard to take. After having worked through the literature on R&D spillovers I decided for an exact topic of my thesis, which should be an empirical piece of work on trade-related R&D spillovers from the OECD to developing countries (i.e. from the North to the South). This leads me to the formulation of the hypotheses I want to check for plausibility and for empirical evidence:

I claim that there exist substantial positive external effects of R&D performed in the OECD countries on the economic output of developing countries. These positive spillovers are carried by imports from the OECD to developing countries to some extent. They are in this sense “trade-related”.

I claim further that a developing country's return to foreign R&D depends on its stock of human capital and on its openness to trade. The higher the level of education the higher the return to foreign knowledge due to a higher “absorptive capacity”. The higher the level of trade the higher the return to foreign knowledge since the more open a country is the more is it able to benefit from foreign R&D activities. This is because knowledge is to some extent a private good and greater openness implies that more foreign products containing knowledge are then flowing into the country.

These are the hypotheses that will be considered in this work.

2 Introduction

The early neoclassical theories of economic growth (e.g. *Solow*, 1956) had no role for an endogenously determined technological progress, although already *Schumpeter* (1911) considered innovation and research to be relevant for economic growth. The role that R&D activities play when it comes to explaining economic growth was thought about in the beginning of the 1990s. *Grossman/Helpman* (1991) is the classical reference concerning this issue. The idea was that higher R&D expenditures rise productivity and finally economic output.

Especially the impact of foreign R&D activities on economic growth soon attracted much interest from economists. Several empirical studies tried to work out this impact by looking for evidence of foreign knowledge spillovers that are captured by trade, particularly imports, either within the OECD or from the OECD to developing countries. The idea was that imports of foreign products are a channel through which foreign knowledge that is contained in these products may enter the domestic economy. *Coe/Helpman* (1995), *Lichtenberg/van Pottelsberghe de la Potterie* (1998), *Coe/Helpman/Hoffmaister* (1997), *Falvey/Foster/Greenaway* (2002) are a subset of a larger number of studies that found strong evidence of a positive relationship between trade-related R&D spillovers and economic performance.

Moreover, human capital was expected to contribute to a country's ability to absorb foreign knowledge since better educated citizens might understand a foreign technology more easily than people with nearly no education. A high-skilled labor force will be more likely to be able to use a foreign technology. Moreover, it will be more likely to be able to copy and imitate foreign products. Supporting evidence was found by *Foster/Falvey/Greenaway* (2005) and *Coe/Helpman/Hoffmaister* (1997) for instance.

I will give a more detailed literature review in section 3, while section 4 will define the variables that will be used and explain why they will be used. Section 5 will shed light on the dataset. Econometric preliminaries will be given in section 6 before providing econometric results in section 7. Section 8 concludes.

3 Literature Review

3.1 Endogenous Growth Models

Economic growth is one of the most crucial issues in economic theory and the question of its origin and of its determinants some of the issues most thought about. The neoclassical view of growth theory concentrated heavily on the role of capital accumulation for creating wealth, while technological progress was assumed to be exogenous, i.e. determined outside the model and not influenced by other model parameters (*Solow*, 1956). The result was that in the long run technological progress is the only factor that can increase output per capita, yet nothing was told about how technological progress could be achieved.

As a response to the inadequacies of the neoclassical growth model, economists began to think about the causes of technological progress, developing the theory of endogenous growth. Today, due to the contributions of endogenous growth theory technology is seen at least to some extent as a result of economic activities and therefore as determined within the model itself (*Grossman/Helpman*, 1991; *Romer*, 1990; *Aghion/Howitt*, 1992). The main view is that economic incentives in the form of future profits lead firms to undertake commercially oriented innovation efforts, which are considered to be the "major engine of technological progress" (*Coe/Helpman*, 1995, 860). Important to mention here is that on the one hand, the innovation process is assumed to benefit from the available knowledge in society, i.e. knowledge is an input to innovation efforts, while on the other hand, knowledge should be regarded as the output of innovation efforts. As a consequence the productivity of an economy depends on the stock of knowledge and on the volume of research activities. Knowledge is at least to some extent a public and non-rival good and therefore bears remarkable growth potentials for a country. The empirical evidence on the importance of domestic R&D for economic productivity growth is convincing (*Griliches*, 1998; *Coe/Moghadam*, 1993).

3.2 R&D spillovers

From seeing domestic R&D expenditures as a determinant of economic growth it is only a small step to consider whether the R&D efforts of foreign countries will also have an impact on domestic economic performance. *Grossman/Helpman* (1991) provide some theory on this issue and point out that foreign knowledge will indeed have positive impacts on an economy's output level. The positive external effect of foreign R&D efforts on domestic productivity growth is the so-called R&D spillover from the donor country (the one that did the research) to the recipient country (the one that benefits from the spillover). This spillover is an obvious consequence of trade, foreign direct investment and other channels of knowledge diffusion. *Grossman/Helpman* identified four major channels through which knowledge diffuses between different countries (*Grossman/Helpman*, 1991; *Coe/Helpman/Hoffmaister*, 1997):

- International trade, especially imports, introduces a large variety of foreign intermediate goods and foreign capital goods to the domestic economy, thereby increasing the productivity of domestic resources. These new products are either new in a horizontal sense, i.e. they are complementary to each other and thus raising the variety of intermediate goods, or they are new in a vertical sense, i.e. they substitute old domestic goods and are thus raising the quality of intermediate goods (for a more detailed and formal description of these two approaches see *Grossman/Helpman*, 1991).
- Channels of communication that are provided by international trade will increase economic productivity by learning new production methods, product designs and organizational methods from the trade partner. This holds for imports and exports as well.
- Imitation of foreign technologies will be stimulated by international contacts (imports and exports) and also contribute to productivity growth.
- Besides from that, an increase in knowledge through international trade will raise a country's productivity in own R&D activities and in that way have an indirect impact on economic growth.

In my empirical work I will focus on international trade, and imports in particular, as the main channel of technology diffusion. Besides from the direct effect of international trade on productivity via the introduction of new and more productive intermediate and capital goods in the manufacturing sector there is also an indirect effect: By exploring the technology of a foreign capital good an economy can gain useful information that will serve as an input into further research activity, thus raising productivity in the R&D sector and also in the manufacturing sector.

Starting from the main idea that a country's R&D efforts may have an impact on the growth rate of another country, a remarkable amount of empirical studies have addressed this idea's plausibility and the significance of the spillover effect during the past two decades. I will now follow by reviewing some of them, focusing on papers that investigated R&D spillovers via the import channel, which is the one that will be considered in this thesis.[1]

3.3 Empirical research on international R&D spillovers

3.3.1 The starting point – *Coe/Helpman* (1995)

In 1995, *Coe* and *Helpman* (*CH* from now on) examined the importance of R&D spillovers using data from 21 OECD countries plus Israel over the period 1971-1990 (*Coe/Helpman*, 1995). They explained a country's total factor productivity (TFP) as a function of the domestic knowledge stock (or domestic R&D capital stock in other words) and the foreign knowledge stock. While the domestic knowledge stock had already been used as an explanatory variable for TFP, the idea that the foreign knowledge stock was also a determinant of TFP was new.

CH concentrated on imports as the major channel of knowledge diffusion and therefore constructed the foreign knowledge stock of each country in the following way, where the time index is omitted for reasons of simplicity:

$$S_i^f = \Sigma_j (m_{ji} / m_i) S_j^d$$

S_i^f … foreign knowledge stock of recipient country i

S_j^d … domestic knowledge stock of donor country j

1 There are also studies dealing with exports as the major channel of technology diffusion: *Funk* (2001) and *Falvey/Foster/Greenaway* (2004) are the main references.

m_{ji} … recipient country i's imports from donor country j

m_i … recipient country i's total imports

A country's foreign knowledge stock is therefore defined as the import-share-weighted average of the domestic knowledge stocks of trade partners. This view suggests that the higher is the share of imports coming from countries with high levels of technology the higher are the potential gains from R&D activities of the trade partners, which is reflected by the import share in the formula above. Going further the authors argue that not only the import shares (giving answer to the question “With which countries does a specific country trade?”) matter but it is at least as important to take into account the overall level of imports of each country. A country having a higher level of imports will be able to benefit more from its trade partners' R&D activities. This is due to the fact that knowledge is not a purely public good in reality (patents, secrets, …). The overall level of imports would be irrelevant if one assumed knowledge to be a purely public good, but since this is not very realistic it is not considered by *CH*. Therefore they interact the foreign knowledge stock of each country with its openness, defined as the ratio between total imports and total output. They then estimate the following equation, where they assume their variables to be cointegrated even though they cannot be sure about that due to the fact that “the econometrics of pooled cointegration are not yet fully worked out” (*Coe/Helpman*, 1995, 870):

$$\log(\mathrm{TFP}_i) = \alpha_i + \beta_i^d \log(S_i^d) + \beta_i^f m_i \log(S_i^f)$$

Notice that they use a level specification and do not difference their data since they want to benefit from the characteristics of cointegrated relationships, which will be explained later.

Their result is a highly significant coefficient of 0.294 associated with the interaction of the foreign knowledge stock and openness, which leads them to the conclusion that “there indeed exist close links between productivity and R&D capital stocks. Not only does a country's TFP depend on its own R&D capital stock, but as suggested by theory, it also depends on the R&D capital stocks of its trade partners” (*Coe/Helpman*, 1995, 875).

3.3.2 Critics on *CH's* foreign knowledge stock – *Lichtenberg et al.* (1998)

Lichtenberg and *van Pottelsberghe de la Potterie* (*LP* from now on) continued the empirical research on international R&D spillovers by considering the way *CH* constructed their foreign-knowledge-stock variable. They discovered that the import-share-weighted version developed by *CH* was subject to an "aggregation bias" *(Lichtenberg et al., 1998)*, which means that a merger between two countries would change the data in an inappropriate way (actually, it made a huge difference whether one measured the knowledge stocks separately and summed them up or whether one measured the knowledge stock of the "new" country after the merger: a merger between two countries would increase the world's knowledge stock, which would make only little sense), which of course is not a convincing argument in favor of the *CH* measure. Therefore using the same dataset as *CH*, *LP* introduced an alternative measure of the foreign knowledge stock that didn't suffer from this aggregate bias, and which can be written as:

$$S_i^f = \Sigma_j (m_{ji} / y_j) S_j^d$$

y_j … GDP of donor country j

Now the central ratio is the one between the recipient's imports from the donor m_{ji} and output of the donor y_j. This view looks at the spillover process from the side of the donor and weights the donor's domestic knowledge stock by the export shares. *LP* showed that their measure performed better in empirical settings due to the substantially reduced aggregation bias, under which the measure of *CH* suffers. Another interpretation of their different foreign-knowledge-stock variable will be discussed later.

LP also corrected for an "indexation bias" that emerged because *CH* indexed their data in an inadequate way, whereas *LP* decided to use the data in absolute values rather than indexed data.

Concerning the results *LP* do not differ much from *CH* in that they confirm their findings. The import channel is highly important for a country to gain from foreign R&D and the more open this country is the more will it benefit from it.

3.3.3 Using randomly matched trade partners – *Keller* (1998)

Keller (1998) found that using random weights, rather than the actual import-share weights, for the foreign R&D capital stock results in a higher R-squared than when using the actual trade shares. He concludes that his "random" weighting scheme performs better empirically than the one introduced by *CH* indicating that the structure of imports is of no importance for the knowledge spillovers flowing into a country. In a response to this *Coe/Hoffmaister* (1999) showed that *Keller's* weights are not in fact truly random, but simple averages with a random error. They use three alternative sets of truly random weights to construct the foreign knowledge stocks and find that the estimated equations explain less (i.e. they provide a lower R-squared) of the variation in TFP than the equations estimated by *CH*. So *Coe/Hoffmaister* (1999) revealed the fact that *Keller's* weighting scheme was not truly random and that if one would have used truly random weights, the R-squared would have been lower.

3.3.4 The North-South dimension – *Coe/Helpman/Hoffmaister* (1997)

After *CH's* basic analysis from 1995 it did not take much time until *Coe/Helpman/ Hoffmaister* (1997) began to focus on R&D spillovers from OECD countries to developing countries. They tried to get some insight about how developing countries can benefit from research activities done in the Northern hemisphere. Due to a lack of data, domestic R&D activities were not included as a regressor in the equation. This was assumed not to be a great problem since the size of R&D expenditures in developing countries is negligible. Foreign knowledge stocks were computed in the same way as in *CH* (import-share-weighted averages). The authors also added data on school enrollment as a proxy for human capital to their equation since they expected it to be an important determinant of total factor productivity growth. They included school enrollment not only as a control variable in their equation, but also tried to account for the role human capital plays in enabling a recipient's population to really acquire the knowledge that are included in imported products. This facet of human capital is summarized under the term "absorptive capacity" and is expressed by an interaction term between the foreign knowledge stock and school enrollment in the estimated equation.

Using a change rather than a level specification in their econometric analysis they arrive at the result that (as within OECD countries) international R&D spillovers embodied in foreign trade substantially increase total factor productivity in developing countries.

3.3.5 Focus on capital goods trade – *Xu/Wang* (1999)

Xu/Wang (1999) suggested that when constructing the import-share-weighted average foreign knowledge stock of a recipient country one should not use total import data but concentrate on the imports of capital goods instead, which are more likely to embody advanced technology that can benefit manufacturing productivity. Focusing on the same countries and time periods as *CH* (1995) the econometric results they get prove them right in that the capital-goods-import-spillover is higher both in value and significance than the simple total-import-spillover measured by *CH*. Moreover, the impact of non-capital-goods-trade is not significantly different from zero indicating indeed that most knowledge is contained in capital goods.

3.3.6 Interpretations of different weighting schemes – *Falvey/Foster/Greenaway* (2002)

Investigating R&D spillovers from the G5 (France, Germany, Japan, UK, USA) to 52 developing countries *Falvey/Foster/Greenaway* (2002, *FFG* from now on) shed light on a new interpretation of the different foreign-knowledge-stock measures proposed by *CH* and *LP*. The authors argue that whether to apply the *CH* or the *LP* measure depends on whether one considers knowledge to be public or private in the donor country and the recipient country, respectively (*FFG*, 2002).

In the *CH* measure the imports from the donor to the recipient m_{ji} are divided by total imports m_i from the recipient indicating that knowledge is private in the recipient country, whereas it is public in the donor country.

$$S_i^f = \Sigma_j (m_{ji} / m_i) S_j^d$$

Contrary to this the ratio of imports from the donor to the recipients m_{ji} and output of the donor y_j in the *LP* measure indicates knowledge to be a private good in the donor country and a public good in the recipient country.

$$S_i^f = \Sigma_j (m_{ji} / y_j) S_j^d$$

If one divided imports m_{ji} by total imports m_i and a donor's output y_j, one would consider knowledge to be private in both countries, whereas an approach not dividing m_{ji} by anything would assume knowledge to be public in both countries.

This suggestion of interpreting the different weighting schemes for the foreign knowledge stock gives *FFG* the possibility to run their regression with all four measures and to check whether all coefficients are significantly different from zero or whether one has to make some assumptions on the nature of knowledge so as to get significant results. Performing this procedure they come to the conclusion that R&D spillovers are surely significant if the knowledge spillover is a public good in the recipient country, where knowledge's nature in the donor country is of no importance. Besides, they find out that “there is only weak evidence that knowledge spillovers affect growth” (*FFG*, 2002, 666), if one considers the knowledge spillover to be private in the recipient country.

3.3.7 Dealing with “Indirect” Spillovers – *Lumenga-Neso/Olarreaga/Schiff* (2001)

The view presented by *Lumenga-Neso/Olarreaga/Schiff* (2001) emphasizes that when defining foreign knowledge stocks, what really matters is not the knowledge *produced* by each donor country but the knowledge *available* in each donor country. Available knowledge will exceed produced knowledge since it also takes into account the knowledge spillovers received by the donor country itself: Imports from country A to country B will increase B's knowledge stock and when country B exports to country C, C will receive a so-called “indirect” spillover from country A, increasing C's knowledge stock.

Therefore the foreign-knowledge-stock measures used so far might overestimate the effect a specific trade pattern has on productivity by overestimating the differences in knowledge stocks in the different donor economies. In reality these differences might be much smaller since trade between the donor countries helps their knowledge stocks to converge in the long run. They end up with the result that as expected “indirect” knowledge spillovers are substantial. In fact they are on average two to three times as large as “direct” spillovers. Besides, total (“direct” plus “indirect”) spillovers are more stable than “direct” spillovers alone and therefore a country's specific trade pattern has a weaker influence on its growth path than suggested by earlier studies. Yet this does not mean that trade does not matter for international R&D diffusion: The authors also come to the conclusion that, while the specific trade pattern loses some of its importance, the evidence on trade as a whole being crucial to technology diffusion gets stronger once “indirect” spillovers are included. This relates to the results of *Keller* (1998) when using average shares being some measure of overall openness, while the true import shares are not included.

3.3.8 Including different control variables

When considering the inclusion of other explanatory variables in the regression equation, one has to distinguish between two different ways this can be done: First, one can simply add this new variable as a control variable to the equation, which will maybe lower the coefficient of the foreign knowledge stock and correct for some model misspecification bias. Yet by doing so, one will not find out much on how the knowledge spillover and this additional regressor are inter-related and how they work together with respect to a rise in economic output. This however can be done using the second way of implementing a new variable into the equation, which is to interact the knowledge spillover with the additional regressor and that way try to reveal the effect that this regressor has on the coefficient of foreign knowledge, i.e. the elasticity of output with respect to foreign knowledge. In what follows I describe some of the literature that considers alternative regressors as control variables and/or as interaction terms together with the foreign knowledge stock:

3.3.8.1 Human Capital

Human capital as a pure control variable in the estimated equation was introduced by *Coe/Helpman/Hoffmaister* (1997, *CHH* from now on), *Engelbrecht* (1997) and *FFG (2005)* amongst others. This did not change the significance of the estimated coefficient of the foreign knowledge stock but led to an expected decrease in its point estimate.

Going further and exploring the inter-relationship between human capital and foreign knowledge one might expect a higher level of human capital to increase the return of the foreign knowledge spillover, which is referred to as “absorptive capacity” and which could be identified by a significant interaction term between human capital and foreign knowledge. *CHH* (1997) find such a significant coefficient in some of their estimation specifications, but generally not in their preferred specifications. *FFG* (2005) also find an insignificant effect of the simple interaction between human capital and foreign knowledge stocks, but they find an effect of education on the return of foreign knowledge when using a threshold model and therefore conclude that a higher stock of human capital increases a country's absorptive capacity and thus its return of foreign knowledge. *Kneller* (2005) also identifies that the foreign knowledge spillover is increasing in human capital, yielding support for the view that “absorptive capacity” matters.

There are several approaches on how to measure human capital. *CHH* (1997) used the school enrollment ratio as a proxy for human capital, i.e. a flow variable, following the concept of *Lucas* (1988) and *Becker* (1964) who suggested that it is the change in human capital which is crucial for economic output since human capital can simply be regarded as another factor of production. *Benhabib/Spiegel* (1994), however, showed that human capital has no significant impact on economic output when included in the estimated equation simply as another factor of production, i.e. as a flow variable. Besides, the school enrollment ratio might not be the best choice since it only measures the proportion of the population that has acquired very low skills (reading, writing, …). These low skills are usually not expected to account for a large part of growth, especially not for a large increase in the return of foreign knowledge spillovers.

Engelbrecht (1997) and *Coe/Helpman/Hoffmaister* (2008) used the average number of school years as a proxy of human capital, which might be a more adequate measure since it also takes into account the differences in productivity between a PhD and a person who quit school after two years. Apart from that it is a stock variable and therefore follows the approach of *Nelson/Phelps* (1966) who pointed out that it is the stock of human capital that matters for a country's economic performance since human capital is no simple factor of production but is a determinant of TFP. This view is supported by the results from *Benhabib/Spiegel* (1994) who showed that human capital levels (contrary to human capital flows) do matter for a country's economic performance since they “directly influence the rate of domestically produced technological innovation” (*Benhabib/Spiegel*, 1994, 166). *FFG* (2005) and *Kneller* (2005) differ only slightly from *Engelbrecht* (1997) and *Coe/Helpman/Hoffmaister* (2008) by focusing on average years of secondary schooling in the population over 25 years old. The reason for using only the part of the population that is older than 25 is due to the fact that the authors want to include the labor force's human capital stock only. However, in developing countries people start to work earlier than in the OECD and therefore concentrating on the population over 15 years old might be a more appropriate approach.

3.3.8.2 Spillover control variables

Xu/Wang (1999) mention another pure control variable that might be important. They argue that trade is not the only channel via which knowledge may diffuse across borders. Scientific literature, international conferences, international patenting, student exchanges, foreign direct investment, and so on, will also stimulate R&D spillovers. So as to control for these non-trade-related spillovers they construct an unweighted spillover variable, which is simply the (unweighted, especially not-trade-weighted) sum of foreign knowledge stocks. Adding this variable to the regression does not change the significance of the trade-related spillover variable yet reduces the size of its coefficient. In addition, they construct a distance-weighted spillover variable contributing to the idea that the size of a non-trade-related spillover may be influenced by the distance between donor country and recipient country. The authors find that the statistical significance of the trade-related spillover variable also remains unchanged when they include this second spillover control variables but its size is significantly reduced.

3.3.8.3 Relative Backwardness

“Relative Backwardness” is an issue that is confronted amongst others by *FFG* (2005). The idea here is that the importance of how far a country is behind the economic leader for knowledge spillovers is ambiguous. On the one hand the further behind a country is compared to the economic leader the larger are potential spillovers since it can benefit more from foreign knowledge in relative terms. On the other hand it might be easier for countries that are close to the economic leader to adapt new technologies since they already have more experience with highly developed and complicated machinery. These aspects are taken into account by the inclusion of a “catch-up” control variable relating a recipient country's GDP per capita relative to US GDP per capita and the inclusion of an interaction term between this “catch-up” variable and the foreign knowledge variable. The authors however do not find a significant effect of the interaction between a country's gap towards the technological leader USA and its foreign knowledge stock. What they find is that the economic gap should be included in the equation as a control variable since it has a direct effect on economic growth. In addition, *FFG* (2005) find significant interaction terms between relative backwardness and human capital suggesting that the ability to “catch up” depends on human capital. *Engelbrecht* (1997) also comes to the conclusion that relative backwardness plays a significant role in the model when being interacted with human capital.

Dealing with the same issue *Benhabib/Spiegel* (1994) found that a country lying behind the so-called “leader nation” in technology but possessing a higher human capital stock will finally catch up and overtake the leader. Moreover, a country with the highest stock of human capital will always end up being the technological leader. In other words, they find a significant interaction term between relative backwardness and human capital and conclude, that “the human capital stock affects the speed of adoption of technology from abroad, in the spirit of *Nelson* and *Phelps* (1996)” (*Benhabib/Spiegel*, 1994, 166). These results therefore suggest that it is the stock of human capital rather than its growth rates that is of importance for economic growth.

3.3.8.4 Institutional variables

CHH (2008) were the first to include institutional variables in their estimated equation. They focused on four different institutions that had been emphasized in the theoretical literature.

The first one is the ease of doing business, an average ranking of countries taking into account the ease of starting a business, dealing with licenses, employing workers, registering property, getting credit, protecting investors, paying taxes, trading across borders, enforcing contracts, and closing a business (*World Bank,* 2007). The second one measures the quality of tertiary education, which is measured by the extent to which tertiary institutions have freedom to manage resources, to decide on the sources and structure of funding and to set objectives, and by the extent to which they are accountable, including various types of evaluation (*Oliveira Martins et al.*, 2007). The third institutional measure consists of the strength of intellectual property rights, as measured by an index of patent protection (*Park/Lippoldt, 2005*). The fourth one is the origin of legal systems in either French, German, Scandinavian, or English law.

Since time series data were not available for the ease of doing business, the authors had to assume this variable to be constant. They created three dummy variables: one for the top group of countries, one for the average group and one for the lowest group. Then they interacted these dummy variables with the foreign knowledge stock and found significant effects of this interaction term on total factor productivity. *CHH* (2008) also found positive impacts of the quality of tertiary education on the return to foreign knowledge. Concerning the intellectual property protection variable, the interaction term between patent protection and foreign knowledge stock had a significant coefficient supporting the view that stronger intellectual property rights can lead to higher economic growth also via an interaction term with the foreign knowledge stock. Finally, the authors found “evidence that countries whose legal systems are based on French and, to a lesser extent, Scandinavian law benefit less from their own and foreign R&D capital than countries whose legal origins are based on English or German law” (*CHH*, 2008, 25).

3.3.8.5 Geographic distance

Starting from the idea that geographic distance might have an impact on the return of foreign knowledge due to higher trade, higher foreign direct investment (FDI), common cultures, shared history, and so on, *Keller* (2002) finds that the productivity effects of foreign R&D decline with the distance between donor and recipient country. This view gets support from *Xu/Wang* (1999) who introduce a distance-weighted spillover variable in their equation and find its coefficient to be significantly different from zero. Also *Kneller* (2005) finds significant impacts of geographic distance on the returns of foreign knowledge.

4 Model specification and variable description

In my thesis I investigate foreign knowledge spillovers in developing countries coming from research activities in the OECD, i.e. North-South spillovers. My sample of recipient countries consists of 47 developing countries:

Algeria	Haiti	Philippines
Argentina	Honduras	Senegal
Bangladesh	India	Singapore
Bolivia	Indonesia	South Africa
Brazil	Israel	Sri Lanka
Cameroon	Jamaica	Sudan
Chile	Kenya	Syria
Colombia	Korea	Thailand
Congo (Zaire)	Malaysia	Togo
Costa Rica	Mali	Trinidad & Tobago
Dominican Republic	Mexico	Tunisia
Ecuador	Nicaragua	Uruguay
Egypt	Niger	Venezuela
El Salvador	Pakistan	Zambia
Ghana	Paraguay	Zimbabwe
Guatemala	Peru	

Since the main part of the world's R&D activity is performed in industrial economies (96% in 1990, *UNESCO*, 1993) and within the OECD the seven largest economies account for nearly all the R&D (92% in 1991, *CHH*, 1997) I decided to choose the five biggest economies (the so-called “G5”) as the donor countries for my econometric work. These are France, Germany, Japan, the United Kingdom and the United States.[2] So my thesis concentrates on international R&D spillovers from the G5 to a sample of 47 developing countries.

This work will be done using panel data due to the higher power of this approach compared to pure time series analysis (for a discussion see *Hsiao*, 1983, 1-5). Panel data allows me to implement the information concerning 47 recipient

2 *FFG* (2002) computed that these five countries performed about 90% of the real R&D expenditures of 15 OECD economies for which they had data (average 1973-1990).

countries in my regression work, whereas a pure time series analysis approach would allow me to look at only one specific country at a time. The advantages of panel data econometrics are therefore quite obvious. The time dimension covers the years between 1973 and 2003, which provides 31 observations per country since the data is annual. The cross-section dimension consists of the 47 recipient countries, yielding 1457 observations as a whole. I will proceed by describing the different variables I use in my regression work.

$ln(GDP_{it})$ is the natural logarithm of output of recipient i in period t and will be the dependent variable. Following *FFG* (2002) and contrary to *CH* (1995) I do not use total factor productivity (TFP) but aggregate output (GDP) as the dependent variable since this might avoid “the errors one might introduce in calculating TFP, and allows a more ready comparison with the majority of growth equations” (*FFG*, 2002). Due to the fact that I am interested in elasticity values with respect to the various regressors I take the natural logarithm of GDP, which allows me to interpret the coefficients as percentage changes of the dependent variable in response to a change in one of the independent variables. Data on GDP is from the World Bank's *World Development Indicators 2008* (*World Bank*, 2009a) and is given in thousands of constant US-Dollars (2000 is the reference year).

Domestic knowledge stocks of the developing countries will not be taken into account since in most of the developing countries in my sample R&D expenditures are negligible. Furthermore, a lack of data would make it impossible to include domestic knowledge stocks of the recipient countries in the regression equation. The assumption that domestic R&D expenditures in developing countries are small enough that they can be ignored is in line with *CHH* (1998) and *FFG* (2002) who argue similarly.

$ln(LP_{it})$ is the natural logarithm of the foreign knowledge stock of a recipient country i in period t computed as suggested by *LP*. It is one variation of the main variable of interest, which is foreign knowledge. The data on exports from the G5 to the recipient countries (m_{ji}) is from the OECD's STAN database. The data was in thousands of current US-Dollars and has therefore been converted into thousands of constant US-Dollars using the GDP-deflator with base year 2000 of each donor country (*World Bank*, 2009a). The five export values were divided by each donor's output so as to get the *LP*-weights (see section 3.3.2). The domestic knowledge stocks of the donor countries were computed using annual real R&D expenditures in thousands of US-Dollars (taken from OECD's ANBERD database) and the so-called “Perpetual Inventory Method” (PIM, explained in detail in Appendix A). So as to get each recipient's foreign knowledge stock LP_{it} I weighted the G5's domestic knowledge stocks with the *LP*-weights obtained above and summed up these products.

$ln(CH_{it})$ is the natural logarithm of the foreign knowledge stock of a recipient country i in period t computed as suggested by *CH*. It is the second variant of measuring foreign knowledge considered. The computation equals the one of $ln(LP_{it})$ except that the *CH*-weights are the ratio between the exports from a donor country to a recipient country (m_{ji}) and total imports coming from the G5 of that recipient country (m_i), with total output of the donor country (y_j) not taken into account.

$open\_pc_{it}$ is a measure of recipient i's openness to trade in period t. It is the ratio of total imports m_i to output y_i of a recipient country in period t times 100 (so as to get a percentage value). Output data and total import data are from the *World Bank* (2009a). I do not take the logarithm here since I would like to know the effect of a 1 percentage point increase in openness on the percentage change in output and not the elasticity.

$ln(cap_{it})$ is the natural logarithm of recipient i's capital stock in period t. It is computed using the PIM and investment data from the *World Bank* (2009a) and serves as a control variable in the equation since it is known that capital is one of the main determinants of economic output.

$ln(pop_{it})$ is the natural logarithm of recipient i's population in period t, another control variable that has to be included so as to avoid model misspecification. The data is again drawn from the *World Bank* (2009a).

$seclev\_educ_{it}$ is the percentage of people over 15 with at least one year of second level education in recipient country i in period t. It is a measure of human capital and should act as a control variable in my equation. Following the concept of *Nelson/Phelp*s (1966) and the results from *Benhabib/Spiegel* (1994) it is a stock variable of human capital, so a country's stock of knowledge rather than its growth in knowledge is assumed to be the crucial factor for economic growth. Since I am interested in the impact of a 1 percentage point increase I do not take the logarithm. I look at the part of the population being older than 15 only since I want to focus on a recipient country's labor force. Contrary to *FFG* (2005) and *Kneller* (2005) who assume the labor force to consist of the over-25-year-olds in a country, the part of people over 15 might be a better proxy for the labor force in a developing country since most people start working in these countries when reaching that age. *Barro/Lee* (2000) provide data on human capital for lots of developing countries at 5-year intervals starting in 1960. So as to get annual human development data I assumed a linear process between *Barro/Lee's* observations in order to compute the missing observations.

$avr\_schoolyears_{it}$ is the average number of school years of people over 15 in recipient country i in period t. It is another measure of human capital (and also a stock variable), of which the logarithm is not taken either so as to find the impact of an increase of one year in average schooling on output. The data also comes from *Barro/Lee* (2000) and is transformed into annual data as described above.

no_school_{it} is the percentage of over-15-year-olds that never went to school in recipient country i in period t. It is the third measure of human capital (and also a stock variable) I will use in my regression work. Since I am interested in the impact of a 1 percentage point increase I do not take the logarithm. *Barro/Lee* (2000) provide the data (approximation as before).

$catch\_pc_{it}$ is the ratio of recipient i's GDP per capita to US GDP per capita times 100 (so as to end up with a percentage value). It is the so-called “catch-up” variable and should take into account the impact of “relative backwardness” on economic output. The notion behind the use of this variable is that the gap between this country and the economic leader (i.e. the US) is important for a country's economic performance. This because a country that is far behind the economic leader might have higher growth rates due to the fact that it can borrow technology from the economic leaders, without having to develop the technology itself. Countries further behind the frontier have more available technology to exploit and therefore may enjoy higher growth rates. According to this idea the output levels of countries should converge (see also section 3.3.8.3. for a discussion).

ln(UW_{it}) is the unweighted spillover control variable of recipient country i in period t introduced by *Xu/Wang* (1999). It consists of the (unweighted) sum of all donor countries' domestic knowledge stocks in period t and should control for non-trade-related knowledge spillovers (FDI, scientific literature, …).

ln(DW_{it}) is the distance-weighted spillover control variable of recipient country i in period t introduced by *Xu/Wang* (1999). The donor countries' domestic knowledge stocks are not simply summed up but also weighted according to their distance to recipient country i. The variable is computed for country i as follows:

$$S_i^f(DW) = \Sigma_{(j=1)}^{5} (W_{ji}/W_i)\, S_j^d$$

where $W_i = \Sigma_{(j=1)}^{5} W_{ji}$

and $W_{ji} = 1/\ln(D_{ji})$

D_{ji} … distance in kilometers between the capitals of j and I

So obviously distance is not modeled in a linear way, which implies that spillovers decline relatively quickly with distance. This result has some support in the empirical literature (see for example *Keller*, 2002) The distance-weighted spillover variable can in this way control for knowledge spillovers that are not carried by imports but by other factors that depend upon geographic distance (one might think of international conferences, student exchange programs, cooperating firms, FDI, migration, and so on).

$patent_{it}$ is the value of a patent right index indicating the strength of intellectual property rights (IPRs) in recipient country i in period t. Patent rights are expected to work on the one hand by encouraging domestic innovation, and on the other through encouraging technology diffusion. Such diffusion is assumed to occur through increased trade, FDI and licensing, though both the theory and evidence of the importance of IPRs for diffusion is mixed (see *Falvey/Foster/Medemovic*, 2004). Data is reported by *Park* (2008), who created an index going from zero to five, with zero indicating no patent protection and five indicating very strong patent protection. He computes the values as an unweighted sum of five separate scores for: coverage (indicating which inventions are patentable), membership in international treaties, duration of protection, enforcement

mechanisms and certain restrictions inventors meet. In each category a country can at most achieve a value of one. Since *Park* only provides data for every fifth year starting from 1960 I assumed the patent protection index to remain constant for each period of five years so as to fill in the gaps in my dataset. This procedure seemed more convincing than assuming a linear process as with human capital data since the process of stronger patent protection mainly depends on jurisdiction and laws and is therefore not a continuous but an abrupt process.

Doing-business-data was drawn from the *World Bank* (2009b) for the year 2008. The variable is included since the easier it is in a country to run a business the higher this country's output is expected to be and therefore more entrepreneurship creating wealth will emerge. The problem is that no time series data are available since the first doing-business-report was published by the World Bank only in the 21st century. So as with *CHH* (2008) I assume that actual data on doing-business is also valid for the time period from 1973 to 2003 and I will examine if it yields a significant coefficient. To do this I create four dummy variables, of which

→ one is for countries that are of rank 1-45 (top level),
→ one is for countries that are of rank 46-90 (above median),
→ one is for countries that are of rank 91-136 (below median),
→ one is for countries that are of rank 137-181 (lowest level)

in the ranking of the *World Bank* (2009b).

I now compute various interaction terms. I am especially interested in interacting the foreign knowledge stock with some other regressor since this tells me how the return of foreign knowledge changes due to some change in other independent variables. Therefore I interacted

- the foreign knowledge stock with openness, so as to check whether countries that are more open to trade can benefit more from foreign knowledge spillovers. This reflects the idea already pointed out by *CH* (1995) and discussed in section 3.3.1 that also the level of a country's imports matters when investigating knowledge diffusion and not only the question with which country a recipient country trades.
- the foreign knowledge stock with human capital (all three measures), following the suggestions given in section 3.3.8.1 and checking whether a higher stock of human capital can raise the return to foreign knowledge by

making workers able to use a foreign technology and also to copy it.

- the foreign knowledge stock with the “catch-up”-variable, so as to find out the direction of the ambiguous impact of “relative backwardness” on the return to foreign knowledge that is also discussed in section 3.3.8.3: A country that is far behind the technological leader might have a higher return to foreign knowledge due to its relatively low economic output and due to its low domestic knowledge stock that makes foreign knowledge more important for economic growth in relative terms. However, a far behind country might also have a lower return to foreign knowledge due to this low domestic knowledge stock that might make it difficult for workers to use and copy the foreign technology.
- the foreign knowledge stock with the patent protection indexation, so as to check whether the results by *CHH* (2008) presented in section 3.3.8.4 that a high degree of intellectual property protection provides higher returns to foreign knowledge are confirmed by my dataset.
- the foreign knowledge stock with the doing-business-dummies following the suggestions brought forward by *CHH* (2008) and discussed in section 3.3.8.4 (Does a higher degree of the ease of doing business increase the return to foreign knowledge?), and, finally,
- the foreign knowledge stock with six regional dummies indicating the recipient countries' localization. This is done to see whether the return to foreign knowledge differs across different regions of the world, which would suggest that there might be cultural differences across the world's regions that are not taken into account by the other interaction terms explained above but that are crucial for the return to foreign knowledge. The regions included are:
  → East Asia,
  → South Asia,
  → Sub-Saharan Africa,
  → North Africa and Middle East,
  → South America,
  → Central America and the Caribbean.

Having done this I end up with a wide set of explanatory variables on which I can base my econometric work.

5 Data description

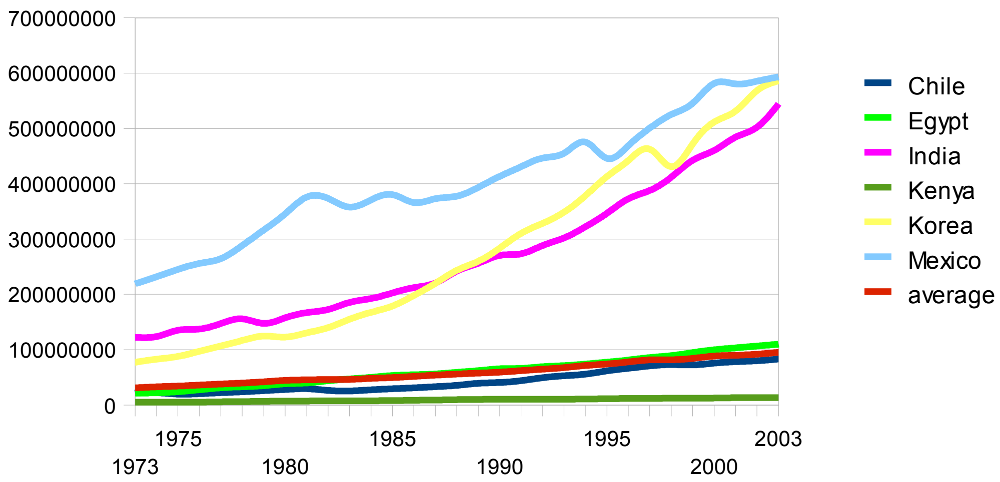
In order to provide a detailed description of the data I decided to focus on six countries, which I expect to be either representative in size and region or particularly interesting due to high experiences of economic growth for instance. Additionally I always will report the average values of the whole sample, i.e. the average of the 47 countries. Chile and Mexico are the two Latin American countries, where Chile is quite small in population and lies in the South of the continent and Mexico is a neighbor of the USA and an important importer of American products, which gave me reason to include it in the subsample. Egypt and Kenya represent the African continent, Egypt being a Muslim country in the North of the Sahara with close proximity to Europe and Kenya being the typical example of an underdeveloped small Sub-Saharan African country with huge poverty problems. India as a South-Asian and Korea as an Eastern-Asia country complete the sub-sample. India was included since it is the biggest country in my sample and Korea takes part due to its great economic growth experience that might be interesting to investigate.

5.1 GDP

The first variable considered is the dependent variable, which is real GDP. Figure 1 shows the large increase in economic output in Mexico, Korea and India, with the average growth path being quite moderate. Chile and Egypt tend to follow this average path while Kenya experienced nearly no economic growth between 1973 and 2003.

FIGURE 1

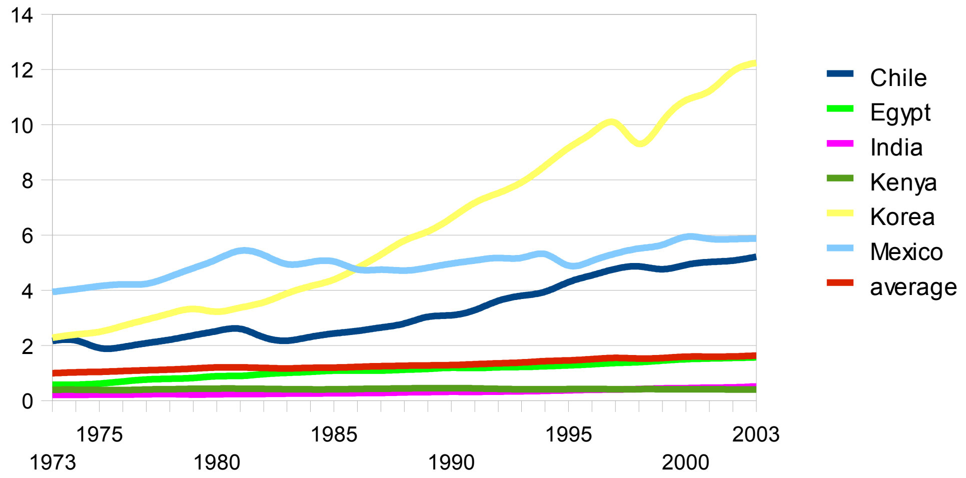
GDP in thousands of 2000 US-$



However, since GDP might rise due to high population growth (especially in India), which I take into account in my regression equation by including population as a control variable, growth of GDP per capita might be the more interesting variable to look at since it is more likely to reflect productivity growth. In figure 2 one can see that the picture now has changed: Korea and Chile are now the countries with the highest growth rate, with Chile growing at a much lower rate. Average growth of GDP per capita for all countries in the sample is low, and similar to the growth rates in India and Kenya. Egypt starting from an average level and Mexico starting from a higher level are countries that experienced moderate above-average growth.

FIGURE 2

GDP per capita in thousands of 2000-US-$



5.2 Foreign knowledge stocks

Figure 3 shows the inflows of knowledge out of the G5 carried by imports (called "foreign knowledge stock"[3]) for each year between 1973-2003. Foreign knowledge stocks are here computed with *LP's* weighting scheme discussed in sections 3.3.2 and 4. The pattern shows that annual inflow of foreign knowledge has experienced a huge increase in Mexico and Korea. This result might to some extent reflect the importance of proximity to the USA (Mexico) and Japan (Korea), two donor countries which are known to be technologically highly developed. Contrary to this, in the other four countries the foreign knowledge stock is relatively stable over all three decades, which is also the case for average values. This might be a surprise yet one has to keep in mind the structure of the *LP* measure of foreign knowledge:

$$S_i^f = \Sigma_j (m_{ji} / y_j) S_j^d$$

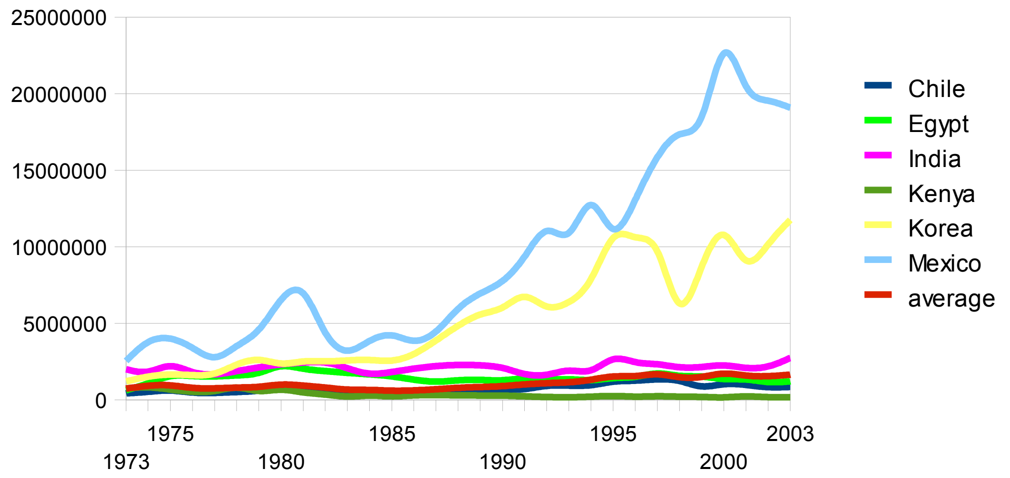
A rise in imports and a rise in the domestic knowledge stock of the donor country

3 Note that this name might be confusing. The foreign knowledge stock actually is a flow variable that measures inflow of foreign knowledge in a specific year. It does not measure the stock of knowledge at the end of a year, which would make it necessary to take the inflows of former years into account. The name comes from the fact that this flow variable is computed using the knowledge stocks of the G5 countries (and here it is really the stocks that are used).

might be outweighed by an increase of the donor country's GDP that appears in the denominator of the fraction. Therefore the foreign knowledge stocks measured by *LP* remained stable on average.

FIGURE 3

FOREIGN KNOWLEDGE STOCKS measured by LP in thousands of 2000-US-$



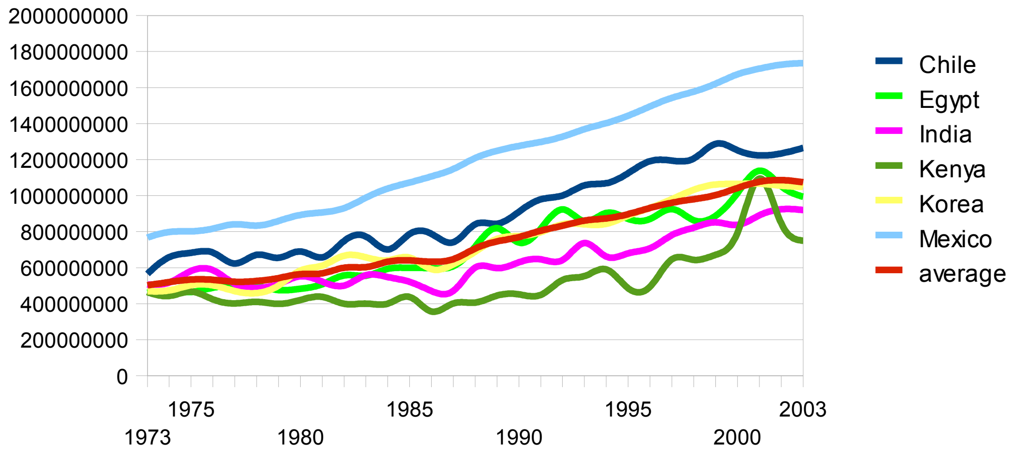
The picture looks different if we compute foreign knowledge stocks using the weighting scheme suggested by *CH* instead of the one used by *LP*:

$$S_i^f = \Sigma_j (m_{ji} / m_i) S_j^d$$

Since donor countries' GDP now no longer appears in the term one can easily identify a slow but stable increase of the average foreign knowledge stock in figure 4. All six countries of my sub-sample share this pattern, where Mexico is on the highest level over all periods followed by Chile. Egypt and Korea coincide with the average path, while India and especially Kenya remain below it.

FIGURE 4

FOREIGN KNOWLEDGE STOCKS measured by CH in thousands of 2000-US-$

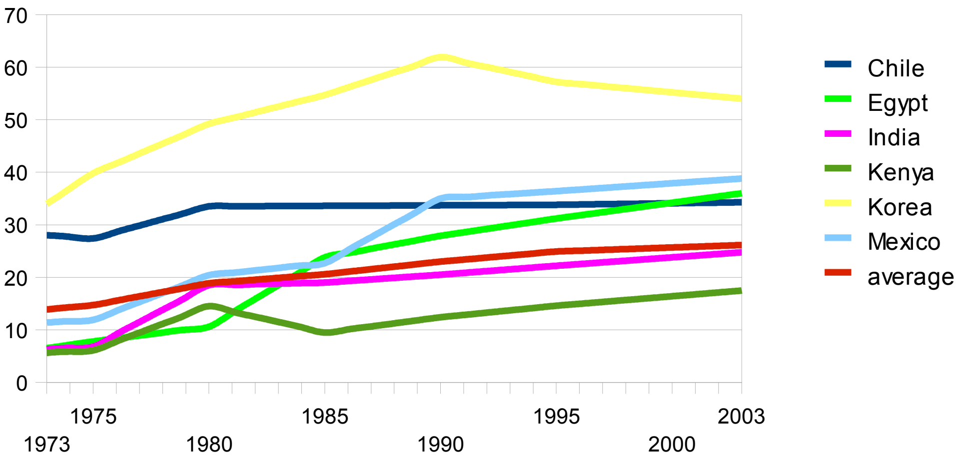


5.3 Human capital

As already mentioned human capital can be measured in different ways, of which I consider three in my empirical work. Figure 5 depicts the percentage of people over 15 years old that have some second level education. Korea experienced a high increase already between 1973 and 1990 and is from that period on an exception in the sample with values of over 50%. India, Egypt, Kenya and Mexico show increasing rates at different levels. Chile had quite high percentages throughout the whole period of over 30%, yet this value only increased slightly since 1980. I also report an average path, which is rising over time. We should be cautious about reading too much in to this however, since the computation is the simple summation procedure of the rates of all 47 countries divided by 47, and not summing all persons over 15 with second level of education in all 47 countries divided by the number of over-15-year-olds in all 47 countries. This could not be calculated due to missing data. Therefore differences in the countries' sizes are neglected in the averages reported.

FIGURE 5

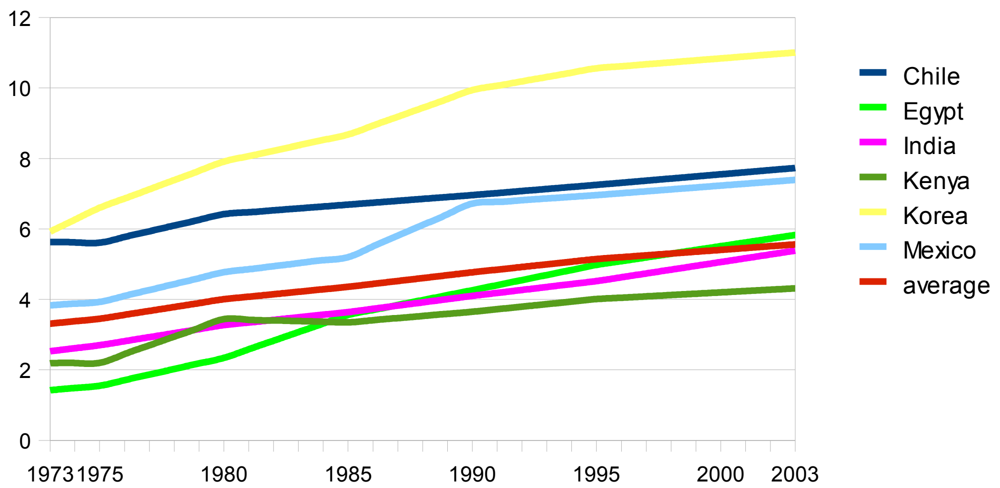
percentage of over-15-year-olds with second level education



A second way of measuring human capital is reporting the average number of school years. The results can be seen in figure 6. Korea, Mexico and Egypt, all starting from different levels in 1973 (6 years, 4 years and 2 years, respectively), succeeded in dramatically increasing the number of school years between 1973 and 2003. Also Chile, India and Kenya show a slight increase over all periods. The average path contains more information now compared to figure 5 since the measured variable is no longer a percentage but an absolute value. Yet still population size of the countries in the sample is not taken into account, which would provide average values for the whole population living in these 47 developing countries.

FIGURE 6

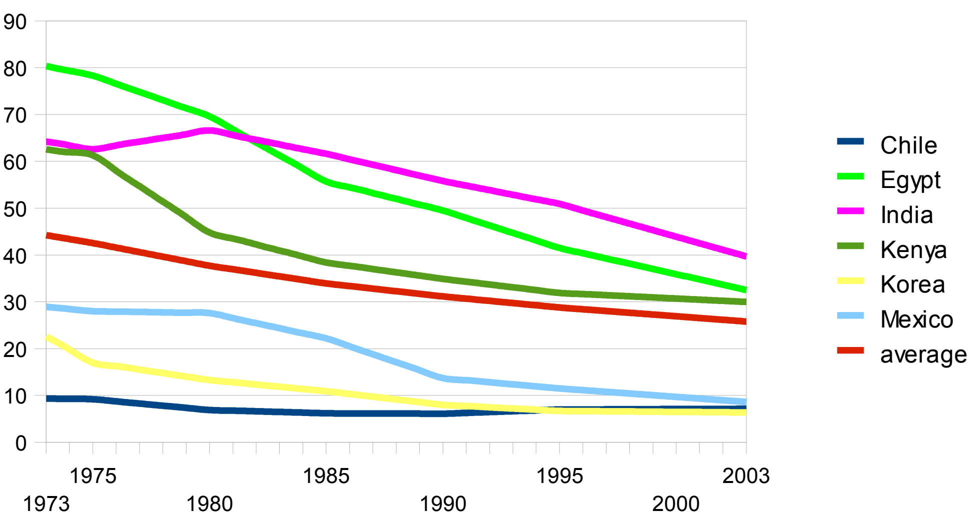
average number of school years of over 15-year-olds



Looking at the pattern of the percentage of people over 15 with no school education at all shown by figure 7 suggests the following. The rates in India, Egypt and Kenya were very high in 1973 (over 60%, even 80% in Egypt) and experienced a sharp decline over the following three decades down to 30-40%. Values in Mexico and Korea fell to 10% during this period, while Chile already had a rate under 10% in 1973. The problems with interpreting the average path is the same as before (figure 5, percentage of second level education).

FIGURE 7

percentage of over-15-year-olds with no school education

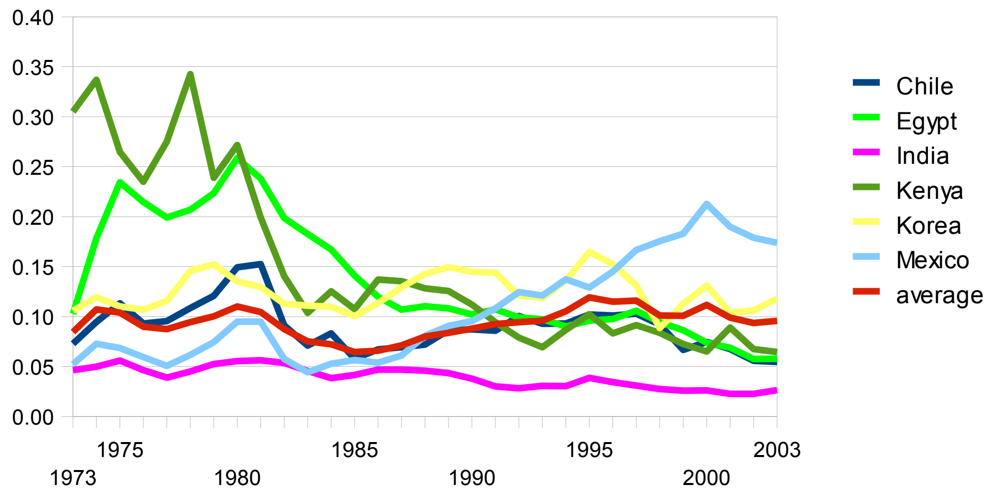


5.4 openness

Openness is measured by the ratio of total imports from the G5 to GDP. Figure 8 provides mixed results. Mexico's openness has been increasing since 1973, while Kenya's rate fell from over 30% to under 10%. The other countries show some fluctuation about a rather constant value: Korea's degree of openness always lies between 10% and 15%, Chile's between 5% and 15%, and India's under 5%. Although openness is again a variable that is reported in percentage terms, the average values now have a meaningful interpretation since they are computed as ratios of the sum of all 47 countries' total imports and the sum of all 47 countries' GDP at a given period and not as a mean across percentages as had to be done for the human capital variables.[4] The graph shows that average openness among the 47 countries in my sample has been fluctuating around 10% between 1973 and 2003. This does not necessarily mean that imports from the G5 did not rise between 1973 and 2003 yet may simply be due to the fact that GDP growth was higher than import growth. Also the share of imports coming from the G5 might have been declining during the last decades.

FIGURE 8

openness measured by imports/GDP



4 Note that this correct procedure only is possible in this context since trade data and output data are available in absolute terms, which is not the case for human capital data (*Barro/Lee* (2000) only report percentages).

6 Econometric preliminaries

6.1 Panel data estimation

In my estimation I will use panel data. Panel data consists of two dimensions, a time dimension and a cross-section dimension. This is a way to broaden a pure time series dataset by including not only one cross-section observation but several. Although this estimation approach is relatively young, “the analysis of panel or longitudinal data is the subject of one of the most active and innovative bodies of literature in econometrics” (*Greene*, 2000, 558).

Compared to pure time series analysis, panel data usually provides much more data on a topic since it uses many individuals, countries, etc. and not just one. Furthermore, causal relationships are more easy to reveal, while pure time series analysis often focuses on forecasting, thereby neglecting causalities.

“The fundamental advantage of a panel data set over a cross section is that it will allow the researcher far greater flexibility in modeling differences in behavior across individuals” (*Greene*, 2000, 559 f). Consider the following equation:

$$y_{it}=\alpha_i+\beta' x_{it}+u_{it}$$

α_i is the individual effect. It is allowed to vary over the different cross-section observations (such as countries) yet it remains constant over time. The only difference between this model and the OLS-estimator is that α_i is not fixed over all observations. However, in most cases the so-called “fixed effects estimator” is a more powerful approach assuming α_i to be a group specific constant term in the model. One simply subtracts the average from each variable (“time demeaning”) so as to end up with

$$y_{it}-\bar{y}_i=\beta'(x_{it}-\bar{x}_i)+(u_{it}-\bar{u}_i)$$

Note that α_i is differenced away. This estimator is called the fixed effects estimator since it does not neglect specific characteristics of the cross-section observations. Another name is within estimator since it measures the variation with respect to time within each cross section observation. This equation could be estimated by pooled OLS. However, the estimated coefficient would only be unbiased if there appeared no serial correlation or heteroscedasticity in the error terms u_i over time and if the regressors were all strictly exogenous (*Wooldridge*, 2002, 461 f).

Another possibility would be to use the so-called between effects estimator that does a regression on averages over time for each cross-section observation (therefore also called group-mean estimator). There are two problems with this estimation approach (*Wooldridge*, 2002, 462): First, by regressing on averages over time for each cross-section unit it neglects much information about the evolution over time. Furthermore it assumes α_i not to be correlated with the regressors, which is a rather strong assumption in our sample. Why should the country-specific effects of some developing countries not also have an impact on their foreign knowledge stock or their human capital stock? At least we cannot reject this hypotheses, which gives us good reasons not to use the between effect estimator.

Finally, the random effects model, contrary to the fixed effects estimator, views the individual specific constant terms as "randomly distributed across cross-section units" (*Greene*, 2000, 567). Since this would mean that my sample was drawn from a large population this approach does not fit my data either. Usually the random effects model is therefore used in microeconomic studies and not when the cross-section units consist of countries. Besides, this approach also assumes α_i not to be correlated with the regressors like the between effects model, which is quite unrealistic.

In my work I therefore used the fixed effects estimator (within estimator) so as to account for the variation over time. The possible bias I mentioned above will be avoided by applying the results of recent research on panel cointegration.

6.2 Unit root tests and cointegration tests in panel data

6.2.1 What is a unit root, what is cointegration?

Since panel data includes a time dimension, problems that are well-known from time series analysis also appear in a panel data setting. The first issue one has to consider when running a regression using panel data is the order of integration of all variables, which is whether a variable is stationary or not.

Definition (*Greene*, 2000, 528):
A series $y_t = y_1; y_2; \ldots ; y_T$ is said to be (…) stationary if it holds that

- $E(y_t) = 0$
- $Var(y_t) = \sigma^2$
- $Cov(y_t; y_s) = 0$ for all t not equal to s.

A non-stationary series is also said to contain a unit root. The question whether a time series is stationary or not has important consequences for the validity of regression results since the estimation of a non-stationary series with OLS may lead to spurious regression results, which means that the coefficients will be significant in the estimation although there is no real relationship (*Granger/Newbold*, 1974; see also *Greene*, 2000, 778 ff for a discussion). The reason lies in the lack of stationarity: "If two time series are each growing, for example, they may be correlated even though they are increasing for entirely different reasons and by increments that are uncorrelated" (*Banerjee et al.*, 1993, 71).

When dealing with panel data the problem is similar, which is not surprising due to the time dimension of panels. Although the OLS-estimator itself is consistent the t-statistics are not correct and therefore one cannot tell anything about inference (*Kao*, 1999; *Entorf*, 1997).

One way to handle this problem is to turn the series into a stationary one by differencing, which yields a stationary series $\Delta x_t = x_t - x_{(t-1)}$ for most economic variables. A series that turns into a stationary one after having been differenced once is called to be integrated of order one, denoted by I(1). As one might imagine, lots of information such as the information on the levels that is contained in the original series gets lost when the series is differenced. Therefore this procedure might only be a last resort when being confronted with non-stationary data.

The other thing that can be done so as to avoid spurious regression results is checking whether the time series that are expected to be correlated are cointegrated. The idea[5] behind cointegration is the following (*Greene*, 2000, 790): If two series x_t and y_t are both integrated of order one (= I(1), i.e. non-stationary) there might be a β such that the error term

$$u_t = y_t - \beta x_t$$

is I(0), i.e. stationary. This need not be the case since it could be that there is no relationship between the two series x_t and y_t and such a β does not exist. One then would expect the error term also to be integrated of order 1. However, if there is such a relationship β between the two, the error term is stationary and then the problem of spurious regressions vanishes. The regression model is now capturing the long-run relationship between the two series, i.e. there is a common trend that both series follow. Since the coefficient itself is not inconsistent in spurious regressions yet only inference values (depending on the error term, which usually is non-stationary if the series are non-stationary too) are not correct, the whole problem disappears if the error term is stationary. Then the two series x_t and y_t are said to be cointegrated. If this is the case then taking a differenced series would not be a good thing to do since by differencing the long-term relationship between x_t and y_t would not be revealed. Instead of doing this one should simply run an OLS regression in levels since no spurious result will be obtained. The critical point is now how to find out if two series are cointegrated or not.

6.2.2 Testing for unit roots in panel data

Before thinking about cointegration one must consider if all series that are suspected to be cointegrated contain a unit root. This is necessary since if both

5 A formal definition of cointegration is given in *Banerjee et al.* (1993), yet since econometric theory is not the core topic of this thesis the intuition behind cointegration seems to be more relevant.

series are stationary there is no need for cointegration since the OLS results are valid anyway and if only one series is stationary it is obviously not possible to find a common trend as described above, i.e. cointegration. Therefore a unit root in both series is necessary for them to be cointegrated. The theory on unit root tests in panel data is quite young and has developed fast during the last 15 years. I will now discuss three types of unit root tests, two of which will be used in my empirical work.

6.2.2.1 *Levin/Lin/Chu* (2002)

The first to consider the issue of panel unit root testing were *Levin/Lin* (1992), whose work was the basis for *Levin/Lin/Chu* (2002, *LLC* from now on). Their starting point was the following AR(1)[6] process, t denoting the time dimension and i denoting the cross-section dimension:

$$y_{it}=(1-\phi_i)\alpha_i+\phi_i y_{(i,t-1)}+u_{it}$$

where the y_{i0} are given and the u_{it} are i.i.d. with $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma^2$. Taking the first difference of y_{it} yields

$$\Delta y_{it}=(1-\phi_i)\alpha_i+\phi_i y_{(i,t-1)}-y_{(i,t-1)}+u_{it}=\delta_i+\beta_i y_{(i,t-1)}+u_{it}$$

where $\beta_i = (\Phi_i - 1)$ and $\delta_i = (1 - \Phi_i)\,\alpha_i$.

If $\Phi_i = 1$, the series is said to contain a unit root, since this means that the value of $y_{i,t-1}$ has a permanent effect also for values of y_i in long future, as can be seen in the upper equations. This is therefore the null hypothesis H_0 of this test: $\beta_i = 0$. The alternative is that $\beta_i < 0$ for all i, indicating that there is no unit root and the effect of a shock diminishes over time. This test can easily be performed by doing the regression and looking for inference of the coefficient β_i. A disadvantage of the test of *LLC* is that the formulation of the alternative is very strong here since it means that none of the cross-sectional time series contains a

6 The reason that an AR(1) process is the logical starting point for a unit root test lies in the properties of a unit root process: The nature of a unit root (this is also its name's origin) lies in the persistence of a shock occurring in time t and also affecting the variable in periods far behind t, being indicated by $\Phi_i = 1$, therefore being called “unit root”.

unit root. However, the framework presented above was a benchmark and gave way to the development of a variety of other test procedures.

6.2.2.2 *Im/Pesaran/Shin* (2003)

The test suggested by *Im/Pesaran/Shin* (2003, *IPS* from now on) differs from the *LLC*-test in that it allows for heterogeneity among the β_i*s*. The alternative hypothesis is now that some of them are smaller than zero, i.e. that not all cross-sectional time series contain a unit root. Rejection of the null hypothesis does therefore not necessarily mean that all cross-sectional observations are free from a unit root. So as to give an intuitive interpretation one could argue that the *IPS*-test in some way is the sum of unit root tests performed on each single cross-sectional time series.

Since both the *LLC*-test and the *IPS*-test are variations of the augmented Dickey-Fuller (ADF) test they share the problem of low power, which means that they often fail to reject the null hypothesis in cases where it actually should be rejected. These tests might therefore result in assuming too many unit roots, which is however less severe than the opposite would be, since spurious regressions only occur when ignoring a unit root and not when simply assuming one in cases where there is none.

Another point that has to be mentioned is the following. Since both tests are somehow a combination of N unit root tests on the single cross-sectional time series they both assume these single tests to be independent of each other, which implies that the residuals of each cross-sectional observation i are uncorrelated. This however might be a problem, although *Breitung/Pesaran* (2005) show that the *IPS*-test is able to capture a specific form of common effects of cross-sectional units.

6.2.2.3 *Pesaran/Smith/Yamagata's* CADF-test (2007)

Pesaran/Smith/Yamagata (2007) developed a panel unit root test procedure named “cross-sectional ADF” (CADF) taking into account that there may indeed appear common time effects between different cross-sectional observations.

Their test is another variation of the ADF-test: Lagged levels and first differences of the averages over all observations in each period $\bar{y}_t$ are added to the equation of the *LLC*-test or the *IPS*-test.

$$\Delta y_{it} = \delta_i + \beta_i y_{(i,t-1)} + c_i \bar{y}_{(t-1)} + d_i \Delta \bar{y}_t + u_{it}$$

The hypotheses correspond to those of *IPS*, allowing *some* cross-sectional units to contain a unit root also under the alternative.

6.2.3 Testing for cointegration in panel data

If unit root tests suggest that there is no unit root, i.e. the data already is stationary, there is no need for differencing or looking for cointegrated series since the usual OLS-procedure will provide meaningful and unbiased results. However, if unit root tests indicate the presence of unit roots one should acknowledge that OLS-results may be spurious. A solution would be differencing the data. Yet since lots of information on the long-term relationship between the series gets lost that way one might prefer estimating a cointegrated equation instead. There are several methods to find out if two series are cointegrated.

6.2.3.1 Using OLS and modifications

One way to test for cointegration appears straightforward: Since two series x_t and y_t are cointegrated if the error term u_t of the equation $y_t = \beta x_t + u_t$ is stationary, one could think of simply performing a unit root test (*LLC*, *IPS*, *Pesaran's* CADF, ...) on the residuals of this regression. The conclusion would be that if the null hypothesis of a unit root can be rejected, the series x_t and y_t are cointegrated. The problem is that this approach only yields efficient results if the regressors are not endogenous (*Pedroni*, 1999, 654). And it is quite likely that the regressors I will use in my thesis are not strictly exogenous but somehow inter-related to economic output. Therefore the simple OLS-procedure with a unit root test on the residuals would not be appropriate.

Two modifications of the simple OLS estimator have been invented during the last three decades. The first is the so-called Fully Modified OLS estimator (FMOLS) which was developed during the 1980s and which is asymptotically efficient since it corrects for endogenous regressors. Despite these advantages the FMOLS-estimator will not be discussed here due to its high degree of sophistication and due to the fact that the second modification has been found to perform better in empirical settings (*Kao/Chiang*, 1999).

The second modification is the so-called Dynamic OLS estimator (DOLS) that was developed by *Saikonnen* (1991). It differs from OLS in that it also includes differences of all regressors in the regression equation so as to correct for the asymptotic inefficiency of OLS. This leads to the following equation in which regressors that are not expected to be cointegrated can be included in differences so as to account for short-run dynamics:

$$y_{it}=\beta' x_{it}+\Sigma_{(k=-K)}^{K}\gamma'_k \Delta x_{(i,t+k)}+u_{it}$$

The residuals u_{it} of this estimator are then tested for a unit root and if the null hypothesis of a unit root (i.e. no cointegration → H_o: no cointegration) can be rejected there is evidence of a cointegrating relationship between y_t and x_t. The DOLS estimator does not only offer a way to test for cointegration yet it also provides unbiased, consistent and asymptotically efficient coefficient estimates under the alternative hypothesis of cointegration. Due to these attractive properties the DOLS estimator will be used in my regression work.

6.2.3.2 *Westerlund's* cointegration test (2007)

Another approach with a totally different test structure was recently brought forward by *Westerlund* (2007). He uses a conditional error correction model where x_t is a pure random walk (and therefore contains a unit root) and tests if the error correction term's coefficient α_i is significantly different from zero, which would indicate some error correction:

$$\alpha_i(L)\Delta y_{it}=\delta_{1i}+\delta_{2i}t+\alpha_i(y_{(i,t-1)}-\beta'_i x_{(i,t-1)})+u_{it}$$

“No error correction” corresponds to “no cointegration” since this would mean that a long-run relationship between x_t and y_t does not exist. Therefore rejection of the null hypothesis “no error correction” (H_o: $\alpha_i = 0$) implies rejection of the hypothesis “no cointegration”. Under the alternative hypothesis (H_A: $\alpha_i < 0$) the two series x_t and y_t are cointegrated. Test statistics are computed based on a simple OLS estimation of α_i. *Westerlund* offers four different test statistics, two of which are group mean statistics (G_α and G_τ) and two of which are panel statistics (P_α and P_τ). “The relevance of this distinction lies in the formulation of the alternative hypothesis” (*Westerlund*, 2007, 712). While the null hypothesis (H_o: $\alpha_i = 0$) is the same in both cases, the alternative hypothesis is H_A: $\alpha_i < 0$ for *all* i when computing the panel statistics P_α and P_τ, whereas it is H_A: $\alpha_i < 0$ for *some* i when computing the group mean statistics G_α and G_τ. Therefore only a rejection of the null hypothesis using the panel statistics P_α and P_τ provides evidence on that all cross-sectional observations are cointegrated.

This procedure has some advantages compared to the OLS-modifications explained above. In particular, *Westerlund's* test also works in the presence of serial correlation between the error terms and weak endogeneity of the regressors. Because of these remarkable characteristics and since there exists a ready-made ado file in STATA I will also report the *Westerlund* statistics below.

7 Empirical Results

7.1 Results of Unit Root Tests

As already pointed out above the first thing to do when dealing with panel data is to check for stationarity in the data using unit root tests. I applied the *LLC*-test and the *IPS*-test to my data[7]. The p-values are reported below. Recall that a p-value below 0.05 indicates that a test with a confidence level of 95% would reject the null of a unit root, i.e. the data is stationary, whereas a higher p-value suggests that the data is non-stationary:

LLC-test:

variable	*LLC*, 1 lag, trend	*LLC*, 2 lags, trend	*LLC*, 4 lags, trend
ln(GDP)	0.012	0.456	0.933
ln(LP)	0.002	0.795	1.000
ln(CH)	0.000	0.061	0.972
open_pc	0.002	0.440	0.998
ln(cap)	0.000	0.001	0.770
ln(pop)	0.000	0.001	1.000
catch_pc	0.000	0.131	0.546
ln(DW)	0.098	0.042	0.017
seclev_educ	0.000	0.000	0.972
avr_schoolyears	0.000	0.000	0.999
no_school	0.000	0.000	0.782
patent	0.272	0.971	1

7 There exist ready-made ado files in STATA to perform these two tests.

IPS-test:

variable	*IPS*, 1 lag, trend	*IPS*, 2 lags, trend	*IPS*, 4 lags, trend
ln(GDP)	0.676	0.905	0.925
ln(LP)	0.020	0.172	0.324
ln(CH)	0.002	0.277	0.841
open_pc	0.032	0.237	0.122
ln(cap)	0.014	0.714	0.794
ln(pop)	0.807	0.594	1.000
catch_pc	0.046	0.652	0.265
ln(DW)	1.000	1.000	1.000
seclev_educ	0.000	0.000	0.841
avr_schoolyears	0.000	0.000	0.103
no_school	0.000	0.000	0.000
patent	1.000	1.000	1.000

Obviously, whether the null hypothesis of a unit root can be rejected strongly depends on how many lags are included in the test equation discussed in section 6.2.2. Including 4 lags implies a high p-value for nearly all variables, making it impossible to reject the null hypothesis of a unit root at a reasonable confidence level. Contrarily, the null hypothesis can be rejected for many variables if only 1 lag is included in the test procedure. The intuition behind that is that the unit root test might need more information than just 1 lag so as to detect a unit root. This might be due to the fact that the logarithm of a variable increases more slowly than the variable itself.[8] Additionally, the way I computed the missing data points for the human capital variables might make it necessary to include 4 lags, a suspicion that is supported by the test results. Therefore I also performed both tests including 4 lags. The loss of degrees of freedom, which accompanies the inclusion of more lags, might not be a great problem since my cross-section contains 47 observations. This is the first reason why I will rely on the 4-lags-tests' results.

The second reason for basing my unit-root-decision on the 4-lag tests is that

8 Indeed, perfoming the tests on GDP, LP, CH, … provided higher p-values than reported above for ln(GDP), ln(LP), ln(CH), …, which is surprising but might occur for the logarithms only show very moderate changes from one period to another.

rejecting a unit root in a case where there actually is one is a more severe mistake than assuming a unit root where there actually is none. This is because the former might lead to spurious regression results, while the latter will have either no effects if cointegration is found or will at least have no effects on the validity of the results if the data is differenced. This is especially important for variables like ln(DW): The *LLC*-test rejects the null hypothesis of a unit root in all three variations of the test, whereas the *IPS*-test provides strong support for the assumption that ln(DW) is non-stationary. Therefore, to be on the safe side, I assume the whole data to be non-stationary.

In all six unit root test specifications a time trend was added to the equation. The reasoning was that stationarity might only be due to the inclusion of a time trend and if the null hypothesis of non-stationarity cannot be rejected even when a time trend is included, then the data surely must contain a unit root. That is exactly what the result looks like, since when including 4 lags, the null hypothesis of a unit root even is rejected for all variables if a trend is included.

7.2 Results of *Westerlund's* test for cointegration

After having come to the conclusion that the variables are all integrated of order one, the decision to make is how to avoid spurious regression results. Due to the fact that differencing is not the right procedure to reveal a long-term relationship between two variables, it seems more attractive to find out if the series are cointegrated.

One way to do this is to run the DOLS estimation described in section 6.2.3.1 and perform a unit root test on the residuals of this equation. If the residuals are stationary, the dependent variable and the regressors can be assumed to be cointegrated. Since this method also provides the coefficient estimations $\hat{\beta}$ for the used regressors, I will provide the results of the unit root test on the residuals together with the coefficient estimation results in section 7.3.

The second approach I decided to apply to my data are the four tests proposed by *Westerlund* (2007) that were discussed in section 6.2.3.2. The null hypothesis is of “no cointegration”, so a low p-value is needed in order to be able to reject the null hypothesis and to assume the variables to be cointegrated. The relationship between ln(GDP), which is the dependent variable, and each possible regressor is investigated separately. The computation was performed using the ready-made program in STATA developed by *Persyn/Westerlund* (2008). The p-values of these tests are reported below. When group mean statistics (G_α and G_τ) are used the alternative hypothesis is that *some* cross-sectional observations are cointegrated with ln(GDP). When panel statistics (P_α and P_τ) are used the alternative hypothesis is that *all* cross-sectional observations are cointegrated with ln(GDP).

regressor	G_τ	G_α	P_τ	P_α
ln(LP)	0.000	0.000	0.893	0.000
ln(CH)	0.000	0.000	0.929	0.000
open_pc	0.000	0.000	0.824	0.000
ln(cap)	0.000	0.000	0.000	0.045
ln(pop)	0.000	1.000	1.000	1.000
catch_pc	0.000	0.000	0.893	0.000
ln(DW)	0.000	0.998	1.000	1.000
seclev_educ	0.000	0.012	1.000	0.007
avr_schoolyears	0.000	1.000	1.000	0.706
no_school	0.000	0.023	0.994	0.021
patent	0.000	0.000	0.200	0.000

The group mean statistics tell us that for nearly all variables there are some cross-sectional units that are cointegrated with ln(GDP). Yet the panel statistics might be more interesting since they investigate if the whole cross-section sample is cointegrated. One panel statistic (P_α) indicates that all regressors except for ln(pop), ln(DW) and avr_schoolyears are cointegrated with ln(GDP), while the other panel statistic tells a completely different story: No regressor except for ln(cap) is cointegrated with ln(GDP).

Due to the fact that the α-statistics have a higher power than the τ-statistics (*Westerlund*, 2007) and due to the fact that both group mean statistics lead to the assumption of cointegration, I will assume all regressors except for ln(pop), ln(DW) and avr_schoolyears to be cointegrated with ln(GDP). Since all these three remaining regressors will enter the equation only as control variables and therefore my interest in the statistical significance of their coefficient is limited, I will use them too (so as to avoid model misspecification), although there is no evidence on them being cointegrated with the dependent variable.

The conclusion that I draw from the results of *Westerlund's* test is therefore that I will not get spurious regression results due to the non-stationarity of the data when using a level specification, which leads me to perform a DOLS estimation in levels.

7.3 Results of a DOLS estimation

The DOLS estimator was already discussed in section 6.2.3.1. It is asymptotically efficient, which is not the case for the OLS estimator due to weak endogeneity of the regressors. It does not only provide the point estimates and t-statistics of all regressors yet it is also a way of checking for cointegration since if the residuals do not contain a unit root, the regressors and the dependent variable are cointegrated. There are some features that are common for all the following estimations:

- In all regression specifications 1363 observations are used since N = 47 and T = 29. T is reduced from 31 to 29 since I use one lag and one lead in my DOLS estimator (so K = 1 in the general equation in section 6.2.3.1). However, I will not report the coefficients of the differenced regressors since they are of no special interest.
- The dependent variable always is ln(GDP).
- Furthermore, a constant is included in all specifications that was always found to be highly significant.
- All regressions are fixed effects estimations (within estimations) as already pointed out in section 6.1.
- ***, ** and * indicate that the coefficient is significantly different from zero at a confidence level of 99%, 95% and 90%, respectively.
- Standard errors are given in parantheses.
- The R-squared is reported for all regressions.
- The p-value of the unit root tests (*LLC* and *IPS*) on the residuals of each

regression is reported in each specification. They are performed using one lag (and no trend) in the test equation since now a high power of the test is important and the inclusion of more lags will lower the power of the test making it difficult to find cointegration although it might actually exist. A p-value smaller than 0.05 makes it possible to reject the null hypothesis of non-stationary residuals and therefore means that the series are cointegrated. However, if the p-value is higher than 0.05 or even 0.1 the DOLS results might be spurious.

7.3.1 Step 1 – simple regression on ln(LP), openness, spillover control

For the first step, human capital, relative backwardness and patent protection are not yet taken into account since the first goal is to find out whether the weighting scheme proposed by LP or the one suggested by CH performs better empirically. What is also investigated in the first step is the impact of openness on output (→ open_pc as a control variable) and its impact on the return to foreign knowledge (→ interaction between open_pc and the logarithm of the foreign knowledge stock: IA_lp_open_pc and IA_ch_open_pc, respectively). Knowledge spillovers not carried by imports are also taken into account by letting ln(DW) and ln(UW) enter the equation. ln(cap) and ln(pop) enter the equation as control variables. The results are reported in TABLE 1 below.

TABLE 1

regressor	1	2	3	4
ln(LP)	0.186*** (0.00836)	0.182*** (0.00842)	0.180*** (0.00839)	0.180*** (0.00838)
IA_lp_open_pc		0.00104*** (0.000298)	0.00120*** (0.000299)	0.00120*** (0.000298)
ln(cap)	0.515*** (0.00962)	0.511*** (0.00966)	0.490*** (0.0103)	0.486*** (0.0104)
ln(pop)	0.497*** (0.0198)	0.474*** (0.0208)	0.241*** (0.0465)	0.228*** (0.0457)
open_pc	-0.0118*** (0.000700)	-0.0260*** (0.00412)	-0.0282*** (0.00413)	-0.0282*** (0.00412)
ln(UW)			0.226*** (0.0423)	
ln(DW)				0.241*** (0.0423)
R-squared	0.926	0.927	0.929	0.929
LLC	0.000	0.000	0.000	0.000
IPS	0.020	0.017	0.010	0.007

Equation 1 in TABLE 1 is the simplest version of the empirical model. It only uses ln(cap), ln(pop) and open_pc as control variables and includes no interaction terms with foreign knowledge. The result is a highly significant estimate of the return of foreign knowledge of 0.186. This means that an increase of the foreign knowledge stock by 1% percent will raise output by 0.186%. The size of the coefficient might however be biased upwards since equation 1 might be too simple. Yet this problem will be faced by adding more regressors to the equation step by step. Note that both unit root tests on the residuals lead to expected low p-values indicating that the regressors and ln(GDP) are indeed cointegrated. The return to capital and population is positive and significant, while a one-percentage-point increase in openness lowers output by 0.01%. The negative sign of this coefficient might be surprising since the benefits of trade are well-known yet one must not forget that the positive knowledge spillover effects that are associated with imports are already captured by the positive coefficient of ln(LP).

Adding IA_lp_open_pc, which is the interaction term between ln(LP) and open_pc, to the equation in regression 2 does not change the significance of any other regressor but lowers the size of the foreign knowledge stock coefficient slightly from 0.186 to 0.182. The coefficient of IA_lp_open_pc itself is significant and positive, as expected. It suggests that a one-percentage-point increase in openness raises the return of foreign knowledge from 0.182 to 0.183 (= 0.182 + 0.001). Besides, an increase in openness may increase the foreign knowledge stock ln(LP) itself, as one always should keep in mind. The p-values of the *LLC*-statistic and the *IPS*-statistic remain low as in regression 1.

Regressions 3 and 4 introduce the spillover control variables ln(UW) and ln(DW) to the equation. Their coefficients are significant and positive, which is no surprise. Moreover, each of them lowers the return on foreign knowledge through imports from 0.182 to 0.180. There is virtually no difference between the unweighted control variable ln(UW) and the distance-weighted control variable ln(DW), suggesting that distance does not play a big role for knowledge spillovers that are not embodied in imports. For my further work I opted for ln(DW) as my spillover-control variable since its coefficient is slightly more significant and since the *LLC*-test and the *IPS*-test perform a little bit better in regression 4.

<u>SUMMARY of step 1:</u>

- Regression 4 is the preferred specification so far.
- The effect of ln(LP) on ln(GDP) is significant and positive (0.180).
- The same holds for the interaction between openness and ln(LP) and ln(GDP).
- Which of the spillover control variables is allowed to enter the equation does not make any difference, yet one of them should enter.
- There is strong evidence of cointegration between the regressors and ln(GDP).

7.3.2 Step 2 – simple regression on ln(CH), openness, spillover control

Now I will repeat the whole procedure of step 1, but instead of *LP's* weighting scheme for the foreign knowledge stock, I will use *CH's* one. This is done so as to find out which of the two schemes performs better empirically. Yet it would not make much sense to base one's decision between the two measures only on their statistical results since as already mentioned in section 3.3.6, the different measures of foreign knowledge stocks have different interpretations. The *LP*-measure considers knowledge to be private in the donor country (G5) and public in the recipient country, whereas it is just the other way around for the *CH*-measure. Due to the fact that intellectual property protection is much more developed in the industrialized world than in developing countries, I expect the *LP*-measure to be a more realistic one when it comes to measure knowledge spillovers.

Nevertheless I report the results one gets when using *CH's* weighting scheme in TABLE 2. The regressors used now are the same as in step 1, except for ln(LP) being replaced by ln(CH) and IA_lp_open_pc being replaced by IA_ch_open_pc.

TABLE 2

regressor	5	6	7	8
ln(CH)	0.0881*** (0.0222)	0.0715*** (0.0232)	-0.0123 (0.0292)	-0.0174 (0.0291)
IA_ch_open_pc		0.00136* (0.000730)	0.00177** (0.000734)	0.00188** (0.000735)
ln(cap)	0.591*** (0.0105)	0.589*** (0.0105)	0.570*** (0.0115)	0.566*** (0.0116)
ln(pop)	0.435*** (0.0337)	0.425*** (0.0352)	0.236*** (0.0564)	0.219*** (0.0558)
open_pc	-0.00147** (0.000605)	-0.0286* (0.0147)	-0.0369** (0.0148)	-0.0390*** (0.0148)
ln(UW)			0.284*** (0.0623)	
ln(DW)				0.307*** (0.0622)
R-squared	0.898	0.899	0.900	0.901
LLC	0.000	0.000	0.000	0.000
IPS	0.022	0.014	0.008	0.007

Comparing TABLE 1 and TABLE 2 by comparing the corresponding regressions 1 and 5 the significantly lower coefficient of ln(CH) is the only difference. The return of foreign knowledge falls from 0.186 to 0.088, if knowledge is considered to be private in the recipient country, which is not surprising. The second difference lies in the role of the spillover control variables. In regressions 7 and 8 one can identify that letting them enter into the equation makes the return on foreign knowledge carried by imports insignificant. This might indicate that ln(CH) is not able to explain more of the variation in output than the spillover control variables: The import weighted spillover variable might be of no further importance, once the unweighted or distance weighted spillover variable is included. This, however, would make ln(CH) a quite uninteresting variable.

I will therefore continue with my regression work using *LP's* weighting scheme and not the one developed by *CH*. There are plenty other reasons for this. The *CH*-measure suffers from an “aggregation bias”, as already pointed out in section 3.3.2. This bias is reduced by *LP*. Second, the assumption that knowledge is private in a recipient country but public in a donor country does not seem to be plausible in my sample. It is more likely that knowledge is private in a G5-country and public in a developing country, which is reflected by the *LP*-measure. Finally, the values of R-squared also are in favor of ln(LP).

<u>SUMMARY of step 2:</u>

- The use of ln(CH) instead of ln(LP) lowers the coefficient substantially, yet it remains significant.
- The spillover control variables ln(DW) and ln(UW) take away this significance of ln(CH), because of which they are not included into the preferred equation 6.
- ln(LP) is chosen instead of ln(CH) for the further estimation work because of the “aggregation bias” of ln(CH) and the private nature of knowledge in developing countries. Therefore regression 4 remains the preferred specification.

7.3.3 Step 3 – exploring the impact of human capital

The three different measures of human capital are now introduced to the regression equation. They are seclev_educ, avr_schoolyears and no_school[9]. Results are given in TABLE 3. There are two interesting aspects of human capital: The first one is its impact on GDP, which is identified by incorporating human capital control variables into the equation. The second is its impact on the return to foreign knowledge, which is accounted for by an interaction between human capital and ln(LP).

9 See section 2.2 so as to see how they are defined.

TABLE 3

regressor	9	10	11	12	13	14
ln(LP)	0.174*** (0.0084)	0.161*** (0.0095)	0.156*** (0.00839)	0.134*** (0.0109)	0.173*** (0.00825)	0.205*** (0.0104)
IA_lp_open_pc	0.0014*** (0.0003)	0.0012*** (0.0003)	0.0015*** (0.0003)	0.0011*** (0.0003)	0.0013*** (0.0003)	0.0010*** (0.0003)
IA_lp_seclev_educ		0.0008*** (0.0002)				
IA_lp_avr_schoolyears				0.006*** (0.00171)		
IA_lp_no_school						-0.001*** (0.0002)
ln(cap)	0.492*** (0.0105)	0.483*** (0.0109)	0.486*** (0.0101)	0.477*** (0.0104)	0.484*** (0.0101)	0.474*** (0.0103)
ln(pop)	0.237*** (0.0456)	0.261*** (0.0463)	0.223*** (0.0440)	0.259*** (0.0453)	0.0367 (0.0496)	0.0641 (0.0495)
open_pc	-0.031*** (0.00415)	-0.027*** (0.00432)	-0.031*** (0.0040)	-0.026*** (0.00423)	-0.029*** (0.00400)	-0.025*** (0.0041)
ln(DW)	0.185*** (0.0437)	0.175*** (0.0438)	0.0774* (0.0440)	0.0680 (0.0441)	0.313*** (0.0418)	0.292*** (0.0416)
seclev_educ	0.0034*** (0.00073)	-0.0065* (0.00351)				
avr_schoolyears			0.074*** (0.0073)	-0.00469 (0.0253)		
no_school					-0.005*** (0.00062)	0.008*** (0.0027)
R-squared	0.930	0.931	0.934	0.935	0.933	0.935
LLC	0.000	0.000	0.000	0.000	0.000	0.000
IPS	0.001	0.001	0.001	0.002	0.003	0.007

Using seclev_educ and avr_schoolyears as proxies for human capital yields similar results. Note at first the higher R-square in all the regressions performed in step 3 and that all specifications yield low enough LLC- and IPS-statistics so as to assume cointegration.

Regressions 9 and 11 show that including them as control variables in the equation lowers the coefficient of ln(LP) from 0.180 (regression 4) to 0.174 and 0.156, respectively, yet does not change its significance. Also the significance of all other coefficients is not altered. The impact of human capital itself on ln(GDP) is significant and positive in both cases: A one-percentage-point increase in seclev_educ raises GDP by 0.003%, while a one-year-increase in average school years raises GDP by 0.07%.

Regressions 10 and 12 investigate the impact of human capital on the return to foreign knowledge by adding IA_lp_seclev_educ and IA_lp_avr_schoolyears to the equation. This lowers the coefficient of ln(LP) again to 0.161 and 0.134, respectively. In addition, the coefficient of the human capital control variables become negative, which happens because the interaction term itself now carries the positive effects of human capital on output. The interaction terms are positive and significant in both regressions 10 and 12: A one-percentage-point increase of seclev_educ raises the return of foreign knowledge carried by imports from 0.161 to 0.1618 (= 0.161 + 0.0008). A one-year increase in avr_schoolyears raises this return from 0.134 to 0.140 (= 0.134 + 0.006).

Regressions 13 and 14 report the results when no_school is the proxy for human capital. The coefficient's signs are now opposite to those above since a lower rate of people with no school education means higher human capital. The results correspond to those of regressions 9 to 12 except for one thing: The higher return on foreign knowledge in regression 14, which is even higher than without including human capital (regression 4) is counterintuitive.

Which regression now to choose? My interest lies in estimating the return of foreign knowledge stocks in developing countries. Therefore no_school might not be the appropriate proxy for human capital due to the fact that it mainly contains information about the basic skills (reading, writing, …) that are acquired in school and these skills are not considered to have a large effect on the "absorptive capacity" of R&D spillovers. This might be the reason for the strange coefficient of ln(LP) in regression 14. The same idea lies behind the argument against avr_schoolyears: This variable assumes the return of education to be constant over all school years. However, it is assumed (see section 4) that especially the second level of education is of high importance for being able to benefit from foreign knowledge. Due to this and due to the fact that the other two proxies do not yield results that are a lot more convincing, I consider seclev_educ to be the adequate proxy for human capital in my context, which is the reason why I will use it when continuing in my estimation process.

<u>SUMMARY of step 3:</u>

- The effect of human capital on GDP is positive. Letting a human capital control variable enter the equation lowers the return of foreign knowledge substantially.
- The effect of human capital on the return of foreign knowledge stocks is positive and significant.
- My preferred specification now is regression 10 since seclev_educ is considered to be the best proxy so as to measure a country's “absorptive capacity” of foreign R&D. This due to a-priori considerations mentioned above and due to the fact that the other two proxies do not perform better.

7.3.4 Step 4 – introducing “relative backwardness”

As already pointed out it is often assumed that the degree to which a country lies behind the world's economic leader (= “relative backwardness”) has an impact on economic output as well as on the return of foreign knowledge (see sections 3.3.8.3 and 4 for a deeper discussion). Therefore I introduce a “catch-up” control variable (called catch_pc, see section 4 for a definition) and an interaction term between this “catch-up” variable and ln(LP) called IA_lp_catch_pc to regression equation 10. Doing this yields the results presented in TABLE 4.

TABLE 4

regressor	10 (preferred)	15	16	17
ln(LP)	0.161***	0.141***	0.140***	0.152***
	(0.0095)	(0.00889)	(0.00901)	(0.00883)
IA_lp_open_pc	0.0012***	-0.000657**	-0.000666**	-0.000246
	(0.0003)	(0.000297)	(0.000303)	(0.000299)
IA_lp_seclev_educ	0.0008***	0.000668***	0.000606**	0.000940***
	(0.0002)	(0.000240)	(0.000242)	(0.000240)
IA_lp_catch_pc			0.000938***	-0.00354***
			(7.03e-05)	(0.000572)
ln(cap)	0.483***	0.425***	0.437***	0.392***
	(0.0109)	(0.0109)	(0.0108)	(0.0118)
ln(pop)	0.261***	0.248***	0.265***	0.205***
	(0.0463)	(0.0428)	(0.0433)	(0.0425)
open_pc	-0.0274***	-0.000956	-0.00108	-0.00628
	(0.00432)	(0.00413)	(0.00421)	(0.00414)
ln(DW)	0.175***	0.300***	0.255***	0.422***
	(0.0438)	(0.0416)	(0.0415)	(0.0453)
seclev_educ	-0.0065*	-0.00494	-0.00402	-0.00898***
	(0.00351)	(0.00323)	(0.00327)	(0.00324)
catch_pc		0.0162***		0.0731***
		(0.00114)		(0.00934)
R-squared	0.931	0.940	0.939	0.943
LLC	0.000	0.000	0.000	0.000
IPS	0.001	0.003	0.001	0.042

Regression 10 is depicted here again in order to make it easier for the reader to compare the results. Regression 15 includes a “catch-up” control variable, regression 16 includes the interaction term IA_lp_catch_pc and regression 17 includes both of them. Note that the p-values of the *LLC* and *IPS* test again provide support for cointegration.

The point estimate of ln(LP) is again lower in all new specifications, with a minimum of 0.140 in regression 16. The R-squared is increased by the inclusion of “relative backwardness” in the equation. The coefficient on catch_up is positive and significant, meaning that the closer a country is to the USA the higher is output. This all seems plausible[10] and might lead to the conclusion that including “relative backwardness” is a good thing to do. However, some characteristics of regressions 15 to 17 are not convincing:

- The argument that a positive coefficient of catch_up indicates that the closer a country is to the US the higher its output is trivial. The “catch-up” variable itself is defined as a country's GDP divided by US-GDP so it really would be surprising if a positive correlation could not be found.
- In all three new specifications the coefficient of IA_lp_open_pc has now turned negative, which means that catch_pc now carries some information that has been carried by open_pc before. This can also be seen by the insignificance of the openness control variable (open_pc): Openness and relative backwardness are expected to correlated in the sense that a country being more open to trade will benefit from it and thereby reduce the gap between itself and the economic leader of the world.
- The same holds for seclev_educ, which is insignificant in regressions 15 and 16, becoming significant again in regression 17. So in fact, there is just one new significant variable replacing an old one. This however makes the model more complicated and might indicate that the equation is over-specified.
- In regressions 16 and 17 we now have three interaction terms of ln(LP), which makes it again a little bit more difficult to determine the exact return of foreign knowledge. And due to the fact, that one of them is negative now (IA_lp_open_pc) one may assume that some of them outweigh each other and do no longer have an own meaning. This might especially occur if the regressors are correlated, which may be the case for openness, relative backwardness and education.

Because of these arguments I will not include “relative backwardness” in my equation since it does not seem to explain a big part of economic output that is not already taken into account by regression 10.

<u>SUMMARY of step 4:</u>

- “Relative backwardness” reduces the point estimate of ln(LP).

10 Note that the dependent variable is output and not output growth. Therefore a positive coefficient is plausible and not unexpected. If a country's GDP is far behind US-GDP, then this country's GDP is low. That this conclusion is quite trivial is another part of the story.

- Its coefficient is positive, which is not surprising since GDP is the dependent variable and appears in catch_pc as well.
- The model might be over-specified when IA_lp_catch_pc is included. An interpretation gets more and more difficult due to all the interaction terms. In addition, the effects of IA_lp_catch_pc might already be captured by the other interaction terms since the coefficient of one of them turns negative now.
- Therefore "relative backwardness" will not be added to the equation and regression 10 remains the preferred one.

7.3.5 Step 5 – the role of patent protection

The idea behind including the degree of patent protection into the regression equation is that a strong enforcement of intellectual property rights might have positive effects both on economic output and on the return to foreign knowledge. The first effect may arise since foreign investors usually want to be assured that their inventions are not exploited or copied by some other firm. The second effect may be due to the fact that for a given level of imports a foreign firm will deliver a greater part of the technology to its trade partner in the recipient country if it can be sure that the trade partner can effectively patent its knowledge. As a counter argument it is brought forward that a strong patent protection might give the importing firm great market power and thereby might reduce imports and output in the long run. To test for these hypotheses I introduce a patent protection control variable (patent) and an interaction term between patent protection and ln(LP) to regression 10, the results of which are given below in TABLE 5.

TABLE 5

regressor	10	18	19	20
ln(LP)	0.161***	0.160***	0.158***	0.147***
	(0.0095)	(0.00956)	(0.00963)	(0.0106)
IA_lp_open_pc	0.0012***	0.00112***	0.00109***	0.000794**
	(0.0003)	(0.000314)	(0.000315)	(0.000328)
IA_lp_seclev_educ	0.0008***	0.000691***	0.000632**	0.000219
	(0.0002)	(0.000263)	(0.000267)	(0.000300)
IA_lp_patent			0.00102*	0.0112***
			(0.000524)	(0.00382)
ln(cap)	0.483***	0.484***	0.485***	0.487***
	(0.0109)	(0.0110)	(0.0110)	(0.0110)
ln(pop)	0.261***	0.273***	0.277***	0.289***
	(0.0463)	(0.0469)	(0.0469)	(0.0472)
open_pc	-0.0274***	-0.0269***	-0.0264***	-0.0226***
	(0.00432)	(0.00433)	(0.00434)	(0.00451)
ln(DW)	0.175***	0.142***	0.134***	0.147***
	(0.0438)	(0.0481)	(0.0479)	(0.0482)
seclev_educ	-0.0065*	-0.00555	-0.00477	0.000293
	(0.00351)	(0.00356)	(0.00360)	(0.00399)
patent		0.0111		-0.142***
		(0.00725)		(0.0528)
R-squared	0.931	0.931	0.931	0.931
LLC	0.000	0.000	0.000	0.000
IPS	0.001	0.002	0.002	0.001

Note at first that the values of R-squared, *LLC*-statistics and *IPS*-statistics are not altered by the inclusion of patent protection. We can continue to assume that the regressors and the dependent variable are cointegrated.

Letting patent protection enter the equation only as a control variable does not have a significant impact on the point estimate of ln(LP)'s coefficient. However, patent's coefficient itself is not significantly different from zero. Therefore, regression 18 is not convincing.

The coefficient of patent, however, is significantly negative when IA_lp_patent enters the equation too. It might be negative since part of the positive impacts of patent protection are now captured by the interaction term IA_lp_patent, which is quite high. Besides, the negative coefficient might reflect the fact that too much patent protection might reduce output in the short run since new innovations cannot be used in the whole economy (i.e. limits the diffusion of technology). But these are only speculations given that the model might already be over-specified in regression 20 and that many regressors might be correlated. This can also be seen by looking at the human capital variables since seclev_educ as well as IA_lp_seclev_educ are now insignificant, indicating that the addition of patent protection did not provide additional information yet simply replaced one significant variable by another: A high degree of intellectual property protection might as well as high levels of human capital simply reflect a high level of technological progress and innovation due to a high correlation between patent protection and innovation, and high human capital stocks and innovation, respectively. Regression 20 therefore is not preferred either.

Conversely, including only IA_lp_patent and not the patent control variable seems to be quite attractive. The additional interaction term is significantly different from zero at a confidence level of 10% and positive, indicating that the impact of patent protection on the trade partner's incentive to share knowledge is of greater importance than the impact of patent protection on greater market power of the importing firm and the negative impact of this market power on economic output. In addition, the inclusion of IA_lp_patent does not turn any other coefficient into an insignificant one except for the one of seclev_educ, which is no big problem since this is only a control variable, while IA_lp_patent explains the impact of patent protection on the knowledge spillover, which I am actually interested in. Adding IA_lp_patent to the equation lowers the coefficient of ln(LP) to 0.158 compared to 0.161 in regression 10. The coefficient of the interaction term itself is significantly different from zero and means that an increase of the degree of patent protection of one point[11] raises the return of foreign knowledge to 0.159 (= 0.158 + 0.001). The other interaction terms remain significant and positive. Together, these features of regression 19 lead me to abandon regression 10.

11 See section 4 for a description of how the patent protection variable is constructed.

<u>SUMMARY of step 5:</u>

- A higher level of patent protection has a positive impact on both economic output and the return on foreign knowledge.
- However, since other variables lose their significance when including patent protection as a control variable, it is only allowed to enter regression 10 as the interaction term between the foreign knowledge stock and patent protection, i.e. IA_lp_patent. Regression 19 is the new preferred equation.

7.3.6 Step 6 – the ease of doing business and regional dummies

The practical problems that appear when measuring a country's ease of doing business has been described already in section 4. The main point to keep in mind is that due to a lack of data, the ease of doing business is assumed to be constant. Therefore one might not give too much weight to the following results. I am only interested in the impact of the ease of doing business on the return to foreign knowledge, so I only include interaction terms of three of the four dummy variables (one for each level, see section 4 for details) and ln(LP). The interaction of ln(LP) with the lowest group (group 4) is omitted, so this is the one to which the coefficient of the other interaction terms refer. The results are given below in regression 21 in TABLE 6..

At last I introduce six dummy variables for each region of the world as already explained in section 4. I interact each of them with the foreign knowledge stock ln(LP) so as to measure the impact of a country's location in the world on the return of foreign knowledge and test for regional effects in regression 22 in TABLE 6. Central America and the Caribbean is the region that is not included in the equation. Therefore, the other region's coefficient show the relative effect compared to a country that lies in Central America or in the Caribbean.

TABLE 6

regressor	19 (preferred)	21 (business)	22 (regions)
ln(LP)	0.158*** (0.00963)	0.165*** (0.0108)	0.170*** (0.0156)
IA_lp_open_pc	0.00109*** (0.000315)	0.000123 (0.000363)	-0.000426 (0.000361)
IA_lp_seclev_educ	0.000632** (0.000267)	0.000886*** (0.000267)	0.000301 (0.000273)
IA_lp_patent	0.00102* (0.000524)	0.000556 (0.000509)	0.00108** (0.000510)
IA_lp_business1		0.0986*** (0.0192)	
IA_lp_business2		-0.0729*** (0.0170)	
IA_lp_business3		-0.0381*** (0.0127)	
IA_lp_east_asia			0.112*** (0.0176)
IA_lp_south_asia			0.0282 (0.0427)
IA_lp_sub_sahara_africa			-0.0128 (0.0156)
IA_lp_north_africa_middle_east			0.000497 (0.0197)
IA_lp_south_america			-0.0621*** (0.0182)
ln(cap)	0.485*** (0.0110)	0.457*** (0.0110)	0.472*** (0.0116)
ln(pop)	0.277*** (0.0469)	0.336*** (0.0460)	0.317*** (0.0475)
open_pc	-0.0264*** (0.00434)	-0.0125** (0.00499)	-0.00628 (0.00495)
ln(DW)	0.134*** (0.0479)	0.137*** (0.0466)	0.127*** (0.0479)
seclev_educ	-0.00477 (0.00360)	-0.00805** (0.00361)	-0.000998 (0.00367)
R-squared	0.931	0.937	0.936

LLC	0.000	0.001	0.000
IPS	0.002	0.016	0.002

The result of regression 21 including data on the ease of doing business is counterintuitive since only the coefficient of IA_lp_business1 is positive. This means that being in the top group of all countries has a positive impact on the return of foreign knowledge, while being in the lowest group gives the second highest return: The negative coefficients of IA_lp_business2 and IA_lp_business3 mean that the rate of return on foreign knowledge actually is lowered by climbing up the ranking in the World Bank's doing business data. Since this result does not make any sense I conclude that it is quite likely that the strong assumption of the ease of doing business being the same for the period 1973-2003 as in the year 2007 is so far away from reality that this institution variable simply cannot be used.

Considering the effects of including the regional dummies in the equation, regression 22 shows that 3 out of 5 regional interaction terms are not significantly different from zero. It has no impact on the return of foreign knowledge if a country lies in South Asia, in one of the two African regions or in Central America or the Caribbean, at least as long as factors like human capital, openness and patent protection are already taken into account. Moreover, the coefficient of ln(LP) is raised from 0.158 (regression 19) to 0.170, which I would not have expected.

The coefficient of IA_lp_east_asia is positive and significant. It indicates that the return of foreign knowledge for East Asian countries is 0.282 (= 0.170 + 0.112) and therefore substantially higher than for the four regions just mentioned (0.170). Since the coefficients of two other interaction terms (IA_lp_seclev_educ and IA_lp_open_pc) have now however lost their significance, the positive East-Asian-coefficient might simply indicate that the information that was formerly contained in these two interaction terms is now captured by the regional dummy. This may be because East Asian countries are often relatively open and have quite high levels of education compared to the other five regions.

The coefficient of IA_lp_south_america is significant and negative. It means that

South American countries' returns to foreign knowledge are on average 0.062 percentage points lower than the ones of Central American or Caribbean countries. While the latters' average return is 0.170, the formers' one is only 0.108. So we see that only two out of five regional interaction terms provide significant coefficient estimates. Besides, other regressors lose their significance once regional dummies are added to the equation. Therefore regression 22 is not preferred to regression 19.

<u>SUMMARY of step 6:</u>

- Neither the doing business data nor the regional dummy variables will be included in the equation. The first since the results of regression 21 do not make sense, which might be due to the assumption of the ease of doing business to be constant over time. The second since most of the information carried by the regional dummies is already taken into account by the other interaction terms.[12]
- Regression 19 is the preferred specification.

7.3.7 Overview over all interaction-term-specifications

The reader might already have noticed that the way the regression results were presented was based on the thinking process: I tried to incorporate one variable after the other into the equation. So as to give a different overview of the results I will report some regressions including the interaction terms of all variables below in TABLE 7. I do not report the results of the regressions using ln(CH) as the proxy for foreign knowledge and I will not deal with the equations including “doing business” or regional dummies either. This may not give any new insights but might make it easier to compare these different specifications.

12 I also ran a regression only including the East-Asian-results yet the results were not convincing either.

TABLE 7

regressor	4	10	16	19
ln(LP)	0.180*** (0.00838)	0.161*** (0.0095)	0.140*** (0.00901)	0.158*** (0.00963)
IA_lp_open_pc	0.00120*** (0.000298)	0.0012*** (0.0003)	-0.000666** (0.000303)	0.00109*** (0.000315)
IA_lp_seclev_educ		0.0008*** (0.0002)	0.000606** (0.000242)	0.000632** (0.000267)
IA_lp_catch_pc			0.000938*** (7.03e-05)	
IA_lp_patent				0.00102* (0.000524)
ln(cap)	0.486*** (0.0104)	0.483*** (0.0109)	0.437*** (0.0108)	0.485*** (0.0110)
ln(pop)	0.228*** (0.0457)	0.261*** (0.0463)	0.265*** (0.0433)	0.277*** (0.0469)
open_pc	-0.0282*** (0.00412)	-0.0274*** (0.00432)	-0.00108 (0.00421)	-0.0264*** (0.00434)
ln(DW)	0.241*** (0.0423)	0.175*** (0.0438)	0.255*** (0.0415)	0.134*** (0.0479)
seclev_educ		-0.0065* (0.00351)	-0.00402 (0.00327)	-0.00477 (0.00360)
R-squared	0.929	0.931	0.939	0.931
LLC	0.000	0.000	0.000	0.000
IPS	0.007	0.001	0.001	0.002

7.3.8 Detailed discussion of the preferred specification (regression 19)

Once again the results of regression 19 are shortly discussed since it is found to be the preferred equation for modeling international R&D spillovers carried by imports.

- The return on foreign knowledge is 0.158 and significantly different from zero. The meaning of this coefficient is that an increase of the foreign knowledge stock by 1% increases output by 0.158%. This return is substantial.

- Each percentage point of openness raises a country's return of foreign knowledge by 0.0012 percentage points. For Example, if a country's openness is 20%, its return on on foreign knowledge is 0.182 (= 0.158 + 20*0.0012). Although this might be considered to be a quite negligible effect, one should not forget that the effect of imports is already contained in the foreign knowledge stock itself at least to some extent.
- Each percentage point of population over 15 with some second level of education raises the return of foreign knowledge by 0.000632 percentage points. This means that for example a 40%-rate of seclev_educ raises the return to 0.183 (= 0.158 + 40 * 0.000632).
- Each achieved level of patent protection raises the return of foreign knowledge by 0.001 percentage points. An already above average patent protection-level of 3 therefore raises the return to 0.161 (= 0.158 + 3*0.001).
- In TABLE 8, the total returns of foreign knowledge (including the interaction terms) are given for the years 1975 and 2000 for each country. Obviously, it is not possible to identify a clear trend. In some countries, the return became lower during these 25 years, while in others it became higher. The average return decreases slightly due to a decrease in openness between 1975 and 2000 that outweighs the gains in education and patent protection (the coefficient of the openness interaction term is significantly higher than those of the education and patent protection interaction terms).

TABLE 8

	total return, 1975	total return, 2000	change in %
Algeria	0.2145	0.1892	-11.79
Argentina	0.1743	0.1851	6.19
Bangladesh	0.1785	0.1715	-3.93
Bolivia	0.1918	0.1755	-8.50
Brazil	0.1695	0.1756	3.55
Cameroon	0.1930	0.1774	-8.08
Chile	0.1910	0.1929	1.00
Colombia	0.1829	0.1862	1.81
Congo (Zaire)	0.1863	0.1742	-6.46
Costa Rica	0.1872	0.1928	2.99
Dominican Republic	0.1964	0.2017	2.72
Ecuador	0.1950	0.1877	-3.73
Egypt	0.1925	0.1904	-1.09
El Salvador	0.1751	0.1905	8.80
Ghana	0.2292	0.1945	-15.15
Guatemala	0.1764	0.1817	2.97
Haiti	0.1800	0.1917	6.52
Honduras	0.1892	0.2281	20.59
India	0.1701	0.1785	4.93
Indonesia	0.1856	0.1870	0.76
Israel	0.2080	0.2037	-2.08
Jamaica	0.2094	0.2122	1.33
Kenya	0.1950	0.1791	-8.15
Korea	0.1986	0.2128	7.15
Malaysia	0.2008	0.2226	10.86
Mali	0.1876	0.1754	-6.52
Mexico	0.1749	0.2112	20.78
Nicaragua	0.1803	0.1883	4.44
Niger	0.1925	0.1721	-10.63
Pakistan	0.1796	0.1793	-0.14
Paraguay	0.1770	0.1859	5.04
Peru	0.1845	0.1886	2.20
Philippines	0.1917	0.2207	15.13
Senegal	0.2097	0.1835	-12.50
Singapore	0.2439	0.2444	0.22
South Africa	0.2013	0.1956	-2.84
Sri Lanka	0.1947	0.2041	4.81
Sudan	0.1978	0.1703	-13.93
Syria	0.2011	0.1808	-10.11
Thailand	0.1807	0.1931	6.87
Togo	0.2295	0.1885	-17.86
Trinidad & Tobago	0.2067	0.2119	2.51
Tunisia	0.2140	0.2016	-5.80
Uruguay	0.1784	0.1893	6.10
Venezuela	0.1819	0.1893	4.07
Zambia	0.2268	0.1796	-20.83
Zimbabwe	0.1627	0.1879	15.45
average	0.1929	0.1920	-0.48

- The coefficients of the control variables ln(cap), ln(pop) and ln(DW) are all highly significant and positive as expected. Omitting one of them would increase ln(LP)'s coefficient significantly and would lead to some form of model misspecification.
- Openness as a control variable enters the equation with a significantly negative coefficient due to the fact that it is also included in the form of an interaction term with foreign knowledge, whose coefficient is positive. Besides, the positive impacts of openness on output are also captured by the foreign knowledge stock, which is computed using import data, as one should remember. Therefore, the negative sign might not have an own interpretation.
- Although the coefficient of the human capital control variable seclev_educ is only insignificantly different from zero, omitting this variable would cause the return of foreign knowledge to increase, which indicates that seclev_educ should be included in the equation as a control variable in any case.
- Note again, that the results of the cointegration tests support the assumption that the regressors and ln(GDP) are cointegrated.

8 Conclusion

In this thesis I sought evidence on the hypothesis that the R&D performance in G5-countries has some influence on the GDP in developing countries. The channel through which this R&D performance was assumed to affect a developing country's GDP is imports. A foreign knowledge stock variable was constructed according to the suggestions of *Lichtenberg/van Pottelsberghe de la Potterie* assuming that knowledge is a public good in the developing country and a private group in the G5-country.

Then regressions of the recipient countries' GDPs on their foreign knowledge stocks were run including a number of control variables and interaction terms. The specifications exploited the attractive properties of cointegrated variables, which allowed me to investigate the long-run relationship between foreign R&D efforts and domestic economic growth without being confronted with the problem of spurious regressions. The specification of the regression equation was worked through carefully and step by step, with many different regressors included in the equation so as to take into account all possible determinants of a country's economic output. Finally however, a preferred equation (regression 19) was found.

The results were clear. Both hypotheses claimed in the introductory section were supported by my empirical work. I found strong evidence of substantial international trade-related R&D spillovers in my dataset. The return of foreign knowledge that is carried by imports is 0.158, yet openness, education, and so on are not taken into account here and might also contribute to the return. The interpretation of these results is a relative one: A return of 0.158 means that an increase of the foreign knowledge stock by 1% increases output by 0.158%.

A further result of my empirical work was that the size of the return of foreign knowledge strongly depends positively on the level of openness and education of a country, but also on the degree of patent protection. Therefore, returns across countries after having taken into account openness, human capital and patent protection vary between 0.17 and 0.22 (see TABLE 8). The average return fluctuates around 0.19 over the time sample.

List of Literature

- Aghion, P. and Howitt, P., 1992, *A Model of Growth Through Creative Destruction*, Econometrica 60, 323-351.
- Banerjee, J., Dolado, J., Galbraith, J. and Hendry, D., 1993, *Co-integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press.
- Barro, R. and Lee, J., 2000, *International Data on Educational Attainment: Updates and Implications*, CID Working Paper No. 42. http://www.cid.harvard.edu/ciddata/ciddata.html
- Becker, G., 1964, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, University of Chicago Press.
- Benhabib, J. and Spiegel, M., 1994, *The Role of Human Capital in Economic Development. Evidence from Aggregate Cross-Country Data*, Journal of Monetary Economics 34, 143-173.
- Breitung, J. and Pesaran, H., 2005, *Unit Roots and Cointegration in Panels*, Discussion Paper 42, Deutsche Bundesbank (Frankfurt).
- Coe, D. and Helpman, E., 1995, *International R&D Spillovers*, European Economic Review 39, 859-887.
- Coe, D., Helpman, E. and Hoffmaister, A., 1997, *North-South R&D Spillovers*, The Economic Journal 107, 134-149.
- Coe, D., Helpman, E. and Hoffmaister, A., 2008, *International R&D Spillovers and Institutions*, IMF Working Paper WP/08/104.
- Coe, D. and Hoffmaister, A., 1999, *Are There International Spillovers Among Randomly Matched Trade Partners? A Response to Keller*, IMF Working Paper No. 99/18.
- Coe, D. and Moghadam, R., 1993, *Capital and Trade as Engines of Growth in France: An Application of Johansen's Cointegration Methodology*, IMF Staff Papers 40, 542-666.
- Engelbrecht, H.J., 1997, *International R&D Spillovers, Human Capital and Productivity in the OECD*, European Economic Review 41, 1479-1488.
- Entorf, H., 1997, *Random Walks with Drifts: Nonsense Regressions and*

Spurious Fixed-Effect Estimation, Journal of Econometrics 80/2, 287-296.

- Falvey, R., Foster, N. and Greenaway, D., 2002, *North-South Trade, Knowledge Spillovers and Growth*, Journal of Economic Integration 17/4, 650-670.
- Falvey, R., Foster, N. and Greenaway, D., 2004, *Imports, Exports, Knowledge Spillovers and Growth*, Economics Letters 85, 209-213.
- Falvey, R., Foster, N. and Greenaway, D., 2005, *Relative Backwardness or Absorptive Capacity: How Does Knowledge Spill Over Across Borders?*, Economics Letters 97, 230-234.
- Falvey, R., Foster, N. and Medemovic, O., 2004, *The Role of Intellectual Property Rights in Technology Transfer and Economic Growth: Theory and Evidence,* UNIDO Working Paper. http://www.unido.org/fileadmin/import/60030_05_IPR_rights_in_technology_transfer.pdf
- Funk, M., 2001, T*rade and International R&D Spillovers among OECD countries*, Southern Economic Journal 67/3, 725-736.
- Granger, C. and Newbold, P., 1974, *Spurious* R*egressions in Econometrics,* Journal of Econometrics 2, 111-120.
- Greene, W., 2000, *Econometric Analysis*, Prentice Hall Inc.
- Griliches, Z., 1979, *Issues in Assessing the Contribution of R&D to Productivity Growth*, Bell journal of Economics 10, 92-116.
- Griliches, Z., 1988, *Productivity Puzzles and R&D: Another Nonexplanation*, Journal of Economic Perspectives 2, 9-21.
- Grossman, G. and Helpman, E., 1991, *Innovation and Growth in the Global Economy*, MIT Press (Cambridge, MA).
- Hsiao, C., 1993, *Analysis of panel data*, Cambridge University Press.
- Im, K., Pesaran, H. and Shin, Y., 2003, *Testing for Unit Roots in Heterogenous Panels*, Journal of Econometrics 115/1, 53-74.
- Kao, C., 1999, *Spurious Regressions and Residual-based Tests for Cointegration in Panel Data*, Journal of Econometrics 90, 1-44.
- Kao, C. and Chiang, M., 1999, *On the Estimation and Inference of a Cointegrated Regression in Panel Data*, Working Paper No. 2, Center for Policy Research of the Maxwell School of Citizenship and Public Affairs, Syracuse University,

(http://www-cpr.maxwell.syr.edu/cprwps/pdf/wp2.pdf).

- Kao, C., Chiang, M. and Chen, B., 1999, *International R&D Spillovers: an Application of Estimation and Inference in Panel Cointegration*, Oxford Bulletin of Economics and Statistics 1999 (Special Issue), 691-709.
- Keller, W., 1998, *Are International R&D Spillovers Trade-Related? Analyzing Spillovers Among Randomly Matched Trade Partners*, European Economic Review 42, 1469-1481.
- Keller, W., 2002, *Geographic Localization of International Technology Diffusion*, American Economic Review 92/1, 120-142.
- Kneller, R., 2005, *Frontier Technology, Absorptive capacity and Distance*, Oxford Bulletin of Economics and Statistics 67/1, 1-23.
- Levin, A. and Lin, C., 1992, *Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties*, University of California at San Diego, Economics Working Paper Series 92/23.
- Levin, A., Lin, C. and Chu, C., 2002, *Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties*, Journal of Econometrics 108/1, 1-24.
- Lichtenberg, F. and van Pottelsberghe de la Potterie, B., 1998, *International R&D Spillovers: A comment*, European Economic Review 42, 1483-1491.
- Lucas, R., 1988, *On the Mechanics of Economic Development*, Journal of Monetary Economics 22, 3-42.
- Lumenga-Neso, O., Olarreaga, M. and Schiff, M., *On “Indirect” Trade-Related R&D Spillovers*, European Economic Review 49/7, 1785-1798.
- Nelson, R. and Phelps, E., 1966, *Investment in Humans, Technological Diffusion, and Economic Growth*, American Economic Review 56, 69-75.
- Oliveira Martins, J., Boarini, R., Strauss, H., de la Maisonneuve, C. and Saadi, C., 2007, *The Policy Determinants of Investment in Tertiary Education*, OECD Economics Working Paper, May 2007.
- Park, W., 2008, *International Patent Protection: 1960-2005*, Research Policy 37/4, 761-766.
- Park, W. and Lippoldt D., 2005, *International Licensing and the Strengthening of Intellectual Property Rights in Developing Countries During the 1990s*, OECD Economic Studies 40, 7-48.
- Pedroni, P., 1999, *Critical Values for Cointegration Tests in*

Heterogeneous Panels with Multiple Regressors, Oxford Bulletin of Economics and Statistics 1999 (Special Issue), 653-670.

- Persyn, D. and Westerlund, J., 2008, *Error Correction Based Cointegration Tests for Panel Data*, Stata Journal, forthcoming.
- Pesaran, H. and Smith, V. and Yamagata, T., 2007, *Panel Unit Root Tests in the Presence of a Multifactor Error Structure*, Cambridge Working Paper in Economics, No. 0775.
- Romer, P., 1990, *Endogenous Technological Change*, Journal of Political Economy 98, 71-102.
- Saikkonen, P., 1991, Asymptotic Efficient Estimation of *Cointegration Regressions*, Econometric Theory 7, 1-21.
- Schumpeter, J., 1911, *Theorie der wirtschaftlichen Entwicklung: eine Untersuchung über Unternehmergewinn, Kapital, Kredit, Zins und den Konjunkturzyklus*, 9. Auflage (1997), Duncker & Humblot, Berlin.
- Solow, R. M., 1956, *A Contribution to the Theory of Economic Growth*, Quarterly Journal of Economics 70, 1956, 65–94.
- UNESCO, 1993, *Statistical Yearbook*.
- Westerlund, J., 2007, *Testing for Error Correction in Panel Data*, Oxford Bulletin of Economics and Statistics 69, 709-748. (http://edocs.ub.unimaas.nl/loader/file.asp?id=1212)
- Wooldridge, J., 2002, *Econometric Analysis Of Cross Section and Panel Data*, MIT Press (Cambridge, MA).
- World Bank, 2008, *Doing Business 2007*.
- World Bank, 2009a, *World Development Indicators 2008*.
- World Bank, 2009b, *Doing Business 2008*.
- Xu, B. and Wang, J., *Capital Goods Trade and R&D Spillovers in the OECD*, Canadian Journal of Economics 32/5, 1258-1274.

Appendix A – "Perpetual Inventory Method"

Here I discuss briefly the "Perpetual Inventory Method" (PIM) that was used to compute capital stocks and the donor countries' knowledge stocks. I will explain the PIM using the example of knowledge stocks. The basic equation that allows me to compute the knowledge stocks S_t for every period given the changes R_t (R&D expenditures) every period is the following, where δ is the depreciation rate that was assumed to be 5 percent (as in *CH*):

$$S_t = (1-\delta) S_{(t-1)} + R_{(t-1)}$$

With this formula one can easily compute the knowledge stocks for all years given the values of *R* for every period and given the initial knowledge stock (for t = 0). Since S_0 is not given I used the following approach suggested by Griliches (1979) to compute it, where *g* is the average annual logarithmic growth of R_t, i.e. R&D expenditures:

$$S_0 = R_0 / (g+\delta)$$

This provides the possibility to compute the knowledge stock for each donor country for each year and therefore to compute the foreign knowledge stocks of the recipient countries. The capital stocks of the recipient countries are computed using the same procedure.

Appendix B – Abstract in English

n my thesis I investigate to what extent there exist knowledge spillovers from the five biggest economies in the world called the G5 (France, Germany, Japan, United Kingdom, USA) to 47 developing countries through imports and can therefore be considered to be “trade-related”. I develop an empirical model that builds upon the work provided by *Coe/Helpman* (1995) and *Lichtenberg/van Pottelsberghe de la Potterie* (1998) to model international trade-related knowledge spillovers from the Northern to the Southern hemisphere. To do this I compute so-called “foreign knowledge stocks”, i.e. a variable that combines the import structure of a developing country and the knowledge stocks of its trade partners from the G5. These “foreign knowledge stocks” then measure the size of the foreign knowledge spillover and are regressed on economic output of the developing countries. A number of control variables are added to the estimated equation and model specification to examine the robustness of the results. I also concentrate on the role that human capital plays when it comes to absorbing foreign knowledge that enters a developing country through imports by including various interaction terms. The dataset is a panel and consists of 47 developing countries and covers the years 1973 to 2003. In the regression work I exploit the attractive characteristics of cointegrated equations for the evidence on cointegration is convincing. I find as a result that R&D spillovers from the G5 to developing countries are substantial and that they contribute to the countries' economic performance. Furthermore empirical evidence suggests that the return to foreign knowledge increases in human capital, which means that a population with high skills is more able to benefit from foreign knowledge because of their higher “absorptive capacity”.

Appendix C – Abstract in German

In meiner Diplomarbeit untersuche ich das Ausmaß externer Effekte von Forschung und Entwicklung in den fünf größten Volkswirtschaften der Welt, den sogenannten G5 (Frankreich, Deutschland, Japan, Vereinigtes Königreich, USA), auf 47 Entwicklungsländer, die mit Hilfe von Importen verbreitet werden und insofern "trade-related", also mit Handel im Zusammenhang stehend sind. Ein empirisches Modell wird entwickelt, dass insbesondere auf der Arbeit von *Coe/Helpman* (1995) und *Lichtenberg/van Pottelsberghe de la Potterie* (1998) aufbaut und versucht, internationale handelsbezogene externe Effekte von Forschung und Entwicklung von der nördlichen auf die südliche Hemisphäre zu modellieren. Um dies zu bewerkstelligen werden sogenannte "foreign knowledge stocks" berechnet, also Variablen, die die Importstruktur eines Entwicklungslandes und den Forschungsstand seiner Handelspartner aus den G5 miteinander verknüpfen. Diese "foreign knowledge stocks" messen nun die Größe des externen Effektes und werden auf die Wirtschaftsleistung der Entwicklungsländer regressiert. Mehrere Kontrollvariablen werden der geschätzten Gleichung beigefügt, das Modell wird sorgfältig spezifiziert. Auch auf die Rolle, die Humankapital bei der Aufnahme dieser externen Effekte spielt, wird gesondert eingegangen. Die Daten sind ein sogenanntes "panel" und enthält 47 Querschnittsbeobachtungen (Entwicklungsländer) und die Zeitperioden von 1973 bis 2003. Da verschiedene Teststatistiken dafür sprechen, werden cointegrierte Gleichungen geschätzt. Das gefundene Resultat zeigt, dass externe Forschungseffekte von den G5 auf Entwicklungsländer bemerkenswert und nicht zu vernachlässigen sind und einen beachtlichen Beitrag zur Wirtschaftsleistung leisten. Auch wird heraus gearbeitet, wie wichtig Humankapital in Entwicklungsländern ist, um von fremden Forschungsergebnissen wirklich profitieren zu können.

Appendix D – curriculum vitae in German

- Name: Wendelin MORITZ
- Geburt: 16.3.1987 in Rum bei Innsbruck, Österreich
- Absolvierung der Reifeprüfung im Juni 2005 am Bundesrealgymnasium Adolf-Pichler-Platz in Innsbruck
- Ab Herbst 2005 Studium der Volkswirtschaftslehre und der Rechtswissenschaften an der Universität Wien