Scale Effects and Labor Productivity

Jürgen Antony

Scale Effects and Labor Productivity

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Wirtschaftswissenschaftlichen Fakultät

der Universität Augsburg

vorgelegt von

Jürgen Antony

Augsburg, im Mai 2006

Scale Effects and Labor Productivity

Erstgutachter:	Prof. Dr. Alfred Maußner
Zweitgutachter:	Prof. Dr. Peter Michaelis
Vorsitzender der mündlichen Prüfung:	PD Dr. Andreas Pyka
Tag der mündlichen Prüfung:	1. August 2006

TO MY PARENTS

Contents

List of Figures

List of Tables

Chapter 1

Scale Effects in Economic Theory

Scale effects play an important role in economics. They are a key assumption in the form of increasing returns to scale in the models of the new trade theory and the new economic geography and an important result, reflecting basic assumptions, of models of the new growth theory. While the former theories try to explain trade behavior between countries or regions and clustering phenomenons of economic activity, the latter in general tries to explain the observation that labor productivity in an economy is growing steadily over the long run.

The topic of this thesis is about scale effects and their implications for productivity, where the emphasis clearly lies on labor productivity measured by per capita production. Since per capita production is an important determining factor of wages, the discussion will be about relative and absolute wages for different types of labor as well. Models will be developed that take up ideas, and hence assumptions, of all the aforementioned theories in order to elaborate on the relationship between scale and labor productivity.

This chapter looks first at some basic considerations about scale effects and their role in production. After a short review of contributions to the early economic literature, the focus will be on more recent strands of the literature as mentioned in the first paragraph. The models used will be briefly explained by looking at the economic intuition of models from trade theory, economic geography and growth theory. Some formal aspects of selected models are given since these models will be used as building stones in the different chapters of this thesis. Finally, the last section of this chapter gives an overview of the contributions of the following chapters.

1.1 Basic Considerations

Throughout this thesis productivity is meant to cover per capita production of workers in real terms. Scale and scale effects denote the extent of the relevant work force and its impact on productivity. A key contribution of this thesis will be on the correct economic definition of scale, i.e. the definition of the relevant work force. Two associated keywords used in economics are increasing returns to scale and economies of scale. The former are present if a proportionate increase in all inputs to a production technology induces a more than proportionate increase in output. Economies of scale are present if average cost of production decreases as output increases. The scale effects which are under investigation in this thesis concern labor productivity measured by per capita production. A scale effect in per capita production is present if an increase in the labor force in the relevant market induces an increase in per capita production. Note that the workforce used to normalize production to yield per capita production need not be the same as the workforce determining the scale of the relevant market. Also the discussion is restricted to aggregative production technologies which are homogenous of degree one in all usual production factors.

One important channel through which scale effects operate is technology. The source or the basic assumption behind this idea is the existence of some indivisibility in production factors, e.g. the existence of fixed costs in production in combination with otherwise constant marginal costs. This simple and basic idea leads to falling average costs per unit of output produced, i.e. the more units of one particular good is produced, the lower is the fraction of the fixed costs attributable to one unit. As an economy gets larger, measured in terms of input factors, the fixed costs distribute themselves, ceteris paribus, onto more and more factor units. If labor is a production factor then output per worker necessarily increases. However, this scale effect, where scale is defined as the amount of labor available to production, is decreasing in the size of the work force, i.e. the marginal effect of one more worker on labor productivity decreases with the size of the work force.

For the marginal contribution of one worker to labor productivity to be in some sense constant, one more idea or assumption is needed. Following only the just mentioned argument, it would be efficient to produce just one type of good, and thus to incur the fixed costs only once, but this good in a large number of units. Obviously this is not what one observes by looking at an economy. Usually fixed costs are present to some extent in production of all existing goods. But there are millions of different goods produced at a time in any economy of the world. Economists have tackled this by arguing that not only costs per unit and the implied price is important, but also that variety matters. The last point is usually formalized by an assumption guaranteeing that, directly or indirectly, utility in an economy is, ceteris paribus, larger the more distinct goods are available at the same time. Here distinct means, that goods are not perfect substitutes and hence a trade off between variety and low average costs exists. The drawback of variety is that usually fixed costs per variant have to be incurred, thus the sum of fixed costs increases if the number of distinct goods grows large. This trade off is usually modelled in an aggregation technology exhibiting imperfect substitutability between different goods produced in the economy. This aggregation technology is thus formalizing the channel through which variety operates to bring scale effects into existence. The key assumption is imperfect substitutability, i.e. there is some benefit to having different goods which runs counter to the need of incurring fixed costs for that number of goods. In the theories mentioned at the beginning of this chapter this trade off is solved by finding the combination of quantity and variety that optimizes objective functions of the economic actors in the economy. The result is that the size of the model economy is a determining factor limiting the variety dimension. This result carries over to the models used in this thesis. These two basic assumptions, fixed costs and a preference for variety, yields in equilibrium a constant proportionate effect of an additional worker on per capita production, since with the extend of the work force the possibilities for variety increases.

These effects have been recognized very early in economics as by the work of Adam Smith and others. We will look at them in the following section before exploring more recent contributions to the literature.

1.2 Early Economic Writings[1]

Adam Smith (1776) took notice in his writings of the presence of scale effects in the form of increasing returns to scale. He does so by noticing first the benefits of the *division of labor* in the often cited example of the pin factory, where the steps necessary to produce a pin are divided into different tasks. Exactly this division of labor into different steps of production is, according to Smith, the source of productivity gains. The channel through which these productivity gains are realized is due to education of the worker in his specialized task and the usability of machinery equipment in it. The sources of both, specialized education and machinery, are, according to Smith, the division of labor. Proceeding further, Adam Smith (1776) elaborates on the limits of the division of labor. He does so by noting that the reason for division is simply the possibility of goods exchange between economic agents, giving each the possibility to specialize in that field of activity where a comparative advantage exists. Since Smith sees this as the ultimate reason for the division of labor, he notes that the division might be limited by exchange possibilities which he terms the extent of the market. He summarizes this in the title of chapter III in book I: "That the Division of Labor is Limited by the Extent of the Market". He later on postulates that the extent of the market is positively influenced by the population size and density, the available natural resources, the stock of capital and transportation possibilities. To put things together, Smith claims that a large market gives many opportunities to specialize which in turn affect productivity positively. This basic idea is what can be found in many modern economic models.

Also dealing with increasing returns to scale are John Stuart Mill (1848) and Karl Marx (1867-94). Both are less concerned with the sources of increasing returns

[1] This section draws to some extent on Vassilakis (1987).

to scale but deal more with their implications for the economy. Vassilakis (1987) formulated their ideas as a proposition stating that an increasing market size leads to stronger concentration and an increase in the scale of production within the individual firm, thereby increasing efficiency. However, nothing is said about the ultimate source of the increasing returns to scale.

Alfred Marshall (1890, 1919), on the one hand, was concerned with the coexistence of increasing returns and the possibility of a market equilibrium with perfect competition, and on the other hand, with the trade off stated above between efficiency, i.e. low average costs, and the degree of diversity in goods produced in the economy. He commented on this trade off by noticing that an growing extent of the market allows for more heterogeneity in goods demanded by the market. Some of these goods are demanded in large scale, allowing for a high degree of division of labor in the production process; some of the goods are demanded in low quantities thereby making only limited use of division of labor. In general, he concludes that an increase in the diversity of goods demanded from the market reduces efficiency due to the decreasing possibility of exploiting increasing returns to scale.

1.3 New Trade Theory

Increasing returns to scale play a crucial role in the new trade theory. Especially the models based on monopolistic competition are very related to the analysis in some of the chapters to come. It was the work of Krugman (1979, 1980) to introduce this approach in international trade theory. The basic assumptions are as follows. Utility of consumers is defined over differentiated consumption goods with a taste for variety, i.e. utility increases, ceteris paribus, if the number of differentiated consumption goods increases. The basic difference between Krugman (1979) and Krugman (1980) is that in the former, the elasticity of substitution between two differentiated consumption goods is allowed to vary and is fixed at a constant in the latter. The source of increasing returns to scale in production are fixed cost which have to be incurred from the producer of the differentiated consumption good. Production afterwards takes place at constant marginal cost caused by the use of labor in production. Since consumption goods are differentiated, competition does not take place within a specific variant. New entrants in the market for consumption goods will rather produce a new variant than to compete in an existing variant. Nonzero profits that can be earned from producing consumption goods lead to market entry by new producers of new variants, inducing downward pressure on the profits of incumbents. In equilibrium, net profits equal zero. If the autarky case for one particular country is examined, the zero profit condition implies a number of differentiated consumption goods that is directly proportionate to the workforce of the economy. Allowing for frictionless trade in differentiated consumption goods between two different countries, where the difference comes from the extent of the workforce, leads to an increase of the number of varieties available in both economies, where the new number of variants is now proportionate to the sum of the workforces. In Krugman (1980), with a constant elasticity of substitution between consumption goods, an increase in the extent of the market through opening up for trade has no influence on the real wage in terms of consumption goods in the economy[2]. However, utility is increasing as the number of variants of consumption goods increases and there is love of variety.

Krugman (1981) extends the idea in Krugman (1980) by introducing a two sector economy into an international trade environment. Utility in one country is symmetrically defined over two indices composed of differentiated consumption goods with a constant elasticity of substitution (Dixit and Stiglitz 1977). Labor is sector specific and production of consumption goods takes place analogous to Krugman (1979, 1980). The important result of the model is that once free trade between two countries in differentiated consumption goods is allowed for, and both countries differ in their endowments of sector specific labor, inter- and intraindustrial trade between countries takes place. Additionally the total number of differentiated goods increases compared with the autarky situation. The reason is, as in Krugman (1979, 1980), that the larger market can cover more fixed cost and hence allows for more variety. Since the model covers two types of labor, it has something to say about

[2]This, however, depends on the price level chosen to compute the real wage; see the section on the new economic geography below for a discussion of this point.

wage inequality, however, the results are not that surprising. Under autarky the wage rate for the scarce type of labor is higher; under free trade, wages for one type of labor equalize due to equalization of prices. The effect of opening up for trade on wages is that the scarce factor happens to experience a reduction in wages and the abundant factor experiences an increase in wages. The effect on utility is however not determined and depends on the elasticity of substitution between consumption goods[3]. Krugman does not comment on effects of labor supply on wages, a question that arises in the context of studies dealing with wage inequality between different types of workers (see chapter two and three on wage inequality). The basic idea of this model is taken up in chapter two on wage inequality and trade where we will see that the Krugman model is very specific in one assumption, namely the specification of the utility function to be the sum of the logs of the consumption indices. This implies an elasticity of substitution between the two types of good of exactly one, which yields very specific results if the reaction of wages on labor supply of different types of workers is studied. Especially it precludes the possibility of increasing relative wages as a response to increasing relative supply. Exactly this happened in the US and the UK where increased supply of high skilled labor was accompanied by an increasing relative wage. In both economies some forces were at work which offset the usual expected substitution effect which would have let the relative wage decline. It will be shown in chapters two and three how changes in the assumption regarding the elasticity of substitution between two types of goods critically affects the relationship between wage inequality and labor supply.

Ethier (1982) is another important contribution to the international trade literature. He considers an environment in which a country is engaged in producing two goods, good one with a constant returns to scale technology and good two with an increasing returns to scale technology. Labor and capital are the only inputs to produce good one. But it is possible to produce another good with the same technology which serves as an input to production of intermediate input factors used in production of good two. The production function for the intermediate input factor is subject

[3] The critical threshold for this elasticity is two.

to increasing returns to scale as in Krugman (1980, 1981). Good two is produced according to a Dixit and Stiglitz (1977) index defined over differentiated intermediate input factors generalized by a factor determining the returns to differentiation. As noted by Bhagwati et al. (1998), there are thus increasing returns to scale due to fixed cost in production of intermediates and there are economies of scale due to the influence of differentiation on the production of good two, i.e. the more variants of intermediate inputs are available, the lower are marginal and average costs of producing good two. Opening up this economy for trade with another country can lead to different results. As mentioned by Bhagwati et al. (1998) there can be multiple equilibria and possibly specialization in production of good one and two between countries. If the last phenomenon does not take place, trade in intermediate input occurs and each country has access to the other countries' variants of intermediate input factors. Ceteris paribus, this makes production of good two in both economies cheaper or equivalently increases productivity.

Another way of modelling international trade in a framework of increasing returns to scale is taken by Helpman (1981). He employs Lancaster's (1979) model of the ideal variety, where consumers are distributed uniformly over potential varieties of manufacturing goods, representing their ideal variety, on a circle. These manufacturing goods are produced with increasing returns to scale. Additionally there is another good which is produced with a constant returns to scale technology. Capital and labor are the only production factors used for producing goods one and two. From this approach and the results of the models cited above, it is already clear that under autarky, the differentiation of manufacturing goods is limited by the size of the economy. This is due to the presence of increasing returns and the size of the economy is determined, due to the assumptions in this model, by the exogenous population size. Opening up the economy for trade with another country leads to intra- and interindustrial trade characterized by the differences of both economies in their factor endowments, i.e. the available amount of capital and labor. Most important, intraindustry trade in differentiated manufacturing goods takes place, thereby increasing the number of variants available in both economies. If one country is labor abundant and the constant returns to scale good is labor intensive, the labor intensive country will be a net exporter of the constant returns to scale good and a net importer of differentiated, capital intensive manufacturing goods. There are effects of trade on per capita units like wages. On the one hand, these effects are due to the price equalization that free trade induces. On the other hand, there are scale effects due to the increased number of variants of differentiated manufacturing goods. This is because the elasticity of demand for these differentiated goods is not constant with the assumed utility of consumers. Helpman (1981) assumes that this elasticity is increasing if the number of variants increases. This leads to an increase in the produced quantity of any variant of manufacturing goods and lowers prices for them. This affects clearly the real wage of workers but Helpman (1981) does not elaborate on this point. If one would assume a constant elasticity of substitution as in Krugman (1980) this last effect would disappear and there wouldn't be any scale effects in per capita figures like wages because no productivity effects associated with an increase of variants available is present. But there is of course a scale effect on utility. Because of an increased number of differentiated manufacturing goods, i.e. the circle is now more densely occupied by variants of goods, the manufacturing goods are now nearer to the ideal variety of the consumer.

To sum up this section, there are broadly three approaches using increasing returns to scale in international trade models. First, the pioneering work of Krugman (1979, 1980), secondly, Ethier's (1982) approach, and lastly, Helpman's (1981) model. All three modelling strategies rely on differentiated goods which are employed by the economy with an elasticity of substitution smaller than infinity. The degree of diversification is always determined by indivisibility of input factors, usually the existence of fixed cost in production, via a zero profit condition for producers of variants. The determining factor of this degree turns out to be the size of the economy, i.e. the country under consideration in the autarky case or in the trade case the extent of the trading countries. Both sizes are associated with the factor endowment of the economy, which is usually the endowment with labor. There is however an important difference between the Ethier (1982) model and the Krugman (1979, 1980) and the Helpman (1981) approach. Ethier (1982) was the first to model the returns due to differentiation, i.e. the economic reward for a greater number of variants of goods produced, in terms of productivity. While the degree of differentiation affects in the Krugman (1979, 1980) and the Helpman (1981) model only utility, it has an influence in Ethier (1982) on productivity in one sector of the economy. In the development process of trade models this did not play an important role because the ultimate reason for these models was to explain trade patterns, especially the presence of inter- and intraindustry trade, and not the presence of scale effects in e.g. per capita production or wages. Thus Ethier (1982) possibly unintentionally obtains this result due to the assumption that the degree of differentiation affects a production function of the economies. This additionally gives insights into the mechanism of generating scale effects, as mentioned above, in per capita units. There are two ingredients needed to achieve them: first, increasing returns to scale to allow for differentiation of goods produced, and second, a reward to this differentiation which is modelled in Ethier (1982) by a production function that gives a higher productivity as a higher degree of differentiation of input factors is present. The first ties differentiation to the extent of the underlying economy, the last makes productivity depending on it. This a direct interpretation of Adam Smith's idea of the benefits of division of labor limited by the extent of the market.

1.4 New Economic Geography

The new trade theory, summarized in the section above, was mainly designed to explain trade patterns between countries, especially intra- and interindustry trade. The models usually belonging to the new economic geography go one step further. The aim of these models is to present a unified theory of trade patterns and geographical localization of production[4]. But still they build on the same assumptions as models of the new trade theory, i.e. the existence of increasing returns to scale. However, it turns out that one additional assumption is needed in this context in

[4]The literature preceding the new economic geography can be titled by location theory. For a summary of early ideas of this literature see Krugman (1998).

order to obtain economically meaningful results. This is the assumption of trade cost, at least for some goods in the models.

Helpman and Krugman (1985, chapter 10) provide a first version of models of the new economic geography. This can be seen as an amendment of the Krugman (1979) model by additional goods which are produced with constant returns to scale or an application of the Helpman (1981) model. To yield new results, they introduce trading costs in different ways, two of them are rather extreme: (i) transport costs for some constant returns to scale industries are prohibitively high, i.e. they are non-tradable, and transport costs for the increasing returns to scale industry are negligible; (ii) transport costs for the increasing returns to scale industry are prohibitively high and negligible for the constant returns to scale industry; and (iii) the increasing returns to scale industry is faced with intermediate transport costs of the "iceberg" type (Samuelson 1954) and there are no transport cost for the constant returns to scale industry. Their subsequent analysis is based on the assumption that utility defined over the different industries is of the Cobb-Douglas form. Their findings can be summarized as follows: In the first case there are essentially no new results. There are now three goods, two produced with constant returns to scale and one of them is non-tradable. The differentiated increasing returns to scale goods can be traded at zero transport costs. The basic conclusions from the Helpman (1981) model apply. In the second case things are more complicated and Helpman and Krugman (1985) are only able to present results for a special case of the model. The assumptions are constant expenditure shares on all goods in the economy and homothetic production functions. The conclusion from this specification is that, if the trade case between two different countries is considered, the larger country, measured by total income, has both a larger number of differentiated goods and a larger quantity produced per variant. Depending on the endowments with capital and labor there might be factor price equalization, but no scale effects in real wages, in terms of goods, or per capita production. The last of the three above mentioned cases is also analyzed for a special case. The (sub-)utility function for differentiated manufacturing goods is assumed to be of the Dixit and Stiglitz (1977) form and is aggregated together with consumption of the constant returns to scale good according to a Cobb-Douglas utility function. Constant returns to scale goods are traded freely and differentiated manufacturing goods are traded at "iceberg" transport costs. With this set-up, specialization in production is a possible outcome, i.e. it may happen that one country specializes completely in producing constant returns to scale goods. If this is the case, the number of variants produced in the other country depends among other things on the total size of both economies, measured by their joint supply of labor. If total specialization does not prevail, the number of variants produced in each economy depends positively on its supply of labor and negatively on the supply of labor in the other economy. However, there are no scale effects on real wages in terms of goods or per capita production since, first, the elasticity of substitution between differentiated products is constant as in Krugman (1980), and second, there are no productivity gains from an increased specialization through a rising number of variants. There is, as in the other utility based models cited above, a scale effect on utility since there is love of variety and variety depends on the sizes of the labor forces.

An important contribution to the field of economic geography is Krugman (1991). He analyses two geographical aspects, first, the location of production, and second, taking mobility of factors into account, the localization of labor. He uses a two sector model in which two types of goods are produced. There is one homogenous good produced with constant returns to scale and there are in the second sector many differentiated goods produced with increasing returns to scale as in Krugman (1979, 1980). Labor is heterogenous, i.e. one type is employed in the constant returns to scale sector and the other in the differentiated sector. Utility of the consumers is given by a Cobb-Douglas utility function aggregating constant returns to scale goods and a consumption index of the Dixit and Stiglitz (1977) form of differentiated goods. Thus the elasticity of substitution between differentiated goods is constant. Both sectors use as the only production factor sector specific labor; full employment is assumed. The autarky equilibrium does not give any surprising new results, the degree of differentiation is determined by the extent of the work force employable in the increasing returns to scale sector. In the next step Krugman (1991) opens up two economies, differing with respect to the work force in the differentiated goods sector, to trade in differentiated and constant returns to scale goods. As in one version of the Helpman and Krugman (1985) model, he assumes no trade costs in constant returns to scale goods, but "iceberg" costs in the case of differentiated goods. Regarding the trade patterns, the outcome of the model is inter- and intraindustry trade. Each economy exports its country specific variants of the differentiated goods and imports the variants produced in the other country. The number of variants produced in each economy is proportionate to the extent of the work force of the differentiated sector. Focusing on per capita figures as wages, it is clear that wage equalization in the constant returns to scale sector takes place due to price equalization. Wages in the differentiated sector however differ because of the presence of transport costs. Note that there, as in the models cited so far, with exception of the Ethier (1982) model, no productivity effects of the degree of differentiation, there is thus no direct scale effect in real wages measured in terms of goods. There is however an influence of the size of the work force in the differentiated sector on wages. The larger the market, the more variants can be produced in the economy and relatively less variants are being imported. This has a direct influence on the consumption price index, the relatively larger number of own variants of differentiated goods makes trade costs relatively less important, lowering the consumption price index and raising the real wage for consumers in the economy. Note that this is only due to the presence of trade costs. If these are zero, the effect disappears and the market size, measured by the extend of the work force, has no influence on the real wage[5].

To elaborate on this point a little bit more consider the following simple model. Let utility of a representative consumer in an economy be given by

$$U = \left(\int_0^N c_i^\rho di \right)^{\frac{1}{\rho}}, \tag{1.1}$$

where c_i is consumption of one variant of the differentiated consumption goods in the

[5] In this case the distinction between the two economies is obsolete because the trade equilibrium will establish the equilibrium of the fully integrated economies with full price and wage equalization.

economy. ρ determines the elasticity of substitution between the different variants which is assumed to be larger than one. The mass of the set of variants is N. The different variants are produced with identical technology given by

$$c_i = a_i l_i, \tag{1.2}$$

where l_i is labor used in the production of the $i$the variant and a_i is a productivity parameter. The total labor force is exogenously fixed at L, clearly full employment implies $\int_0^N l_i di = L$. Assume further that different variants are produced from technology monopolists who maximize profits and who are equally owned by the consumers maximizing utility. Then prices are given as a mark-up on marginal cost. These marginal costs are given by the wage rate w of the economy. Prices for the different variants are then given as (see e.g. Fujita et al. 2000)

$$p_i = \frac{1}{\rho}\frac{w}{a_i}. \tag{1.3}$$

It is clear from equation (1.3) that real wages (in terms of one variant of the consumption goods) is fixed by ρa_i. However, consumers consume a basket of goods according to their utility function. A price index commonly used in the context of models using the Dixit and Stiglitz (1977) index is $p = \left(\int_0^N p_i^{-(1-\rho)/\rho} di\right)^{-\rho/(1-\rho)}$. This index is equivalent to the unit cost function if a Dixit and Stiglitz (1977) index is used as production function and production costs are minimized. However, this index captures what is called as returns to differentiation, i.e. the index decline ceteris paribus if the set of variants grows larger. If one uses this index to compute real wages as $\frac{w}{p}$ and lets the set of variants increase, the real wage as just defined would steadily increase while it would be constant in terms of any good in the economy. This is probably why Krugman (1991) defines another price index for computation of the real wage in an environment similar to this one. He calls this "the true price index"[6] and defines is according to $\tilde{p} = \left(\int_0^N \frac{1}{N} p_i^{-(1-\rho)/\rho} di\right)^{-\rho/(1-\rho)}$. Clearly the

[6]Krugman (1991) treats the special case where he has two sets of variants where variants in each set are priced identically. Therefore he uses as weights instead of $\frac{1}{N}$ the share of each set in the total set of variants and has only two prices instead of a continuum of prices.

effect of the size of the set of variants disappears now and his measure of the real wage is $\frac{w}{p}$. This is important as the set of variants is usually determined by the extent of the work force in more elaborate models. Fujita et al. (2000) deviate from this procedure in determining the real wage as they take account of the degree of differentiation given by the set of variants. This discussion boils down to merely the distinction whether the returns to differentiation are utility or production based. If the latter is the case, as in Ethier (1982), then the returns of differentiation are real in a sense that they can be measured by data on wages and prices. If the former is the case the effect is more subjective in a sense that it can not be detected in such data.

The next question addressed is the localization of production factors. Krugman (1991) analysis the case where labor employable in the differentiated goods sector is mobile between the two countries engaging in trade. The result is, that convergence or divergence, in the sense of a stable distribution of the work force in the differentiated sector between the two economies or a total concentration of this type of labor in one of the countries, can take place, depending on the actual values of the parameters in the model.

Another interesting application of increasing returns to scale technologies can be found in Krugman and Venables (1995). As other studies from the field of the new economic geography it is primarily concerned with location of production in a world with trade possibilities. In particular they examine the role of transport costs in the determination of the location of global production. Regarding the assumptions about consumer preferences the authors assume that preferences are defined by an expenditure function of the Cobb-Douglas type. This is nothing else than another representation of Cobb-Douglas preferences. Regarding production technologies, there are important deviations with respect to other publications in this field. In each of the two economies considered, there is the possibility of production of two goods, good one and good two. Good one is produced with constant returns to scale and can be traded freely while good two occurs in many varieties and is produced with increasing returns to scale with a cost function as in e.g. Krugman (1979,

1981) and is therefore a differentiated good. The variants of this good can, on the one hand, be used as consumption goods or, on the other hand, as inputs in the production of good two according to a Dixit and Stiglitz (1977) index. The other factor used in producing good two is labor which is also the sole input in producing good one. Labor is mobile between these two sectors. The focus of this study is not on per capita terms like wages or per capita production. The set-up of the model is too complicated to yield closed from solutions for these figures and for the relationship between the extent of the work force and number of variants of good two. As such especially the influence of the extent of the market measured by labor force in the two economies has not been analyzed. But the differences to the other models summarized in this section is now that the degree of differentiation in the good two sector, i.e. the number of variants, has a direct impact on productivity as in Ethier (1982) through the production technology in the good two sector.

Building on the just cited study, Puga and Venables (1997) set up a model which is slightly more complicated than the one in Krugman and Venables (1995) but uses essentially the same building stones. What makes this study interesting in the context of this thesis is that the authors examine the impact of an increase of the labor force on different results of the model. Especially they are interested in the agglomeration effects of a change in the size of the work force[7]. There are two goods as in Krugman and Venables (1995), good one is produced with constant returns to scale using labor and land while good two is differentiated and is produced with increasing returns to scale using labor and and several composite goods which are aggregated according to a Cobb-Douglas function. The increasing returns to scale stem from fixed costs in terms of output of the particular variant of good two. The composite goods come from all the economies engaged in trade, for each economy one. They are obtained from Dixit and Stiglitz (1977) indices of all differentiated inputs from all countries, i.e. the variants of good two. Consumers have preferences as in Krugman and Venables (1995) represented by an expenditure function of the Cobb-Douglas type over good one and the different composite goods of all

[7]The authors interpret growth in the work force as technical change altering the efficiency units of labor.

economies. Interpreting their results is more complicated since they assume land as an additional input factor in the production for good one, which is held constant in their analysis. This automatically implies that, given changes in the labor force, aggregate production of good one has non-constant returns to scale. However, there is an effect of the size of the labor force on wages which is caused by the combination of increasing returns to scale in production and returns to differentiation in good two production. The model is too complicated to be solved for per capita production, wages or the degree of differentiation given by the number of good two producers. But besides trade costs, the aforementioned assumptions certainly play a role in determining the relationship between the size of the work force and wages.

1.5 New Growth Theory

The beginning of the new growth theory is seen in the publications of Romer (1987, 1990) where he proposed a new approach to motivate endogenous growth. To clarify things assume the following macroeconomic production function that gives output of an economy at a particular point in time t according to a Cobb-Douglas specification

$$Y_t = (A_t L)^\alpha K_t^{1-\alpha},$$

where L is labor input which is constant over time, K_t is the capital stock of the economy and A_t is the level of a labor augmenting technology[8]. Growth occurs when A_t is growing with t. Such behavior of A_t as a result of endogenous decisions in a decentralized economy is motivated by the fundamental contribution of Romer (1990) that assumes an underlying functional specification of production in the form of

$$Y_t = H_{Y,t}^\alpha L^\beta \int_0^{A_t} x_{i,t}^{1-\alpha-\beta} di,$$

[8]Although in this Cobb-Douglas set-up it is not meaningful to distinguish between labor and capital augmenting technologies, this terminology is used because it corresponds to the underlying idea of the Romer model.

where $H_{Y,t}$ is human capital used in production, L is ordinary labor and $x_{i,t}$ is the input of the $i$the variant of an intermediate input factor which is produced by a one-to-one technique from capital goods. These capital goods are produced from the aggregated capital stock K_t of the economy. Human capital and labor are equally distributed over the consumers in the economy. A_t is the upper bound of the set of available intermediate input factors, determining the level of technology. Romer's main argument for this specification is that it captures one important aspect of technology: Once an innovation has been made, i.e. the introduction of a new variant of the intermediate input factors, it benefits all potential users, i.e. the workers L in the above formulation. Although this formulation says something about technology, it does not say anything about growth. Consider a simple accumulation equation for K_t in continuous time with $\dot{K}_t$ denoting $\frac{\partial K_t}{\partial t}$

$$\dot{K}_t = Y_t - C_t,$$

where C_t is consumption expenditure out of total income at time t. Leaving A_t and L constant at the moment. Romer (1991) shows by simple calculus that the growth rate of the capital stock and hence of output must converge to zero. Thus for long run growth to take place, one must assume that A_t is growing over time. This is modeled in Romer (1990) as

$$\dot{A}_t = \delta A_t H_{A,t}, \tag{1.4}$$

where δ is a productivity parameter and $H_{A,t}$ is human capital devoted to R&D, and $H_{A,t} + H_{Y,t} = H$ with H the time invariant stock of human capital. From this immediately it follows that $\frac{\dot{A}_t}{A_t} = \delta H_{A,t}$, i.e. the growth rate of technology is directly proportionate to the stock of human capital devoted to R&D. In equilibrium this growth rate is also the growth rate for final output, the capital stock and consumption. To close the model an additional assumption about the determinant of the interest rate of the economy is needed. This is done by integrating a traditional Ramsey problem for capital accumulation into the model. The representative consumer of the economy maximizes lifetime utility given by

$$U_t = \int_0^\infty \frac{c_t^{1-\sigma} - 1}{1-\sigma} e^{-\rho t} dt, \tag{1.5}$$

where σ is the inverse of the intertemporal elasticity of substitution of consumption, c_t, and $\rho > 0$ is the constant rate of time preference. Utility (1.5) is maximized subject to the intertemporal budget constraint

$$\dot{a}_t = w_t + r_t a_t - c_t,$$

where a_t denotes assets, w_t is wage income from labor and human capital and r_t is the net interest rate. Besides a transversality condition, the optimum is characterized by the well known Keynes-Ramsey rule

$$\frac{\dot{c}_t}{c_t} = \frac{1}{\sigma}(r_t - \rho).$$

On the balanced growth path the growth rate of individual, c_t, and aggregate consumption, C_t is identical and constant as is the interest rate. This growth rate is given by

$$g = \frac{\delta H - \Lambda \rho}{\alpha \sigma + 1} \quad \text{with} \quad \Lambda = \frac{\alpha}{(1-\alpha-\beta)(\alpha+\beta)}, \tag{1.6}$$

and hence the equilibrium amount of human capital devoted to R&D is

$$H_{A,t} = \frac{H - \Lambda \rho / \delta}{\Lambda \sigma + 1}.$$

From (1.6) it is obvious that total available human capital positively influences the growth rate of the economy. Romer (1990) argues that this formulation of the model has an advantage over an alternative formulation where human capital is not explicitly recognized. This means that ordinary labor instead of human capital is used in equation (1.4) and the coefficient β is set to zero. This would have the effect that the growth rate of the economy is directly proportionate to the extent of the labor force. Now it is directly proportionate to the stock of human capital of the economy. But standard economic thinking about human capital is that it can be accumulated over time, as in the Uzawa-Lucas model (Uzawa 1965 and Lucas 1988), and is not stationary. This would lead to ever-increasing growth rate in this formulation of the model as can bee seen from equation (1.6). Generally this result of the model is interpreted in such a way that the scale of resources devoted to R&D influences the growth rate of an economy, an effect that was termed by Jones (2005) as the "strong" scale effect.

Although the modelling strategy of Aghion and Howitt (1992) is not used in its original formulation in this thesis, but ideas based on it, it is worth to give the basic intuition of their model. Aghion and Howitt (1992) choose not to instrumentize horizontal differentiation as in Romer (1990) as the engine of growth, but vertical innovations. Because of this, they termed their model a model of creative destruction inspired by Schumpetrian ideas. As time goes by, new intermediate input factors, which are used in final good production with a fixed limited degree of differentiation, are discovered with an increased level of productivity. New versions replace old ones, thereby creating steady state growth. This basic idea of increased productivity has been extensively used in models of the second generation type which are subject of the following paragraphs. Although the growth mechanism of Aghion and Howitt (1992) is different from that in Romer (1990), the model also exhibits the aforementioned "strong" scale effect.

It is this scale effect which was addressed by Jones (1995b), and was later termed by Jones (2005) as the "strong" scale effect, which is now known as the Jones critique. Jones (1995b) presented time series evidence that this scale effect is not supported by the data. In fact, after World War II, resources devoted to research and development (R&D) in developed countries clearly show a positive time trend which can not be found in the time series of the growth rates of these economies[9]. This criticism has led economists to develop growth models of the second generation type, which are not subject to the Jones critique.

[9]Indeed growth for example for the US economy slowed down during the 20th century.

The growth models of the second generation type group themselves into endogenous and semi-endogenous models. The former are models in which policy parameters, i.e. parameters reflecting government action to influence growth, have an impact on the long-run growth rate of an economy. The latter are characterized by long-run growth which is beyond the influence of policy measures. Both types of models have the absence of any scale effects in R&D expenditures on long-run growth rates of an economy in common.

Models of this type can be found in Jones (1995a), Kortum (1997), Segerstrom (1998), Young (1998), Peretto (1998), Dinopoulos and Thompson (1998), Aghion and Howitt (1998, ch. 12) and Howitt (1999). Jones (1999) gives an excellent overview of the mechanisms at work in these models. The first group consists of the models in Jones (1995b), Kortum (1997) and Segerstrom (1998) which are characterized by limited state dependence in a growth equation comparable to equation (1.4), i.e. the elasticity of the stock of knowledge in creating new ideas is strictly less than one. This leads inevitably to the result, that long-run growth rate of an economy is proportionate to the growth rate of resources devoted to R&D, typically some type of labor. As long as e.g. the population growth rate is seen as exogenous, these models belong to the semi-endogenous class since policy can not change the growth rate of the economy in the long run. Young (1998), Peretto (1998), Dinopoulos and Thompson (1998), Aghion and Howitt (1998) and Howitt (1999) have in common that they account for both horizontal differentiation as in Romer (1990) and vertical innovations as in Aghion and Howitt (1992). The model of Howitt (1999) belongs to the class of endogenous growth models since government action, e.g. a subsidy of R&D, influences long-run growth rates of the economy. The basic mechanism for eliminating the scale effect in the growth rate of the economy is the same in these models. As in later chapters the idea of the Young (1998) model is used extensively, the basic structure of that model is presented in the following.

Young (1998) assumes a production technology given by a Dixit and Stiglitz (1977) index as in Grossman and Helpman (1991a), i.e. final output of the economy at time t is given by

$$Y_t = \left(\int_0^{A_t} (\lambda_{i,t} x_{i,t})^{\alpha} di \right)^{\frac{1}{\alpha}},$$

where A_t determines the mass of the set of differentiated input factors indexed by $i \in [0, A_t]$, $\lambda_{i,t}$ is the quality level of the $i$the variant of the input factors and $x_{i,t}$ denotes the used quantity. $\alpha \in (0,1)$ determines the elasticity of substitution between different variants. Time is discrete and the representative consumer maximizes lifetime utility given by

$$U_t = \sum_{\tau=t}^{\infty} \frac{\ln c_\tau}{(1+\rho)^{\tau - t}},$$

subject to the intertemporal budget constraint

$$a_{t+1} = (1 + r_t) a_t + w_t - c_t,$$

where a_t are assets, r_t is the interest rate, w_t is wage income and c_t is consumption expenditure. The key assumption of the model is that the development of A_t is not governed by a growth equation like (1.4), but is determined by market entry behavior of monopolistic competitive firms producing different variants of input factors. To produce a particular variant in period t, a firm has to incur a fixed R&D investment in $t-1$ depending on the chosen quality level. These fixed costs are in terms of labor and must be financed until the variant is sold in period t. Capital accumulation of consumers serves to finance these investments. The fixed costs are given by

$$F_{j,i} = \begin{cases} f e^{\mu \lambda_{i,t} / \bar{\lambda}_{t-1}} & \text{if} \quad \lambda_{i,t} \geq \bar{\lambda}_{t-1}, \\ \eta e^{\mu} & \text{otherwise}, \end{cases} \tag{1.7}$$

where $\bar{\lambda}_{t-1}$ is the maximum quality of variant i in the past. If variant i has never been produced in the past $\bar{\lambda}_{t-1}$ is the average of maximum quality of all produced variants in the past. f, η and μ are exogenously given parameters. Once the fixed costs have been covered, the input factors can be produced at constant marginal costs using labor with a unit productivity as the sole input.

Producers of input factor variants maximize profits net of the fixed costs by setting the price and the quality level. In this model, this gives a standard mark-up of prices over marginal costs and a rule for the evolution of the quality level over time

$$\frac{\lambda_{i,t}}{\bar{\lambda}_{t-1}} = \frac{1}{\mu}\frac{\alpha}{1-\alpha}.$$

This rule in turn determines the fixed costs of every producer. Now, producers will enter the market for differentiated input factors as long as there are positive net profits. In equilibrium a zero profit condition holds which gives the result, that A_t is directly proportionate to the extent of the homogenous labor force, L_t, i.e. $A_t = \eta L_t$[10]. A constant fraction s of the labor force works in producing the input factors while the fraction $1-s$ is evolved in performing R&D for the next generation input factors[11]. This result gives rise to another important implication of these type of models, an effect termed by Jones (2005) as the "weak" scale effect. Since labor market clearing demands $x_{i,t} = \frac{sL_t}{A_t}$, final goods production with symmetric quality levels can be written as $Y_t = (\eta L_t)^{\frac{1-\alpha}{\alpha}} \bar{\lambda}_t s L_t$. In other words per capita production is increasing in the extent of the labor force with elasticity $\frac{1-\alpha}{\alpha}$ and the growth rate of per capita production depends positively on the population growth rate and the rate at which the quality level grows. This "weak" scale effect is common to all growth models of the second generation type (see Jones 1999).

1.6 Growth and Economic Integration

First generation models of growth strongly influenced thinking about growth and its determinants. It is therefore not surprising, that a large strand of the growth literature related to trade and economic integration from the end of the 80ies and the beginning of the 90ies focused on the impact of economic integration on growth rates. In the following the most influential publications from these fields are reviewed

[10] η is a constant given by a combination of exogenous model parameters, see Young (1998) for details.

[11] s also depends only on exogenous model parameters.

in part to motivate research on the relationship between size of integrated economies and their per capita terms.

Helpman (1988) gives an introduction into the economic arguments involved in looking at interactions of trade and growth. He reviews the literature of both trade and growth theory up to the late 80s in order to motivate the use of models incorporating increasing returns to scale, as they were mentioned in the above paragraph about the new trade theory and the economic geography. This approach seems meaningful since both new trade models and models of the new growth theory use the same tools and ideas in formulating theoretical models.

Grossman and Helpman (1991b), building on Grossman and Helpman (1990), develop an endogenous two sector growth model with intertemporal utility defined over two goods. These goods are either produced from labor or human capital combined with a CES index of sector specific intermediate input factors according to a Cobb-Douglas production function. The two sector goods are tradeable; the intermediate input factors are not. The economy considered is small and takes world prices for the two goods as given. Intermediate input factor producers are able to set a price over marginal costs, giving an incentive to innovate. The growth equation for the measure of intermediate input factors is similar to the one in Romer (1990). The growth rate of the measure of intermediate input factors is linear in the amount of human capital devoted to R&D. The remainder of this article is devoted to the effects of trade and R&D policies on growth of the small open economy. Due to the structure of the model it belongs to the first generation models with scale effects in the growth rate of the economy and hence it can not be used to formalize hypothesises on the relationship between the size of the economy or trade on per capita figures.

Rivera-Batiz and Romer (1991) study the impact of economic integration of two similar economies in the context of the original Romer (1990) model. Thereby they analyze the influence of the accumulation equation of technology, i.e. the measure of differentiated intermediate input factors. They distinguish two different cases: the knowledge-driven and the lab-equipment specification

$$\begin{aligned}\dot{A}_t &= \delta A_t H_{A,t},\\ \dot{A}_t &= R_t,\end{aligned}$$

where $H_{A,t}$ is human capital used in R&D and R_t is a constant fraction of final output devoted to R&D. Both specifications clearly imply a scale effect in the growth rate of an economy caused by the amount of the resources devoted to R&D[12]. To study the effects of economic integration of two economies Rivera-Batiz and Romer (1991) analyze several sets of assumptions about flows of goods and ideas between the two countries and examine when an effect of integration on the growth rate of both economies occurs. Clearly, because of the implied scale effects of the model, it is suited for studying scale effects in growth rates and not per capita terms.

Ventura (2005) develops a series of theoretical growth models taking explicit account of trade between different economies. As this work is rather detailed and comprehensive, only the intuition behind the theoretical models is presented here. The framework is an overlapping generation model with a constant savings rate and many sectors; preferences over different final goods is specified to be Cobb-Douglas. Production in any sector is characterized by a Cobb-Douglas production function with labor, capital and a CES index of intermediate input factors as inputs. The production function is augmented by a productivity factor. Intermediate input factors are produced according to a Cobb-Douglas production technique from labor and capital after covering fixed costs in terms of intermediates. Intermediate input factor producers enter the market until net profits are driven down to zero. Labor is homogenous and not sector specific. The model is solved for the deterministic and the stochastic case; the latter treats the productivity factor as a stochastic variable in the sector production functions. With this set-up a number of globalization scenarios are closer examined. Of interest, in this context, is the effect of increased market size if the economy is opened up for trade and its impact on per capita

[12] In the lab-equipment specification this scale effect arises because R_t, a fraction of output, growth at same rate as A_t.

terms. However, the model is too complicated to establish a fully reduced form for e.g. per capita production in terms of exogenous numbers such as labor. This is because labor is used in different sectors and stages of total production. The cases considered are totally free trade in all goods and mobility of production factors: free trade in goods with no factor mobility between countries, free trade in some goods and/or some intermediate input factors but with labor and capital immobility. There are different effects in the models making it difficult to identify the role of market size. In the first case there are factor reallocations to countries and sectors where the productivity factor is highest, reallocation of production to countries where the productivity factor is highest, and a positive effect of increased market size on productivity through a larger measure of intermediate input factors. The second case can under certain conditions lead to factor price equalization. There are effects of reallocation of production between countries inducing shifts of factors to specific sectors and thereby creating market size effects. This is because not all sectors are operating in all countries and factors are concentrated among a subset of sectors which gain in size and induce a larger measure of intermediate input factors. The third case is rather difficult to summarize because there a number of possible combinations for trade frictions to be included in the model. The main conclusion is that market size is influenced by globalization and that there are effects operating through the measure of intermediate input factors.

1.7 Motivation for Further Research

As has been shown in the preceding paragraphs, scale effects are important building stones and implications of modern economic models. Models from the new trade and the new economic geography as well as the new growth theory use them as important assumptions which directly translate to the observed results. But of course there are more implications of scale effects than the ones in the above cited literature. The models in the following chapters will all be based on growth models described above and will show the effects of scale on important economic issues.

The first two chapters will be on the development of wage inequality between different types of labor, i.e. between high and low skilled workers. The basic impetus for this research comes from an article by Acemoglu (1998), on directed technical change. This can serve as an explanation for rising wage inequality between high and low skilled workers in the recent decades in the U.S. and the UK economies. The basic argument of directed technical change is that after an increase in the relative supply of high skilled workers, the market for technologies directed to them extents. This makes innovations for this market relatively more profitable, inducing technical change in favor of high skilled workers. Acemoglu (1998) uses a first generation growth model in the spirit of Aghion and Howitt (1992) extended to the two sector case to account for two types of labor. In a later publication Acemoglu (2001) used the Romer (1990) model in a two sector version and obtained identical results. The mechanism in the model is directly related to the "strong" scale effect of the Aghion and Howitt (1992)/Romer (1990) model. A larger market for innovations directed to high skilled workers makes innovations in this direction more profitable pulling more R&D resources into this sector of the economy. Since the growth rate depends on the scale of R&D resources, it will increase as long as relatively more R&D resources are devoted to the high skilled sector. Although this is a temporaryly effect until a new equilibrium is reached where increased R&D activity in the high skilled sector has eliminated the extra profitability, higher growth rates for productivity in the past have permanent effects on the future level of productivity benefiting the high skilled. If this technology effect is strong enough it can under certain circumstances overcome the usual substitution effect which would lead the relative wage for high skilled to fall as a response to an increase in relative supply. The condition for this effect to occur depends on the elasticity of substitution between high and low skilled workers or sectors.

The approach taken by Acemoglu (1998, 2001) has the drawback in that it can not distinguish between skill and sector biases in technologies because skill and sector coincide. Acemoglu (2001) comments on this issue in an appendix and tries to resolve the problem by sketching a model in which two sectors both use high and low skilled workers but with different intensities. However, he is not able to solve the model in detail due to its complexity. Chapter two steps in at this point and proposes an alternative modelling strategy, which is a two sector Romer (1990) type endogenous growth model where both sectors employ high and low skilled workers. The market equilibrium determines the endogenous factor bias of technology within sectors and the sector bias between sector. It is thus a model of directed skill and sector specific technical change. The new result is that a rise of the relative wage of high skilled workers in response to an increase in relative supply depends not on the elasticity of substitution between high and low skilled workers, but on the elasticity of substitution between high and low skilled intensive goods. This is a distinction which can not be made with the models in Acemoglu (1998, 2001).

The model is extended to cover some situations which might occur in an open economy context. The two-country case is analyzed in two different set-ups: First, the case of one large and one small open economy trading with each other is explored, technical change is driven by the large economy. Second, the case of two equal sized trading countries is explored where technical change is determined jointly by both countries. The results of these two model extensions show that the results from the basic model of the closed economy carry over to the two open economy models. Due to factor price equalization, the behavior of wages is identical in both countries.

As mentioned in the previous paragraph growth models of the first generation have been criticized because of the “strong” scale effect. The result of the model of the first chapter relies on the existence of this scale effect. Therefore this model might be criticized as well. Acemoglu (2001) showed that the main result of his directed technical change theory survives if a model of the second generation is used. In particular he employes the Jones (1995a) model. As will be discussed in detail in chapter two, the Jones (1995a) model is a hybrid model, showing “strong” scale effects off the balanced growth path and only “weak” scale effects on the balanced growth path. Thus it might be questioned whether the inclusion of this type of growth model is a real robustness test of the theory of directed technical change. Chapter two is treating this issue by extending the work of Young (1998) to the two sector case with high and low skilled labor. Since the Young (1998) model is free of any "strong" scale effects, this a good way to test the hypothesis of directed technical change in the context of a second generation growth model. The model is developed with the use of different production technologies, i.e. the Romer (1987) technology and the Dixit and Stiglitz (1977) index used as a production function as in Grossman and Helpman (1991a). Furthermore, the Dixit and Stiglitz (1977) index is extended by a formulation due to Ethier (1982) that captures the so called returns of differentiation

$$Y_t = A_t^{\nu - \frac{1-\alpha}{\alpha}} \left(\int_0^{A_t} x_{i,t}^{\alpha} di \right)^{\frac{1}{\alpha}},$$

where A_t determines the set of variety of differentiated input factors, whose quantity for one variant i is denoted by $x_{i,t}$. $\frac{1}{1-\alpha}$ is the elasticity of substitution between differentiated input factors and ν gives the returns to differentiation. This formulation is more flexible in determining how the degree of horizontal differentiation is influencing productivity. It turns out that besides the elasticity of substitution between high and low skilled products, the returns to differentiation, ν, play a critical role. An extension of the model deals with the so called Krugman (1994) hypothesis which states that growing wage inequality, on the one hand, and rising unemployment on the other are just two sides of the same coin. While the US and the UK have experienced rising wage inequality in the past decades between high and low skilled workers, in continental Europe unemployment especially among the low skilled has emerged, possibly due to relative wage rigidity between these two types of labor. The aim of this extension is to confront the Krugman hypothesis, which was developed in the context of a simple labor market model, with a more complex economic environment. Especially it is confronted with endogenous technology which reacts to the skill structure of the economy and trade between two countries pursuing different wage policies. The result is that this hypothesis might be true if one is willing to make strong assumptions about the development of goods prices in the context of an open economy model, where in one country wages are set to clear the labor market while in the other country relative wages between high and low skilled are kept constant.

It will be argued in chapter three that the "strong" scale effect, which has been used in chapter two to explain the relationship between labor supply and the development of wage inequality, can be replaced by the "weak" scale effect. Doing so yields essentially the same results. In addition, a new growth model without any scale effect is presented and it will be shown that by using this model any relationship between labor supply and wage inequality disappears.

Having looked at the effects of scale effects on the development of labor productivity and wages, the question arises whether at least the "weak" scale effect driving the results of the second chapter is present in reality. The empirical literature on the existence of these scale effects might lead to acceptance of them, but there are some limitations. The starting point for an empirical analysis is the theoretical result that larger economies should exhibit a higher per capita production. This is, however, only valid for closed economies and things might be different in the case of open economies. To clarify things the third chapter develops a multi-region/country growth model of the second generation type using the idea of Young (1998) as in chapter three. Regions/countries are allowed to trade and capital is perfectly mobile, however, trade frictions exist in case of trade in goods. The result of this model is that the "weak" scale effect has its analog in an open economy context. The critical scale variable now is not longer the size of an economy, but its own size and the size of trading partners, corrected for trade frictions. This is why small and large open countries can achieve the same per capita production by "importing" economic size through openness to trade.

The model yields empirical tractable results which are tested in the empirical section of the forth chapter. The analysis is undertaken using a cross section of 88 countries in the year 2000, and on the regional level using data for Europe on 221 regions of the "old" 15 countries of the European Union and county data on 3075 main land counties in the US. The model estimated is a spatial econometric model where the scale variable explaining per capita production is a weighted sum of region/country population sizes. As weights, the inverse great circle distance between countries and regions is used. In the case of the country cross-section the economies used to compute the scale variable are the G7 countries because they are usually considered to be the major source of technology and therefore the determinant of per capita production. On the regional level the scale variable was defined over all regions since those under consideration are all from well developed countries. The estimation techniques used, control for possible endogeneity and spatial autocorrelation. The empirical results show clearly the significance of the scale variable both in the country and the regional case. On the regional level the scale variable seems to have a larger impact on per capita production than on the country level.

Chapter five summarizes the findings of the thesis and gives major conclusions. The role of scale effects in theoretical economics is stressed and in the light of the empirical results they seem to be a good description of the world. Comments are given about future research possibilities both theoretical and empirical. The former considers economic questions where scale effects might play a critical role and are so far not implemented by the existing literature. On the empirical side some ideas on extensions of the models estimated in chapter four are given.

Chapter 2

"Strong" Scale Effects and Wage Inequality

This and the following chapters deal with wage inequality between different types of labor; especially the focus is on wage inequality between high and low skilled workers. This is a direct application of what the "strong" scale effect growth models of the first generation have in common. A variant of the Romer (1990) model, using the R&D specifications in Romer and Rivera-Batiz (1991), is extended to the two sector case with heterogenous labor, i.e. high and low skilled. The model is an extension of the work in Acemoglu (1998, 1999a,b and 2001) and yields new results with respect to the relationship between employment and skill structure of the economy and wage inequality. Data from the NBER-CES Manufacturing Database are used to confirm the predictions of the model.

2.1 Introduction

The growing wage inequality between high educated and less educated workers in the U.S. and other major countries has been a field of high interest for economists. For a recent review of the corresponding literature see Acemoglu (2002). The relative number of college graduates in the American working population has increased from 6 percent in 1939 to over 28 percent in 1996. By the same time the proportion of workers not having a high school degree dropped from 68 to 10 percent (Autor, Katz and Krueger, 1998). Despite this rise in the relative supply of educated workers, it is well known that the wage mark up for education, measured by the college wage premium, has also increased during this period (Acemoglu (1998)), with the exception of the 1970s where the college premium actually fell. This premium, compared with workers having only a high school degree, enabled college graduates to earn a 55 percent higher wage in the 1970s. During the 1970s this difference fell to 41 percent but increased thereafter to 62 percent in 1995. These numbers can also be verified by looking at the wage distribution for the manufacturing sectors in the U.S.. The National Bureau of Economic Research provides data on 459 4-digit SIC (Standard Industry Classification) manufacturing industries for the years 1958 to 1996[1]. Unfortunately, no direct figures for high and low skilled workers are given, but the database distinguishes between non-production and production workers, which can be used as proxies for high and low skilled workers (see Berman, Bound and Machin 1997). For these two types of employees, the numbers of industry employment and wages are given and are used in the following to calculate the relative wage for high skilled workers. This relative wage is computed by using aggregated numbers of the high and low skilled wage, obtained from the weighted sum of wage costs per worker from all 459 industries. As weights, the fraction of the industries in total manufacturing employment were used. Figure 2.1 shows the development for this relative wage for the time period 1958 to 1996. As can be seen, the relative wage fluctuates around a constant mean until the beginning of the 1970s, declines afterwards before it increases steadily during the 1980s and the early 1990s. The coincidence with the development of the relative supply can be seen from figure 2.2. The figure shows the aggregated ratio of non-production to production workers as a weighted sum of all industries, weights are again the employment shares of the industries. The relative supply of non-production workers steadily increased, but more slightly until the end of the seventies and stronger thereafter. Thus, we have a parallel development of relative supply and wages of high skilled workers.

[1]The NBER-CES Manufacturing Database, see http://www.nber.org/nberces/nbprod96.htm.

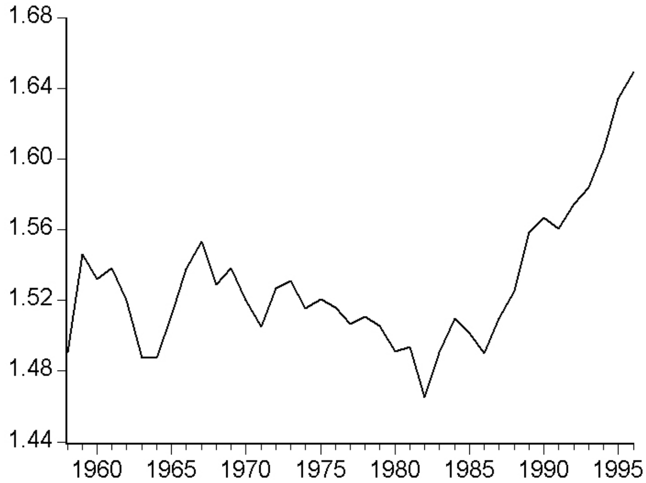

Figure 2.1: Relative wage of non-production workers

Ratio of aggregated wage costs of non-production and production workers in US manufacturing industries. Data source: National Bureau of Economic Research.

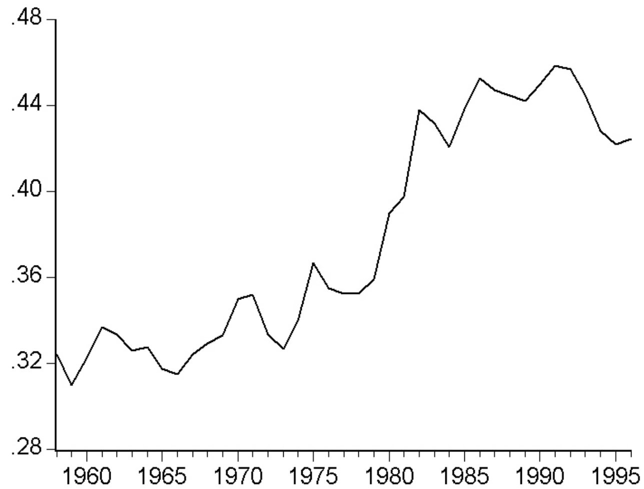

Figure 2.2: Relative supply of non-production workers

Ratio of aggregated non-production and production employment in US manufacturing industries. Data source: National Bureau of Economic Research.

One popular view in light of these facts is that the technological change which took place during the last decades was skill-biased, favoring the high skilled relatively more than the low skilled workers[2]. Other theories explaining the rising relative wage for the high skilled focus on the institutional change that took place in the labor market. Notably these are the declining minimum wage and the declining unionization (see Freeman (1991), DiNardo, Fortin and Lemieux (1995) or Lee (1999)). Yet another possible reason for the rise in the relative wage is the impact of increasing trade with less developed countries, see among others e.g. Feenstra and Hanson (2001). Although these arguments might be important, the focus of this paper lies on the change of technology which, as will be shown, can have major consequences on the distribution of wages. The aforementioned skill bias hypothesis obviously raises the question why the technological development was shaped in favor of the high skilled. Are there good economic reasons guiding the research sectors of an economy to invent relatively more technological advances for the high skilled? This question is addressed by the literature concerning the so-called directed technological change (see Acemoglu (1998, 1999a,b, 2001), Acemoglu and Zilibotti (2001) and Kiley (1999)). The models in these articles argue that the direction of technological development is influenced by the demand of producing firms using the technology. It is assumed that the economy consists of two sectors, one uses only high skilled workers and the other only low skilled. Both sectors produce intermediate goods which are combined in the final stage of production to yield the final output. Further the articles focus on the situation where the technology used in the different sectors is skill specific, i.e. the high skilled work with a different set of production technology than the low skilled. In other words skill and sector have the same meaning in these models. If the number of potential users of one specific technology increases, the profit for the research facility that invents that technology increases as well. This is the so-called market size effect. This effect is accompanied by the price effect: If the sector employing only high skilled personal is growing through an increase of the high skilled proportion of the work force and technological advances, the relative price of

[2]Also objections against this hypotheses can be found in the literature, see e.g. Card and DiNardo (2002)

its output used in final goods production decreases because of the usual substitution process. This effect counteracts the market size effect and the overall effect depends on the absolute value of the elasticity of substitution in the final production stage. The main result of these articles is that, if the mentioned elasticity of substitution is larger than a certain threshold, then the directed technological change leads to an increase of the wage mark up for high skilled if the proportion of the high skilled in the working population increases. Although the suggestion of the cited articles is very appealing, it has the drawback that the above models make an important strong assumption: The skill bias in technological progress is at the same time the sector bias. This is because only high skilled workers are present in the high skilled sector and only low skilled workers are employed in the low skilled sector. This rules out the possibility of the occurrence of different skill and sector biased technological changes. The literature concerning the implication of sector and skill biased technological change comes mostly from the field of international economics. The analysis in this literature focuses on the effects of a different skill bias of technological change across sectors. Xu (2001) analyzes exogenous skill and sector biased technological change in a two-country, two-goods, two-factors Heckscher-Ohlin model. He shows how changes in the exogenous technology parameters affect the relative factor prices under different sets of assumptions about the trade environment under which the economy acts. Krugmann (2000) also addressed the question of how relative factor prices of high and low skilled workers are affected by sector specific skill biased exogenous technological change in a Heckscher-Ohlin model. Recently Haskel and Slaughter (2002) used a model with exogenous technological changes which can take the form of sector specific skill biased technological change as well as sector specific skill neutral technological progress. These authors show how these different sources of technological changes affect the relative wage of high and low skilled workers. What is missing seems to be a unifying approach that takes into account the different mentioned technological changes as well as their possible endogeneity. The model developed in this paper aims to add to the literature by filling this gap. This is done by presenting a framework which allows for different endogenous skill biased technological changes in different sectors as well as different endogenous technological changes which are skill neutral but sector biased. Therefore it might be termed a model of directed sector and skill specific technological change. It is shown how an increase of the relative wage of the high skilled workers can happen in response to a change in the skill structure of the economy. In particular this change in the skill structure corresponds to a rise in the ratio of total high skilled to low skilled workers accompanied by a higher growth rate of high skilled employees in a "high-tech" sector than in a "low-tech" sector. The difference between these two sectors being that high skilled workers are ceteris paribus more productive in the "high-tech" than in the "low-tech" sector. Furthermore it will be shown that some of the above cited models of directed technological change can be seen as special cases of the model in this chapter, and therefore might be interpreted carefully. Section two sets up the basic model and section three examines the direction of endogenous skill and sector specific technological change. Section four compares the relationship of the presented model and some existing models. Some quite interesting open economy extensions are presented in section five. Finally the last section draws some conclusions.

2.2 The Basic Model

2.2.1 The Production Technology

To analyze how the technological changes mentioned in the introduction can affect the relative wage of high and low skilled workers, a two sector, two factor model is used. Firms in both sectors of the economy are producing, using both high and low skilled workers denoted by H_j and L_j respectively. The index $i = H, L$ corresponds to a "high-tech" and a "low-tech" sector, the index j denotes firm j. Firm j in sector i produces output $Y_{j,i}$ with the following production technology:

$$Y_{j,i} = \left\{ \left[\left(\int_0^{A_{i,j}} x_{i,j,l}^{\beta} dl \right) L_j^{1-\beta} \right]^{\rho} + \mu_i \left[\left(\int_0^{A'_{i,j}} x_{i,j,h}^{\beta} dh \right) H_j^{1-\beta} \right]^{\rho} \right\}^{\frac{1}{\rho}} . \tag{2.1}$$

Besides the use of labor there are also other inputs involved in production. For each type of labor there is a continuous set of technological equipment denoted by $A_{i,j}$ and $A'_{i,j}$ that can be used. The quantity of each particular machine to be used by firm j together with low and high skilled labor is denoted by $x_{i,j,l}$ and $x_{i,j,h}$. To keep the analysis tractable it is assumed that these machines fully depreciate after use in one particular point of time. The two sets of technological equipment play an important role in the model. Similar as in Stiglitz (1969) it is assumed that they together form the support $[0, a_i]$ and that $A_{i,j} = [0, \gamma_{i,j} a_i]$, $A'_{i,j} = [\gamma_{i,j} a_i, a_i]$ and $\gamma_{i,j} \in [0, 1]$. These two disjoint sets might be interpreted as the technological resources which areexclusive to each type of labor. Since this is a fundamental assumption of the model it seems necessary to elaborate on this issue a little bit more. To justify the assumption it is first necessary to think about these two disjoint sets maybe not literally as sets of machines, as in usual growth models, but more of technological resources. As will be shown later, the demanded quantity of each variant, $x_{i,j,l}$ and $x_{i,j,h}$, will be the same regardless with which kind of labor it is combined with it. Therefore the assumption can be interpreted as a budgetary problem.Aas will be shown later, firm j is willing to spend a certain fraction of its revenues on technological resources and has to decide how much of this budget it will devote to the high and low skilled departments. The technological equipment and labor are combined according to a Cobb-Douglas production function with output elasticities β and $1 - \beta$. This intermediate output which comes from the high and low skilled departments of firm j is then combined according to a CES production function to yield the final output $Y_{i,j}$ of. The production technology in this final stage is characterized by two parameters μ_i and ρ. μ_i is a sector specific parameter which determines the productivity of the high skilled intermediate output relative to the low skilled intermediate output. It is plausible to assume that $\mu_H > \mu_L \geq 1$, i.e. high skilled intermediate products are more productive in the "high-tech" sector and are in general not less productive than low skilled intermediate products. The parameter ρ determines the elasticity of substitution $\sigma = \frac{1}{1-\rho}$ between high and low skilled intermediate products. If $\sigma = 0$ then there is no substitution possible and the sector stage of production is Leontief. The case $\sigma = 1$ is the Cobb-Douglas case and with $\sigma = \infty$ the final production stage is linear and the two intermediate products are perfect substitutes. If $\sigma < 1$ then the two intermediate products might be termed as in Acemoglu (2001) as gross compliments, if $\sigma > 1$ they are gross substitutes. For now it is only assumed that $\sigma > 0$ is fulfilled, which seems to be a quite reasonable assumption. Finally, it is clear that the production function (2.1) has constant returns to scale with respect to all four inputs $x_{i,j,l}$, $x_{i,j,h}$, H_j and L_j. From the preceding discussion of the production technology it is obvious that the parameter $\gamma_{i,j}$ characterizes the state of skill bias in technology. A rise in $\gamma_{i,j}$ is by construction low skilled labor augmenting whereas a rise in $1 - \gamma_{i,j}$ is necessarily high skilled labor augmenting. To proceed in the analysis it is necessary to make some assumptions about the environment in which firm j acts. To simplify the computation, let firm j be one of many in its sector so that competition between these firms is perfect. Furthermore, if there is a large number of firms, the wages for high and low skilled labor can be seen as exogenous to the individual firm. All firms are faced with the same wages which are identical for high and low skilled workers regardless in which sector they are employed. The wages are entirely determined by the labor market which is also assumed to be perfectly competitive so that wages adjust to clear this market. No unemployment can occur. Regarding the market for technological resources I abstract from perfect competition. This is necessary to motivate a research sector which invents and sells the different variants $x_{i,j,l}$, $x_{i,j,h}$. The firms engaging in research will benefit from their inventions by a comparative advantage in marginal costs for producing their particular variant.

2.2.2 The Demand for technological Resources

First, to simplify the notation, some additional terms should be introduced. The sector stage of production combines the high and low skilled intermediate products which are termed in the following by $Y_{H,i,j} = (\int_{A_{i,j}} x_{i,j,h}^{\beta} dh) H_j^{1-\beta}$ and $Y_{L,i,j} = (\int_{A_{i,j}} x_{i,j,l}^{\beta} dl) L_j^{1-\beta}$, the unit costs of firm j, associated with the production of these intermediate products, are denoted by $c_{L,i,j}$ and $c_{H,i,j}$. The production function

(2.1) can then be formulated as

$$Y_{i,j} = \left[Y_{L,i,j}^{\rho} + \mu_i Y_{H,i,j}^{\rho}\right]^{\frac{1}{\rho}}, \tag{2.2}$$

and has a usual corresponding unit cost function if the intermediate products are used in a cost minimizing way. This unit cost function is equal to the price of final output in each sector due to the assumption of perfect competition:

$$P_{i,j} = \left[c_{L,i,j}^{-\frac{\rho}{1-\rho}} + \mu_i^{\frac{1}{1-\rho}} c_{H,i,j}^{-\frac{\rho}{1-\rho}}\right]^{-\frac{1-\rho}{\rho}}, \tag{2.3}$$

where $P_{i,j}$ is the price of final output of firm j in sector i. Later it will be clear that all firms in one sector will charge the same price. Let $\chi_{H,j,h}$ and $\chi_{L,j,l}$ be the prices for the variants of $x_{i,j,h}$ and $x_{i,j,l}$ and let these prices be identical for all firms in one sector. The demand for each variant is determined by the first order condition which equates the marginal product of each variant with its price. Since the variants combined with one kind of labor enter the production function symmetrically it is clear that the demand for all these variants by firm j will be identical. Standard calculation gives the following demand functions:

$$x_{H,j,h} = H_j c_{H,i,j}^{\frac{1}{1-\beta}} \beta^{\frac{1}{1-\beta}} \chi_{H,i,h}^{-\frac{1}{1-\beta}}, \tag{2.4}$$

$$x_{L,i,l} = L_j c_{L,i,j}^{\frac{1}{1-\beta}} \beta^{\frac{1}{1-\beta}} \chi_{L,i,l}^{-\frac{1}{1-\beta}}, \tag{2.5}$$

which have a constant price elasticity. Since the inventor and producer of the particular variant has a comparative advantage in marginal costs, he sets the price as a mark-up on marginal costs of production. Let this mark-up be denoted by $\tilde{\gamma} > 1$ and marginal costs be constant at one[3], then the price for each unit of a variant is $\chi_{H,i,h} = \chi_{L,i,l} = \tilde{\gamma}$.

[3]The price of the final good of the economy is also normalized to one. This means that final output is used to produce the variants of technological equipment.

2.2.3 Determinants of the Relative Wage for High Skilled

Given perfect competition it is clear that the relative wage, the ratio of the high to the low skilled wage, is determined by the ratio of the corresponding marginal products

$$\frac{w_H}{w_L} = \mu_i \left(\frac{1-\gamma_{i,j}}{\gamma_{i,j}}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{c_{H,i,j}}{c_{L,i,j}}\right)^{\frac{\beta}{1-\beta}\frac{\sigma-1}{\sigma}} \left(\frac{H_j}{L_j}\right)^{-\frac{1}{\sigma}}. \tag{2.6}$$

What still needs to be determined is the ratio between the costs of high and low skilled intermediate products. Since in the first production stage the variants $x_{i,j,h}$ and $x_{i,j,l}$ are combined with the corresponding kinds of labor, using a Cobb-Douglas production technology, this ratio is given by

$$\frac{c_{H,i,j}}{c_{L,i,j}} = \left(\frac{\gamma_{i,j}}{1-\gamma_{i,j}}\frac{w_H}{w_L}\right)^{1-\beta}. \tag{2.7}$$

Equation (2.7) assumes that the input factors in the first production stage are used in a cost minimizing way. Using (2.6) and (2.7) the relative wage can be written as

$$\frac{w_H}{w_L} = \mu_i^{\frac{\sigma}{\sigma(1-\beta)+\beta}} \left(\frac{1-\gamma_{i,j}}{\gamma_{i,j}}\right)^{\frac{(\sigma-1)(1-\beta)}{\sigma(1-\beta)+\beta}} \left(\frac{H_j}{L_j}\right)^{-\frac{1}{\sigma(1-\beta)+\beta}}. \tag{2.8}$$

From this equation it can first be seen that the relative wage decreases if the relative skill structure measured by the ratio of high to low skilled workers in firm j increases. This is the standard substitution effect. The effect of high and low skilled augmenting technological change depends on the elasticity of substitution between high and low skilled intermediate products. If $\sigma > 1$, then the two just mentioned input factors are gross substitutes and high (low) skilled augmenting technological change will also be high (low) skilled biased. If $\sigma < 1$, the input factors are gross compliments and high (low) skilled augmenting technological change will be low (high) skilled biased. If production in the final stage is Cobb-Douglas, $\sigma = 1$, then high and low skilled augmenting technological change has no bias. The analysis so far is quite standard and the results are not very surprising. In the following section we will analyze what happens if the skill bias represented in the model by a change in the parameter $\gamma_{i,j}$ becomes endogenous.

2.3 Endogenous Technological Change

In this section first the implications of endogenous skill specific technological change will be examined. With the obtained results in hand, the model will then be extended by taking also into account that sector specific technological change is driven by market forces.

2.3.1 Skill Specific Technological Change

Using the preceding results, the unit cost function corresponding to the production function (4.23), given that firm j is a cost minimizer, can be written as

$$P_{i,j} = \tilde{\gamma}^{\beta}\beta^{-\beta}\Bigg[\left(\frac{w_L}{\gamma_{i,j}a_i(1-\beta)}\right)^{-(\sigma-1)(1-\beta)} + \mu_i^{\sigma}\left(\frac{w_H}{(1-\gamma_{i,j})a_i(1-\beta)}\right)^{-(\sigma-1)(1-\beta)}\Bigg]^{-\frac{1}{\sigma-1}}. \tag{2.9}$$

Note that in (2.9) the price for the variants of technological equipment is already substituted and factored out. The use of the variants of technological equipment has the effect of lowering the effective wage costs by a factor which is given by the number of these variants used with each kind of the two types of labor. Facing this relationship, firm j is now assumed to endogenously determine the amount of technological resources which is to be devoted to each kind of labor by choosing the appropriate value for $\gamma_{i,j}$. Differentiating (2.9) with respect to $\gamma_{i,j}$, setting this derivative equal to zero and using Shepard's Lemma, the following condition must be satisfied by firm j

$$\frac{1-\gamma_{i,j}}{\gamma_{i,j}} = \frac{w_H H_j}{w_L L_j}. \tag{2.10}$$

That is, the ratio of the technological resources devoted to high and low skilled workers is equal to the ratio of their wage costs. Checking the second derivative it turns out that this can be cost minimum or maximum depending on whether

$\sigma < 2 + \frac{\beta}{1-\beta}$ or $\sigma > 2 + \frac{\beta}{1-\beta}$. The economic reasoning behind the first case is, given the relative wage, if firm j decides to hire more high skilled workers relative to low skilled workers, the marginal product of each variant $x_{i,j,h}$ increases. This makes it profitable for the firm to use more of these variants in combination with high skilled labor. At some point this incentive stops because high and low skilled workers are relatively essential in the production process. If the second case is true, it pays for firm j to concentrate only on one skill group because the elasticity of substitution is so high that the other skill group can be easily replaced. Since the first case is the interesting one, the focus of the following analysis will be on that case. Section 4 will examine a special case where the second case occurs. Combining equations (2.8) and (2.10), it follows that the ratio of technological resources for high and low skilled workers can be written as a function of the relative wage and μ_i alone. This means that the distribution of technological resources is identical for all firms in one sector:

$$\frac{1-\gamma_i}{\gamma_i} = \mu_i^{-\frac{\sigma}{(\sigma-1)(1-\beta)-1}} \left(\frac{w_H}{w_L}\right)^{\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}}. \tag{2.11}$$

There are some interesting special cases arising from different values of the elasticity of substitution. If high and low skilled intermediate products are gross compliments then a higher relative high skilled wage means a higher ratio of technological resources for the high skilled employees. The same applies for the productivity parameter μ_i. The amount of variants of technological equipment used in combination with high skilled labor will always be larger than 0.5 if wages for high skilled are higher than for low skilled and will be higher in the high-tech sector. If high and low skilled intermediate products are gross substitutes, it is still true that a higher productivity parameter μ_i leads to a higher ratio of technological resources for the high skilled. But the opposite is true with respect to the relative wage. If this relative wage is increasing, more variants of equipment will be devoted to the low skilled workers. If γ_i is identical for all firms in sector i, it can be seen from equation (2.10) that the skill composition must also be identical for all firms in sector i. The relative wage is then entirely determined by the ratio of high to low skilled workers in each sector.

$$\frac{w_H}{w_L} = \mu_i^\sigma \left(\frac{H_i}{L_i}\right)^{(\sigma-1)(1-\beta)-1}, \tag{2.12}$$

where H_i and L_i denote the high and low skilled workforce in each sector. Since the technology distribution parameter γ_i is identical for all firms in one sector as is the relative wage, all firms in one sector employ the same skill composition. This skill composition must then also be equal to the skill composition of the whole sector. Equation (2.12) is exactly the result of Acemoglu (1998), but here it applies only to one sector and not economy wide. Note that the possibility of a rising relative wage in response to an increase in the ratio of high to low skilled workers can never occur on the sector level because of the implied condition $\sigma < 2 + \frac{\beta}{1-\beta}$. In the following sections it will be shown that this is not hindering the relative wage for the high skilled from rising if sector specific technological change is taken into account. To guarantee that the relative wage is identical in each sector it must further be assumed that high and low skilled workers can freely choose in which sector they work. The ratio of the relative skill compositions of the two sectors is given by

$$\frac{H_H/L_H}{H_L/L_L} = \left(\frac{\mu_H}{\mu_L}\right)^{-\frac{\sigma}{(\sigma-1)(1-\beta)-1}}. \tag{2.13}$$

From this equation it can be seen that the high-tech sector always has the higher ratio of high to low skilled workers. Substituting the relative wage as determined by equation (2.10), in equation (2.11) it turns out that the relative distribution of the variants of technological equipment can be written as

$$\frac{1-\gamma_i}{\gamma_i} = \mu_i^\sigma \left(\frac{H_i}{L_i}\right)^{(\sigma-1)(1-\beta)}. \tag{2.14}$$

Regarding the distribution of the variants of technological equipment, the following conclusions can be drawn. If $\sigma < 1$, the "high-tech" sector has the higher ratio of high skilled to low skilled employees. In addition to this, this sector has the higher productivity parameter μ_i. From equation (2.11) it is clear that the "high-tech" sector has the higher relative distribution of the variants of technological equipment. So the effect of the productivity parameter μ_i outweighs the counteracting effect of the higher relative skill composition in (2.14). From where the elasticity of substitution is larger than one, both effects work in the same direction

2.3.2 Sector Specific Technological Change

At the heart of the following analysis lies the unit cost function of the firms in the two sectors. Since all firms in each sector are identical, it is sufficient to concentrate on the sector production function. After the endogenous adjustment of the distribution parameter γ_i and taking in account the demand for variants of technological equipment, the production function can be written in reduced form as[4]

$$Y_i = a_i \beta^{\frac{\beta}{1-\beta}} \tilde{\gamma}^{-\frac{\beta}{1-\beta}} P_i^{\frac{\beta}{1-\beta}} \left[L_i^{(\sigma-1)(1-\beta)} + \mu_i^{\sigma} H_i^{(\sigma-1)(1-\beta)} \right]^{\frac{1}{(\sigma-1)(1-\beta)}} . \tag{2.15}$$

To determine the corresponding cost function it is also necessary to compute the expenditures for the different variants of technological equipment. Surprisingly, it turns out using equations (2.4), (2.5), (2.9), (2.10) and (2.15), the demand for each variant is the same, regardless of which kind of labor it is combined with[5]

$$x_{i,l} = x_{i,h} = \frac{1}{a_i} \beta \tilde{\gamma}^{-1} P_i Y_i. \tag{2.16}$$

Now using equations (2.12), (2.15) and (2.16), the unit cost function can be computed as

$$\begin{aligned} P_i = a_i^{-(1-\beta)} (1-\beta)^{-(1-\beta)} \beta^{-2\beta} \Bigg(& w_l^{\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}} \\ & + \mu_i^{-\frac{\sigma}{(\sigma-1)(1-\beta)-1}} w_H^{\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}} \Bigg)^{\frac{(\sigma-1)(1-\beta)-1}{\sigma-1}} . \end{aligned} \tag{2.17}$$

[4]The reduced form can be obtained by using (2.1), the marginal product for each type of labor and the demand for the variants of technological equipment (2.4) and (2.5)

[5]The demand for each variant for all firms in one sector depends linearly on the produced sector output. Therefore the demand functions can easily be aggregated on the sector level.

Up to now the sector wide technology parameter a_i was treated as exogenous and before turning to the case where a_i becomes endogenous let us see what equation (2.17) can tell about the development of wages. Building the total differential of (2.17) for each sector and subtracting the results yields the following relationship between the development of wages, the development of prices and the states of sector technology

$$\hat{w}_H - \hat{w}_L = \frac{1}{\omega_H - \omega_L}\left[\frac{1}{1-\beta}(\hat{P}_H - \hat{P}_L) + \hat{a}_H - \hat{a}_L\right]. \tag{2.18}$$

From now on the notation $\hat{z}$ denotes the growth rate of the variable z. In equation (2.18) ω_i denotes the wage bill share of the high skilled in sector i, $\omega_i = \frac{w_H H_i}{w_L L_i + w_H H_i}$ which can equivalently be written by using equation (2.12) as

$$\omega_i = 1 - \left[1 + \mu_i^{-\frac{\sigma}{(\sigma-1)(1-\beta)-1}}\left(\frac{w_H}{w_L}\right)^{\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}}\right]^{-1}. \tag{2.19}$$

The only difference between the two wage shares of the high skilled in the two sectors comes from the productivity parameter μ_i. Here again a discussion about the effects of the elasticity of substitution is in order. From equation (2.11) and (2.19) it can be seen that the wage share is influenced by the same forces as the relative distribution of the variants of technological equipment and therefore the same arguments apply. If high and low skilled intermediate products are gross compliments, a higher productivity parameter μ_i implies a higher skilled wage share in the "high-tech" sector. A rise in the relative wage of the high skilled increases this wage share by less in the "high-tech" sector than in the "low-tech" sector in percentages. The change in absolute value is larger in the "high-tech" sector if the high skilled wage share is smaller than 50 percent in the "low-tech" sector. In the second case where high and low skilled intermediate products are gross substitutes but $\sigma < 2 + \frac{\beta}{1-\beta}$, a higher μ_i still implies a higher high skilled wage share but a rise in the relative wage of the high skilled now lowers this share. Furthermore it lowers ω_i in the "high-tech" sector by less than in the "low-tech" sector in percentages.

In absolute value this change is larger for the "high-tech" sector than for the "low-tech" sector if again the wage share for the high skilled is smaller than 50 percent in the "low-tech" sector. To summarize, if the elasticity of substitution is smaller than $2 + \frac{\beta}{1-\beta}$, the term $\omega_H - \omega_L$ must be positive. A rise in the relative price of the "high-tech" good and a rise in the relative state of technology given by $\frac{a_H}{a_L}$ have a positive impact on the relative wage of the high skilled. We have seen in the preceding section that the skill composition has effects which influence the skill bias in technological change. The next step in the analysis is now to find out what effects the skill composition of the workforce has on the development of the sectoral technological change. Using equation (2.16) the instantaneous profits of the inventors of new variants of technological equipment for the "high-tech" and the "low-tech" sector are given by $\pi_i = \frac{1}{a_i}\beta\frac{\tilde{\gamma}-1}{\tilde{\gamma}}P_iY_i$ and the value of the discovery of a new variant is determined by the dynamic programming equation $rV_i - \dot{V}_i = \pi_i$, where r is the interest rate which is possibly time variable. This equation relates the discounted present value of future profits V_i to the flow of profits π_i. The term $\dot{V}_i$, the derivative of V_i with respect to time, reflects the possibility that the present value might be time varying. Focusing on a balanced growth path where the present values are constant, they are given by

$$V_i = \frac{1}{r}\frac{\tilde{\gamma}-1}{\tilde{\gamma}}\beta P_i^{\frac{1}{1-\beta}}\left(L_i^{(\sigma-1)(1-\beta)} + \mu_i^{\sigma} H_i^{(\sigma-1)(1-\beta)}\right)^{\frac{1}{(\sigma-1)(1-\beta)}}. \tag{2.20}$$

In the literature of directed technical change there are two quite well known effects present. First the price effect: A higher price of the final output using the particular variant increases the profits of the inventing monopolist. Second, the market size effect: The larger the market for a variant, i.e. the larger the number of workers who are to use the technology, the higher the profits for the inventor. If the number of employees in one sector increases, naturally output of this sector will increase as well. As a consequence of this higher supply of final products, its relative price will fall, so the price effect works in the opposite direction as the market size effect. Note that regardless of the value of the elasticity of substitution, the market size effect is always positive. A more interesting formulation of equation (2.20) can be obtained by using (2.17) and (2.12) to yield

$$V_i = \frac{1}{a_i r} \frac{\beta}{1-\beta} \frac{\tilde{\gamma}-1}{\tilde{\gamma}} (w_H H_i + w_L L_i), \tag{2.21}$$

which states that the profits from inventing a new variant increase with the wage bill of the sector. In other words it will be more profitable to invent for the sector which has the higher wage costs. To determine the possible sector bias in the development of new variants of technological equipment one first has to make some assumptions about the environment under which these new variants are to be discovered. The innovation possibilities frontier can take two forms following the literature on endogenous growth. A first possibility is the so called lab equipment specification of Rivera-Batiz and Romer (1991). In this specification only final output is used in the production of new blueprints for new variants. The second possibility is the knowledge based R&D specification of Rivera-Batiz and Romer (1991). Here long run balanced growth is produced via positive spill over effects from past research which increase the productivity of current R&D activities.

Sectoral Technological Change with the Lab Equipment Specification

Only final output is used in the production of new designs for the variants of technological equipment. Since in the model there are two kinds of sector output, one has to decide how these two types of goods are used in the production of new ideas. One possible assumption would be that only "high-tech" products are used in research, yet another is that a certain combination of the two goods enter the production of new variants of technological equipment. To keep the analysis tractable this last possibility will be used; I will return to this issues later on. To close the model there needs to be one more assumption about the financing of R&D activities. It is reasonable to assume that the households or consumers of the economy save a part of their income which is then used by the research sector for R&D. To keep the analysis simple only consumers with a constant marginal propensity to save are considered[6]. The innovation possibility frontier in the lab equipment specification takes the form $\dot{a}_i = \eta_i R_i$, where $\dot{a}_i$ is the time derivative of the number of variants of technological equipment in sector i. R_i is the quantity of the combination of final "high-tech" and "low-tech" products used for research activities concerning new variants for sector i. η_i is a parameter determining the productivity of R&D. From investing one unit in R&D, η_i new variants will be discovered and the profit stream induced by them has a present value on the balanced growth path of $\frac{\eta_i \pi_i}{r}$. If the R&D sector is to be profit maximizing, it coordinates its research activities so that the present value of the profit streams for innovations to each sector equalize. This implies that on the balanced growth path $\eta_H \pi_H = \eta_L \pi_L$ is satisfied, which might be called the technological market clearing condition. Using equation (2.21) it turns out that on the balanced growth path it must be true that

$$\frac{a_H}{a_L} = \frac{\eta_H}{\eta_L} \frac{w_H H_H + w_L L_H}{w_H H_L + w_L L_L}, \tag{2.22}$$

which is analogous to the optimality condition for the distribution of variants of technological equipment. To see what effect a change of the skill composition and a change of the wage structure have on the development of new variants of technological equipment on the balanced growth path, it is useful to totally differentiate equation (2.22). This leads to

$$\hat{a}_H - \hat{a}_L = (\omega_H - \omega_L) \frac{(\sigma - 1)(1 - \beta)}{(\sigma - 1)(1 - \beta) - 1} (\hat{w}_H - \hat{w}_L) + \hat{H}_H - \hat{H}_L, \tag{2.23}$$

where the result (2.12) is used and the fact that, if high and low skilled wages are each the same in both sectors, then $\hat{H}_H - \hat{L}_H = \hat{H}_L - \hat{L}_L$ has to be fulfilled. Further it is useful to note again that the terms $\omega_H - \omega_L$ and $(\sigma - 1)(1 - \beta) - 1$ carry the opposite sign (as long as $\mu_H > \mu_L$ is satisfied). Equation (2.23) says first that, if high and low skilled intermediate products are gross compliments, a rise in the relative wage of the high skilled leads to a rise in the relative state of the "high-tech" sector technology. If "high-tech" and "low-tech" products are gross substitutes the

[6]The choice of the preferences of the consumers has no influence on the results since they are determined alone by the production side of the economy.

opposite occurs. But more importantly if the term $\hat{H}_H - \hat{H}_L$ is positive, i.e. the high skilled work force in the "high-tech" sector grows faster than in the "low-tech" sector, this leads unambiguously to an increase of the relative state of technology of the "high-tech" sector. From equation (2.18) it is obvious that it is still to be determined how relative prices react in response to changes in the skill composition in order to draw conclusions about the net effect on wages. For this it is necessary to make some assumptions about how the output of the "high-tech" and "low-tech" sector is used in the economy. Assume that in a very final stage the output of the "high-tech" and the "low-tech" sectors is combined to yield a final good Y which is then used for consumption and R&D activities. The price for this final good is assumed to equal its marginal costs[7] and is normalized to one.

$$Y = [\delta Y_H^{\alpha} + (1-\delta) Y_L^{\alpha}]^{\frac{1}{\alpha}} , \tag{2.24}$$

where the parameter δ determines how important the high-tech and low-tech products are in the production of the final good Y. The parameter α is assumed to lie in the interval $(-\infty, 1]$. Equation (2.24) then implies relative prices given that the two types of goods are used in a cost minimizing way

$$\frac{P_H}{P_L} = \frac{\delta}{1-\delta} \left(\frac{Y_H}{Y_L} \right)^{-(1-\alpha)} . \tag{2.25}$$

Now using the production function in reduced form, equation (2.15), the present value of the discovery of new variants (2.20) and the technological market clearing condition, the relative price on the balanced growth path can be written as a function of technological terms:

$$\frac{P_H}{P_L} = \left(\frac{\delta}{1-\delta} \right)^{\frac{\varepsilon}{\varepsilon-1}} \left(\frac{a_H}{a_L} \frac{\eta_L}{\eta_H} \right)^{-\frac{1}{\varepsilon-1}} . \tag{2.26}$$

In (2.26) $\varepsilon = \frac{1}{1-\alpha}$ denotes the elasticity of substitution between "high-tech" and "low-tech" products in the final stage of production. The important implication is

[7] This assumption can be justified within a framework of many firms engaging in the production of Y which stand in perfect competition with each other.

that if “high-tech” and “low-tech” products are gross substitutes in production of the final good, then the relative price of high-tech products depends negatively on the ratio of the states of technology in the “high-tech” and “low-tech” sector. To use equation (2.26) to complete the analysis of this section it is necessary to compute the relationship between the growth rates of the variables

$$\hat{P}_H - \hat{P}_L = -\frac{1}{\varepsilon - 1}(\hat{a}_H - \hat{a}_L). \tag{2.27}$$

With this result at hand the effect of a change in the skill composition of the working population via technological adjustment of the economy on relative wages can be computed. This is done using equation (2.18), the reaction of wages in response to technological and price changes; equation (2.23), the reaction of the sector technologies in response to changes in the skill composition of the working population; and equation (2.27), the reaction of prices.

There exists a balanced growth path on which consumption, final output, sector output and the number of variants of technological equipment in the two sectors grow with the same constant rate. Provided that the parameters of the model fulfill the condition stated below, this balanced growth path is stable.

$$\varepsilon < 2 + \frac{\beta}{1-\beta} + \frac{1}{\phi}\left(\sigma - 2 - \frac{\beta}{1-\beta}\right),$$

where ϕ is always negative, depends on the levels of endogenous variables of the model but stays constant on the balanced growth path. See appendix A at the end of the chapter for details. In response to a shock in the skill composition of the working population the relative wage of the high skilled adjusts by

$$\hat{w}_H - \hat{w}_l = \frac{(\sigma-1)(1-\beta)-1}{\omega_H - \omega_L}\frac{(\varepsilon-1)(1-\beta)-1}{(1-\beta)(\sigma-\varepsilon)}(\hat{H}_H - \hat{H}_L), \tag{2.28}$$

to reach the new steady state value. Equation (2.28) follows from the results (2.18), (2.23) and (2.27).

Since the first term on the right hand side of (2.28) is always negative, the sign of the second term determines whether the relative wage of the high skilled rises when $\hat{H}_H - \hat{H}_L$ is positive. There are several cases where this can occur: First, if the elasticity of substitution in the sector production is smaller than in the final stage, $\sigma < \varepsilon$, this requires ε to be larger than $2 + \frac{\beta}{1-\beta}$. Second, if the elasticity of substitution in the final stage does not possess such a high value, $\varepsilon < 2 + \frac{\beta}{1-\beta}$, the elasticity in the sector production has to be larger than in the final stage, $\sigma > \varepsilon$. Figure 2.3 shows the parameter regions where the mentioned rise in the relative wage of the high skilled can occur.

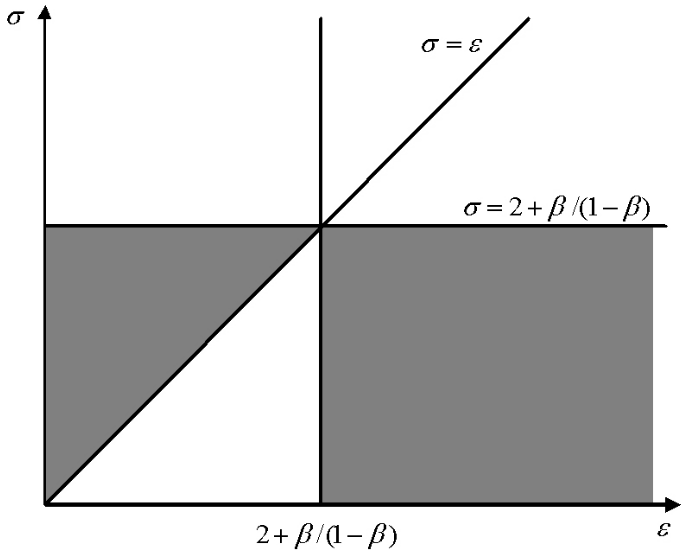

Figure 2.3: Elasticity of substitution

Regions (shaded) for the elasticities of substitution where a rising relative wage for the high skilled occurs when the high skilled working population in the ”high-tech” sector grows faster than in the ”low-tech” sector.

The empirical evidence for this result is also supportive. Again the NBER-CES Manufacturing Database compiled by the National Bureau fo Economic Research and the US Census Bureau’s Center for Economic Studies is used. High and low skilled workers are again identified by using the non-production and production status of employees. The model presented above distinguishes between only between two sectors, one “high-tech” and “low-tech” sector. In reality there are many more sectors and this distinction can hardly be made. As a proxy for the degree of correspondence to the "high-tech" sector of the economy, the ratio of non-production to production workers is used; a high ratio implies a high affinity to the "high-tech" section. A testable hypothesis of the model is that industries belonging more to the "high-tech" section should have experienced a higher growth rate of high skilled employment than industries belonging more to the "low-tech" section.

The data is used in a rolling cross section regression of the 10 years continuously compounded growth rate of the non-production employment in each industry on the log of the ratio of non-production to production labor 10 years ago for all 459 industries in the years from 1968 to 1996. The coefficient of the lagged log ratio of non-production to production workers is interpreted to represent the relationship (2.28), which should be positive in order for the economic story of the model to be true. This analysis should be seen more as descriptive than as a standard econometric estimation of a structural economic model since nothing in the above argument demands the estimated coefficient to be constant over time. The resulting development of the coefficient, together with a 95% confidence corridor, is displayed in figure 2.4. Standard errors are computed using the White covariance estimator which is robust with respect to industry heterogeneity in the error term of the regression.

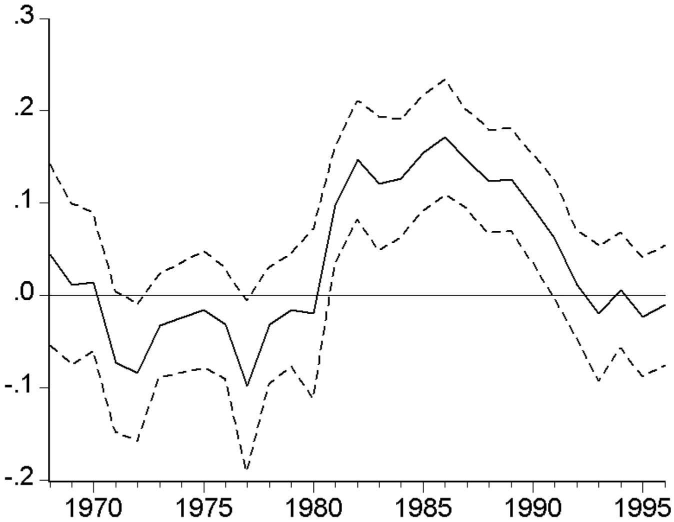

Figure 2.4: Growth of non-production employment and "high-tech" affinity

Development of the slope coefficient (with 95% confidence corridor) of a rolling window regression of the 10 years growth rate of non-production employment and the 10 years lagged log ratio of non-production to production employment.

Comparing figure 2.4 with figure 2.2 gives the following insights. The relationship between non-production employment growth and "high-tech" affinity is significantly positive from 1981 to 1991. The steepest increase in relative non-production employment is starting a little bit earlier but in general there is a coincidence with the above relationship to be positive. Thus this descriptive exercise is supportive of the outcome of the model, if the model is in accordance with both a higher growth rate for high skilled employment in the "high-tech" sector and an overall increase in high skilled labor supply. This is the question which is addressed next.

Using the identities for the total number of high and low skilled workers, $H = H_H + H_L$ and $L = L_H + L_L$, it turns out that the development of the relative wage is given by

$$\hat{w}_H - \hat{w}_L = \frac{[(\sigma - 1)(1-\beta) - 1][(\epsilon - 1)(1-\beta) - 1]}{(\omega_H - \omega_L)(\sigma - \varepsilon)(1-\beta) - \frac{(\varepsilon - 1)(1-\beta) - 1}{\frac{H_H}{H} - \frac{L_H}{L}}} \left(\hat{H} - \hat{L}\right).$$

If the above stated stability condition is to be fulfilled, the only possibility for this relationship to be positive is the case where the elasticity ε is larger than $2 + \frac{\beta}{1-\beta}$ which is a standard result for this type of model specification (see Acemoglu 1998).

Sectoral Technological Change with the Knowledge-Based R&D Specification

In this section a more general formulation for the innovation possibility frontier will be used. The so called knowledge-based R&D specification of Rivera-Batiz and Romer (1991) makes assumptions about the state dependence of the productivity of research activities. The difference with respect to the lab equipment specification in the preceding section is that R&D is now conducted by scientists rather than by use of final output. A natural assumption is that these resources are a scarce factor and can not be accumulated over time. Then, in order to achieve sustainable growth in the steady state, there must be state dependence in the R&D process, i.e. past discoveries must create a positive spill-over effect onto current research. This process, often illustrated by the metaphor that scientists can "stand on shoulders of giants", leads to the ever increasing productivity of scientists giving rise to a constant growth rate in the number of variants of technological equipment. A flexible formulation, also used by Acemoglu (2001), for the innovation possibility frontier is

$$\dot{a}_H = \eta_H a_H^{\frac{1+\kappa}{2}} a_L^{\frac{1-\kappa}{2}} S_L \quad \text{and} \quad \dot{a}_L = \eta_L a_L^{\frac{1+\kappa}{2}} a_H^{\frac{1-\kappa}{2}} S_H. \tag{2.29}$$

Here S_H and S_L are the numbers of scientists engaging in the discovery of new variants of technological equipment for the "high-tech" and "low-tech" sectors. It is assumed that $S_H + S_L = \bar{S}$ and that $\bar{S}$ is constant over time. The parameter $\kappa \in [0, 1]$ determines the degree of state dependence. If $\kappa = 0$ then there is no state dependence and neither does an increase in a_H or a_L make R&D activities of scientists relatively more productive. The opposite case is where there is extreme state dependence and current research in one sector makes future R&D in that sector relatively more productive. This alternative specification of the innovation possibility frontier has an important impact on the technological market clearing condition. This condition states that the impact of one researcher should lead in both sectors to the same profits, $\eta_H a_H^\kappa \pi_H = \eta_L a_l^\kappa \pi_L$. Note that the case of no state dependence leads to the same market clearing condition as in the case of the lab equipment specification and therefore to the same result. On the balanced growth path where each type of R&D is equal profitable, equation (2.22) now becomes

$$\frac{a_H}{a_L} = \left(\frac{\eta_H}{\eta_L}\frac{w_H H_H + w_L L_H}{w_H H_L + w_L L_L}\right)^{\frac{1}{1-\kappa}}. \tag{2.30}$$

Changes in the relative wage bill of the sectors now lead to larger effects on the relative state of the sector technology. Totally differentiating equation (2.30) yields

$$\hat{a}_H - \hat{a}_L = \frac{(\omega_H - \omega_L)}{(\sigma-1)(1-\beta)-1}\frac{(\sigma-1)(1-\beta)}{1-\kappa}(\hat{w}_H - \hat{w}_L) + \frac{1}{1-\kappa}(\hat{H}_H - \hat{H}_L), \tag{2.31}$$

which now replaces equation (2.23). Relative prices in terms of the technology parameters and variables are now determined by

$$\frac{P_H}{P_L} = \left(\frac{\delta}{1-\delta}\right)^{\frac{\varepsilon}{\varepsilon-1}} \left(\frac{a_H}{a_L}\right)^{-\frac{1-\kappa}{\varepsilon-1}} \left(\frac{\eta_H}{\eta_L}\right)^{\frac{1}{\varepsilon-1}}, \tag{2.32}$$

and therefore equation (2.27) now becomes

$$\hat{P}_H - \hat{P}_L = -\frac{1-\kappa}{\varepsilon-1}(\hat{a}_H - \hat{a}_L). \tag{2.33}$$

Finally the relationship between relative wages, the relative state of technology and the skill composition of the workforce, determined by the unit costs of producing in the two sectors, equation (2.18), remains valid. To put things together, the reaction of the relative wage for the high skilled in the steady state in response to a change in the skill composition of the workforce is now given by (2.18), (2.31) and (2.33). In reduced form this yields

$$\begin{aligned} \hat{w}_H - \hat{w}_L =& \frac{(\sigma-1)(1-\beta)-1}{\omega_H - \omega_L} \times \\ & \times \frac{(\varepsilon-1)(1-\beta)-(1-\kappa)}{(1-\beta)[(\sigma-\varepsilon)(1-\kappa)-(\varepsilon-1)(1-\beta)(\sigma-1)\kappa]}(\hat{H}_H - \hat{H}_L). \end{aligned} \tag{2.34}$$

For the relative wage of the high skilled to rise, if $\hat{H}_H - \hat{H}_L$ is positive, there are two possibilities to consider. First, if $\varepsilon > 1 + \frac{1-\kappa}{1-\beta}$ then the denominator of the second term on the right-hand side of equation (2.34) has to be negative. This leads to the condition $(\sigma - \varepsilon)(1 - \kappa) - (\varepsilon - 1)(\sigma - 1)\kappa < 0$. If instead the elasticity of substitution in the final stage of production is smaller, i.e. $\varepsilon < 1 + \frac{1-\kappa}{1-\beta}$, then the opposite has to be true. These are unfortunately highly non-linear restrictions on the parameters of the model and the stability conditions for all values of κ are not always easily interpretable. I therefore abstract in the following from the general formulation and focus on the special case of extreme state dependence, $\kappa = 1$[8]. Appendix B at the end of the chapter shows that for stability the condition $\varepsilon < 1$ has to be fulfilled. Now turning to equation (2.34). If $\kappa = 1$, the second term of the right-hand side of equation (2.34) becomes $-\frac{1}{(\sigma-1)(1-\beta)}$. Therefore the relative wage of the high skilled increases if σ is larger than one if the growth rate of the high skilled workforce in the "high-tech" sector is larger than in the "low-tech" sector, provided the stability condition is satisfied. Reducing equation (2.34) to a relationship between the development of the relative wage and the change in the total skill composition of the workforce yields

$$\hat{w}_H - \hat{w}_L = -\frac{(\sigma - 1)(1 - \beta) - 1}{(\omega_H - \omega_L)(\sigma - 1)(1 - \beta)\left(\frac{H_H}{H} - \frac{L_H}{L}\right) - 1}\left(\hat{H} - \hat{L}\right).$$

For this relationship to be positive it requires the elasticity σ to be greater than $1 + \frac{1}{(\omega_H - \omega_L)(1-\beta)(H_H/H) - L_H/L)}$. Note that $\frac{H_H}{H} - \frac{L_H}{L}$ is always positive and that σ also has to fulfill the condition $\sigma < 2 + \frac{\beta}{1-\beta}$, so that this case is only to occur if the "high-tech" and "low-tech" sectors are very intensively using the corresponding type of labor. Finally, in the equilibrium, consumption, output and the number of variants of technological equipment grow at the same constant rate.

[8] Note that the other extreme of no state dependence gives the same results as in the lab equipment specification.

2.4 A Special Case

In this section a special case of the derived model is examined and related to the existing literature. Acemoglu (1998) and (2001) use models of directed technological change which can nicely be nested into the above framework[9]. First, in these articles it is assumed that there are two sectors in the economy. One uses only high skilled workers, the other only low skilled. This can be achieved in the present model using the lab equipment R&D specification by setting the elasticity of substitution of sector production to a value greater than $2+\frac{\beta}{1-\beta}$ and imposing that the productivity parameter μ_L equals zero. This obviously leads to the “low-tech” sector using only low skilled workers. Cost minimization with respect to the parameter γ_H then leads to a corner solution for the “high-tech” sector. If unemployment is ruled out, the only possibility is that the “high-tech” sector only employs high skilled workers. Consequently all variants of technological equipment will be used only with the one type of labor. These assumption simplify equation (2.23) to yield

$$\hat{a}_H - \hat{a}_L = \hat{w}_H - \hat{w}_L + \hat{H}_H - \hat{L}_L. \tag{2.35}$$

Together with the results (2.18) and (2.27) now the reaction of wages in response to a shock in the skill composition of the working population becomes

$$\hat{w}_H - \hat{w}_L = [(\varepsilon - 1)(1-\beta) - 1](\hat{H}_H - \hat{L}_L). \tag{2.36}$$

This is exactly the final result of the above cited articles: The relative wage for the high skilled increases with a growing relative number of high skilled workers if the elasticity of substitution in the final stage of production is larger than $2 + \frac{\beta}{1-\beta}$.

[9]Acemoglu (1998) uses the quality ladder approach instead of a growth model with horizontal differentiation. However his model can be formulated with horizontal differentiation and still leads to the same result.

2.5 Directed Technological Change in the Open Economy

So far, the analysis concentrated on the case of the closed economy. This section will focus on an open economy specification of the model. A relevant scenario for this is that the relative state of technology measured by $\frac{a_H}{a_L}$ is a global variable, i.e. it is identical for all countries. The incentives for innovations are set worldwide by all the potential users of the variants of technological equipment regardless of which country they come from. Another reasonable assumption is that wages for high and low skilled workers are determined in local labor markets; the factor labor is immobile between countries. Unfortunately this realistic setup is in general too complicated to be analyzed within the above model. It is however possible to examine some special cases which are also quite interesting. The following analysis focuses on the two country case where the two countries can freely trade "high-tech" and "low-tech" products. The first scenario to be examined is the case of two identical countries having the same relative and absolute supply of high and low skilled workers in both sectors at the beginning. The second scenario is that of one small and one large country, where small means that the relative incentives for innovations can be approximated by the situation in the larger country.

2.5.1 Two Identical Countries

At the outset, both countries face the same situation. The relative state of sector technology is globally determined and is therefore identical for both countries. Furthermore, the two countries have the same relative and absolute supply of high and low skilled workers and hence the same skill structure for both sectors. This means that the relative distribution of technological equipment is identical in the two countries and the costs of production in the two sectors are identical as are the wages for high and low skilled employees. In addition to this, it is assumed that all parameters determining the production processes in the two economies are identical too. To begin the analysis, turn to the innovation possibility frontier. For reasons of tractability, this section will only use the lab equipment specification. Taking into account the global market for innovations, the technology market clearing condition is now given by

$$\frac{a_H}{a_L} = \frac{\eta_H}{\eta_L} \frac{w_H^1 H_H^1 + w_L^1 L_H^1 + w_H^2 H_H^2 + w_L^2 L_H^2}{w_H^1 H_L^1 + w_L^1 L_L^1 + w_H^2 H_L^2 + w_L^2 L_L^2}, \tag{2.37}$$

where the super-script 1 and 2 denotes country one and country two. Note that at the beginning of the analysis, the corresponding variables for the two countries have the same value. Totally differentiating equation (2.37) yields

$$\begin{aligned} \hat{a}_H - \hat{a}_L = \frac{1}{2}(\omega_H - \omega_L)\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}(\hat{w}_H^1 - \hat{w}_L^1 + \hat{w}_H^2 - \hat{w}_L^2) + \\ + \frac{1}{2}\left(\hat{H}_H^1 - \hat{H}_L^1 + \hat{H}_H^2 - \hat{H}_L^2\right), \end{aligned} \tag{2.38}$$

which is quite analogous to equation (2.23) for the closed economy case. Here the relative state of technology is influenced by the development of relative wages and the supply of high skilled workers in the two sectors of the two countries. Consider the development of relative prices. Since the products of sector production can be traded freely, their prices have to be identical in the two countries

$$\frac{P_H}{P_L} = \frac{\delta}{1-\delta}\left(\frac{Y_H^1 - E_H^1}{Y_L^1 + E_L^2}\right)^{-\frac{1}{\varepsilon}} = \frac{\delta}{1-\delta}\left(\frac{Y_H^2 + E_H^1}{Y_L^2 - E_L^2}\right)^{-\frac{1}{\varepsilon}}. \tag{2.39}$$

In equation (39) the two terms E_H^1 and E_L^2 denote the net exports of "high-tech" and "low-tech" products of country one and country two respectively[10]. Note that at the beginning the exports are both equal to zero since the two economies are identical. This, however, changes when the skill structures of the economies change. From these two equations three conditions for the growth rate of the variables of interest can be deducted. First, there is a condition guaranteeing the equality of prices in the two economies, which determines the necessary net exports of the intermediate

[10] Of course the choice of the origins of the exports is arbitrary, so no assumptions about the signs of the two terms are made.

products

$$-\frac{dE_H^1}{Y_H^1}+\frac{dE_L^2}{Y_L^2}=-\frac{1}{2}(\omega_H-\omega_L)\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}\left[(\hat{w}_H^1-\hat{w}_L^1)-(\hat{w}_H^2-\hat{w}_L^2)\right]-\\ -\frac{1}{2}\left[(\hat{H}_H^1-\hat{H}_L^1)-(\hat{H}_H^2-\hat{H}_L^2)\right]. \tag{2.40}$$

Since there are no net exports and imports before changes in the skill composition of the working populations take place, the terms dE_H^1 and dE_L^2 give the absolute value of exports of the two intermediate goods. Furthermore, before the shock occurs $Y_H^1=Y_H^2$ and $Y_L^1=Y_L^2$ is true. Equation (2.40) says that the sum of the export shares of production of "high-tech" and "low-tech" products is a function of the development of relative wages and the relative composition of the high skilled workforce of the two countries. If the relative wage for the high skilled in country one increases faster than in country two, the export shares increase if high and low skilled intermediate products are gross compliments in the sector production. They decrease in the case of these intermediate products being gross substitutes. If the ratio of high skilled workers in the high-tech sector to the high skilled workers in the low-tech sector grows faster in country one than in country two, this unambiguously leads to an increase in the trade activities between these two economies. In addition to the equal price condition, there are two conditions showing the development of prices in terms of the variables of the two countries

$$\hat{P}_H-\hat{P}_L=-\frac{1}{\varepsilon-1}\Bigg((\omega_H-\omega_L)\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}(\hat{w}_H^1-\hat{w}_L^1)+\\ +\hat{H}_H^1-\hat{H}_L^1-\frac{dE_H^1}{Y_H^1}+\frac{dE_L^2}{Y_L^2}\Bigg), \tag{2.41}$$

$$\hat{P}_H-\hat{P}_L=-\frac{1}{\varepsilon-1}\Bigg((\omega_H-\omega_L)\frac{(\sigma-1)(1-\beta)}{(\sigma-1)(1-\beta)-1}(\hat{w}_H^2-\hat{w}_L^2)+\\ +\hat{H}_H^2-\hat{H}_L^2+\frac{dE_H^1}{Y_H^1}-\frac{dE_L^2}{Y_L^2}\Bigg). \tag{2.42}$$

Whether a rise in the relative wage of the high skilled in one country leads to an increasing relative price depends on whether "high-tech" and "low-tech" products are gross substitutes or compliments and whether the elasticity of substitution in the final stage of production is larger or smaller than one. If this elasticity is smaller than one in absolute value, a rise in the ratio between the high skilled workers in both sectors leads to a rising relative price of "high-tech" products. Exports of "high-tech" and "low-tech" products are working to equalize the development of prices in both countries as described by equation (2.40). Finally, to close this two economy model, still two additional equations are needed. As in the model of the closed economy, these are the two total differentials of the unit cost function (2.18) for the two countries. Now using equations (2.38), (2.40), (2.41), (2.42) and equation (2.18) for both economies leads to the following result. The development of the relative wage for the high skilled is identical in both countries and this development is given by

$$\begin{aligned}\hat{w}_H^1 - \hat{w}_L^1 = \hat{w}_H^2 - \hat{w}_L^2 = & \frac{1}{2}\frac{(\sigma-1)(1-\beta)-1}{\omega_H-\omega_L}\frac{(\varepsilon-1)(1-\beta)-1}{(\sigma-\varepsilon)(1-\beta)}\times \\ & \times\left[\left(\hat{H}_H^1-\hat{H}_L^1\right)+\left(\hat{H}_H^2-\hat{H}_L^2\right)\right].\end{aligned} \tag{2.43}$$

This result is analogous to the development of the relative wage in the closed economy using the lab equipment specification for the innovation possibility frontier and the same arguments apply. The same conditions apply especially to the relative wage to rise in response to a higher growth rate of the high skilled working population in the high-tech sector than in the low-tech sector.

2.5.2 One large and one small Country

This section deals with the situation of one large and one small economy engaging in trade with each other. As in the preceding section "high-tech" and "low-tech" products can be freely exchanged. The parameters of the model are identical for the two economies but their supply of high and low skilled workers now differs. Assuming that the larger economy is relatively more important, the incentives for innovations can be approximated solely by the profits obtained in the larger economy. New variants of technological equipment are produced again by the lab equipment specification. Therefore the relative state of sector technology is determined by an equation of the type of (2.22) for the large economy. Consequently the relative price of high-tech and low-tech products is also determined in the larger country. From this it directly follows that for the larger economy all results of the closed economy apply. However with respect to the smaller country things are quite different. A changing skill composition in the small economy now does not have any effect on the relative state of technology, but a changing skill composition in the larger country has spill-over effects onto the smaller country via a change in the state of the relative sector technology. For the small economy this leads to a change in the relative wage in response to a change in the skill composition of the larger country. Using equations (2.18), (2.23) and (2.27) this effect is

$$\hat{w}_H^S - \hat{w}_L^S = \frac{(\sigma-1)(1-\beta)-1}{\omega_H^S - \omega_L^S} \frac{(\varepsilon-1)(1-\beta)-1}{(\sigma-\varepsilon)(1-\beta)} \left(\hat{H}_H^L - \hat{H}_L^L\right), \tag{2.44}$$

where the super script “S” and “L” denote the variables of the smaller and the larger country. Equation (2.44) is analogous to the closed economy case and the same arguments apply. Therefore, if one country is the technology leader the effects of a change in the skill composition in this economy carry over to the small country.

2.6 Conclusion

The model presented in this paper has examined the impact of changes in the skill composition of the workforce in different sectors via induced technological change on the relative wage of the high skilled workers. The direction of this technological change is endogenously determined and can have a different skill and sector specific component. The bias in the development of relative technology for skills and sectors comes from the different profitability of new discoveries with respect to sectors and the distribution of them with respect to skill groups.

Endogenizing the innovation process with the lab equipment specification for R&D yields nice results which can easily be interpreted. For the relative wage of the high skilled to rise in response to a higher growth rate in the high skilled workers in the "high-tech" sector than in the "low-tech" sector, all that matters is the sign of the difference of the elasticity of substitution in the sector production and the final production stage. Furthermore, the model is stable for a reasonable range of parameter constellations. In addition, considering the total number of high skilled to rise, this requires a large elasticity of substitution in the final production stage as usual in this kind of models.

However, it gets more complicated using a more flexible formulation of the innovation possibility frontier. With the so called knowledge-based R&D specification one introduces an additional parameter. The model now has four exogenous elasticities and the conditions for stability and the aforementioned rise in the relative wage of the high skilled all involve a non-linear combination of three parameters. Examining the special case of extreme state dependence has shown that nevertheless there exists a parameter region which leads to a rising relative wage and guarantees a stable balanced growth path of the model. Furthermore, it has been shown that quite popular models of the literature on directed technological change can be seen as special cases of the presented model. Although the case of the open economy is of major interest, only some special cases can be examined within the model of this chapter. It has been shown that in the case of two identical economies, the effects of a change in the skill composition of the work force are smaller but carry over from one country to another symmetrically. The effect on wages is the same in both countries because the relative state of sector technology is determined globally and both economies can trade in sector products. If the relative state of sector technology is determined solely in one large country, effects of a change in the skill composition of the large country spill over onto the small country.

2.7 Appendix A: Stability Conditions for the Lab Equipment Specification

To simplify the analysis it is assumed that the labor market is always in equilibrium, i.e. the relative wage is always the same in both sectors. Furthermore it is assumed that relative prices in the economy adjust instantaneously. Free entry into the R&D sector implies that the reward for research is not larger than the cost, $1 \geq \max[\eta_H V_H, \eta_L V_L]$. The case $1 > \eta_i V_i$ only occurs, if and only if, $R_i = 0$ and consequently $\dot{a}_i = 0$. In order for the variants of technological equipment to expand in both sectors one would need $1 = \eta_i V_i(t)$ for an interval of time. Now, from dynamic programming it follows that $r(t)\eta_i V_i(t) = \eta_i \pi_i(t) + \eta_i \dot{V}_i(t)$ and that $1 = \eta_H V_H(t) = \eta_L V_L(t)$ is only possible for $0 = \dot{V}_H(t) = \dot{V}_L(t)$. For this to be true it would be necessary that $\eta_H \pi_H(t) = \eta_L \pi_L(t)$ for that interval of time. However, from equation (2.15), (2.20) and (2.25) it follows that always

$$
\begin{aligned}
\frac{P_H}{P_L} &= \left(\frac{\delta}{1-\delta}\right)^{\frac{1-\beta}{1-\alpha\beta}} \left[\frac{\left(L_H^{(\sigma-1)(1-\beta)} + \mu_H^{\sigma} H_H^{(\sigma-1)(1-\beta)}\right)^{\frac{1}{(\sigma-1)(1-\beta)}}}{\left(L_L^{(\sigma-1)(1-\beta)} - \mu_L^{\sigma} H_L^{(\sigma-1)(1-\beta)}\right)^{\frac{1}{(\sigma-1)(1-\beta)}}} \frac{a_H}{a_L}\right]^{-\frac{(1-\alpha)(1-\beta)}{1-\alpha\beta}} \\
&\equiv \left(\frac{\delta}{1-\delta}\right)^{\frac{1-\beta}{1-\alpha\beta}} \left(\frac{\tilde{H}}{\tilde{L}} \frac{a_H}{a_L}\right)^{-\frac{(1-\alpha)(1-\beta)}{1-\alpha\beta}}, \qquad (2.45) \\
\frac{\pi_H}{\pi_L} &= \left(\frac{\delta}{1-\delta}\right)^{\frac{1}{1-\alpha\beta}} \left(\frac{a_H}{a_L}\right)^{-\frac{1-\alpha}{1-\alpha\beta}} \left(\frac{\tilde{H}}{\tilde{L}}\right)^{\frac{\alpha(1-\beta)}{1-\alpha\beta}}. \qquad (2.46)
\end{aligned}
$$

Totally differentiation of $\frac{\tilde{H}}{\tilde{L}}$ gives

$$
\hat{\tilde{H}} - \hat{\tilde{L}} = \frac{\omega_H - \omega_L}{(\sigma-1)(1-\beta)-1}(\hat{w}_H - \hat{w}_L) + \hat{H}_H - \hat{H}_L. \qquad (2.47)
$$

The adjustment of wages is given by equation (2.18)

$$
\hat{w}_H - \hat{w}_L = \frac{1}{\omega_H - \omega_L}\left[\frac{1}{1-\beta}\left(\hat{P}_H - \hat{P}_L\right) + \hat{a}_H - \hat{a}_L\right], \qquad (2.48)
$$

and prices react according to equation (2.49) by

$$
\hat{P}_H - \hat{P}_L = -\frac{(1-\alpha)(1-\beta)}{1-\alpha\beta}\left(\hat{\tilde{H}} - \hat{\tilde{L}} + \hat{a}_H - \hat{a}_L\right). \qquad (2.49)
$$

Now using equations (2.47), (2.48) and (2.49), the change in $\frac{\tilde{H}}{\tilde{L}}$, given the relative state of technology ($\hat{a}_H - \hat{a}_L = 0$), can be computed as

$$\hat{\tilde{H}} - \hat{\tilde{L}} = \frac{[(\sigma-1)(1-\beta)-1]\,(1-\alpha\beta)}{[(\sigma-1)(1-\beta)-1]\,(1-\alpha\beta)+1-\alpha}\left(\hat{H}_H - \hat{H}_L\right), \tag{2.50}$$

and the corresponding change in the relative profitability of R&D is given by

$$\hat{\pi}_H - \hat{\pi}_L = \frac{[(\sigma-1)(1-\beta)-1]\,\alpha(1-\beta)}{[(\sigma-1)(1-\beta)-1]\,(1-\alpha\beta)+1-\alpha}\left(\hat{H}_H - \hat{H}_L\right). \tag{2.51}$$

Equation (2.50) shows how a shock to the skill composition affects $\tilde{H}$ and $\tilde{L}$. If $\hat{H}_H - \hat{H}_L$ is positive, then the relative change in the profits can either be positive or negative. It will turn out that this does not matter for the stability of the balanced growth path. If $\hat{\pi}_H - \hat{\pi}_L$ is positive (negative) this will give rise to technological adjustment $\hat{a}_H$ and $\hat{a}_L = 0$ ($\hat{a}_L > 0$ and $\hat{a}_H = 0$). Now assume that after the shock $\left(\hat{H}_H - \hat{H}_L > 0\right)$ occurred, the total high skilled and low skilled working population is constant, $dH_H + dH_L = dL_H + dL_L = 0$. The assumption that the labor market is always in equilibrium implies the conditions $\hat{w}_H - \hat{w}_L = [(\sigma-1)(1-\beta)-1]\left(\hat{H}_H - \hat{L}_L\right)$ and $\hat{H}_H - \hat{L}_H = \hat{H}_L - \hat{L}_L$. Using these conditions equation (2.50) now becomes

$$\hat{\tilde{H}} - \hat{\tilde{L}} = \hat{H}_H\left[(\omega_H - \omega_L)\left(1 - \frac{H}{L}\frac{L_L}{H_L}\right) + \frac{H}{H_L}\right], \tag{2.52}$$

$$\tag{2.53}$$

where $H = H_H + H_L$ and $L = L_H + L_L$. Equations (2.47), (2.48) and (2.49) now give the effects of the technological adjustment on $\frac{\tilde{H}}{\tilde{L}}$

$$\hat{\tilde{H}} - \hat{\tilde{L}} = \frac{\alpha(1-\beta)\phi}{[(\sigma-1)(1-\beta)-1](1-\alpha\beta)+(1-\alpha)\phi}\left(\hat{a}_H - \hat{a}_L\right), \tag{2.54}$$

with

$$\phi = \frac{(\omega_H - \omega_L)\left(\frac{H_L}{H} - \frac{L_L}{L}\right) + 1}{(\omega_H - \omega_L)\left(\frac{H_L}{H} - \frac{L_L}{L}\right)}.$$

It can be shown that ϕ is always positive. With (2.54) it is now clear from equation (2.46) that the relative profitability responds to technological adjustment by

$$\hat{\pi}_H - \hat{\pi}_L = \left(\frac{\alpha(1-\beta)}{1-\alpha\beta}\frac{\alpha(1-\beta)\phi}{[(\sigma-1)(1-\beta)-1](1-\alpha\beta)-(1-\alpha)\phi} - \frac{1-\alpha}{1-\alpha\beta}\right). \quad (2.55)$$

To stabilize the economy now $\hat{\pi}_H - \hat{\pi}_L$ has to be negative if $\hat{a}_H - \hat{a}_L$ is positive during adjustment and vice versa. But this requires in both cases the first term on the right hand side to be negative. For this to be true the following condition has to be satisfied

$$\varepsilon < 2 + \frac{\beta}{1-\beta} + \frac{1}{\phi}\left(2 + \frac{\beta}{1-\beta} - \sigma\right).$$

If this condition is satisfied, the ratio $\frac{\pi_H}{\pi_L}$ always returns to its equilibrium level $\frac{\eta_L}{\eta_H}$ after a shock occurs.

In equilibrium it is then true that

$$\frac{\eta_L}{\eta_H} = \left(\frac{\delta}{1-\delta}\right)^{\frac{1}{1-\alpha\beta}} \left(\frac{a_H}{a_L}\right)^{-\frac{1-\alpha}{1-\alpha\beta}} \left(\frac{\tilde{H}}{\tilde{L}}\right)^{\frac{\alpha(1-\beta)}{1-\alpha\beta}} \quad (2.56)$$

has to be fulfilled. But then it must also be true that $\hat{a}_H - \hat{a}_L = (\varepsilon - 1)(1 - \beta)\left(\hat{\tilde{H}} - \hat{\tilde{L}}\right)$. At the same time equation (2.54) has to be satisfied. In general this system of two equations has only the solution $\hat{a}_H - \hat{a}_L = \hat{\tilde{H}} - \hat{\tilde{L}} = 0$. It has an infinite number of solutions if it happens that $\varepsilon = 2 + \frac{\beta}{1-\beta} + \frac{1}{\phi}\left(2 + \frac{\beta}{1-\beta} - \sigma\right)$ which is the borderline of the stability conditions. Also $\hat{\tilde{H}} - \hat{\tilde{L}} = 0$ implies that $\hat{H}_H = \hat{H}_L$ (equation (2.50)). If H is to be constant, then this is only possible for $\hat{H}_H = \hat{H}_L = \hat{L}_H = \hat{L}_L$ which leads directly to $\hat{\tilde{H}} = \hat{\tilde{L}} = 0$. In equilibrium therefore a_H and a_L grow at the same rate θ and $\tilde{H}$ and $\tilde{L}$ do not change. From the production functions (2.15) and (2.24) it can be seen that then also Y_H, Y_L and Y have to grow at this rate if prices are constant. However, they must be constant since the relative price in equilibrium can be expressed as a function of $\frac{a_H}{a_L}$ (equation (2.26)) and the price of final output is normalized to one. The produced final output is used to finance wages and the investments in technological equipment. Since the demand for technological equipment is linear in sector output (equation (2.16)) these investments grow at the rate θ. Consumers receive interest payments as a reward for their savings. The zero profit condition for the R&D sector implies that all profits must be used for these interest payments. Since these profits are a fixed proportion of the turnover of the monopolists, which in turn is linear in the sector output, they must grow at rate θ. Therefore, the total income of the consumers, which equals total output minus investments in technological equipment plus interest payments, must grow also with rate θ. It is assumed that consumers have a constant marginal propensity to save which implies that consumption and savings grow at the same rate as income. On the balanced growth path savings are divided into spending for R&D leading to equal growth rates of the number of variants which implies $\eta_h \frac{R_H}{a_H} = \eta_L \frac{R_L}{a_L}$ and all savings go into R&D, $S = R_H + R_L$. This yields $R_H = \left(\frac{\eta_L/a_L}{\eta_H/a_H + \eta_L/a_L}\right) S$ and $R_L = \left(\frac{\eta_H/a_H}{\eta_H/a_H + \eta_L/a_L}\right) S$. But a_H and a_L grow at the same rate so the first terms on the right-hand sides are constant on the balanced growth path and R_H and R_L grow at the rate θ. Finally from the innovations possibility frontier it can be seen that this growth rate has to be constant since $\frac{\dot{a}_i}{a_i} = \eta \frac{R_i}{a_i}$ and the right hand side is a constant because R_i and a_i grow with rate θ.

2.8 Appendix B: Stability Conditions for the Knowledge-Based R&D Specification

Essentially the same steps have to be taken as in the proof of Appendix A and the same arguments apply. The only difference is that equilibrium is now defined by

$$\frac{\eta_L}{\eta_H} = \left(\frac{a_H}{a_L}\right)^{\kappa} \frac{\pi_H}{\pi_L}. \tag{2.57}$$

To reach this equilibrium after a shock which lead to a positive $\hat{\pi}_H - \hat{\pi}_L$ and therefore to $\hat{a}_H > 0$ and $\hat{a}_L = 0$, $\hat{\pi}_H - \hat{\pi}_L$ has not only got to be negative but at least as large as necessary to compensate the effect $\kappa\left(\hat{a}_H - \hat{a}_L\right)$ from the adjustment process. The necessary condition therefore is

$$\left(\frac{\alpha(1-\beta)}{1-\alpha\beta}\frac{\alpha(1-\beta)\phi}{[(\sigma-1)(1-\beta)-1](1-\alpha\beta)-(1-\alpha)\phi} - \frac{1-\alpha}{1-\alpha\beta}\right) < -\kappa. \tag{2.58}$$

In the special case $\kappa = 1$ this leads to the condition $\varepsilon < 1$ to be satisfied if the model should be stable. If this condition is satisfied the economy arrives at the equilibrium condition given in equation (2.57) and stays there because then R&D for each scientist is equally profitable in both sectors. This demands that in equilibrium the following relation is satisfied

$$\hat{\pi}_H - \hat{\pi}_L = -\kappa\left(\hat{a}_H - \hat{a}_L\right). \tag{2.59}$$

But equation (2.55) of Appendix A also has to be fulfilled at all points in time. These two equations imply that if equation (2.58) is fulfilled with equality, which is the borderline case of stability, the system has a unlimited number of solutions. If this possibility is ruled out, the only solution is $\hat{\pi}_H - \hat{\pi}_L = \left(\hat{a}_H - \hat{a}_L\right) = 0$ which states that the number of variants grow at the same rate. Furthermore from the innovations possibility frontier it directly follows that this rate is constant. It is then easy to verify that output and consumption also grow at this rate.

Chapter 3

"Weak" Scale Effects and Wage Inequality

The last chapter dealt with wage inequality within a first generation endogenous growth model, where the implied "strong" scale effect played a crucial role for the results of the model. However, as mentioned in the introductory chapter, "strong" scale effects are looked at very critical by the literature (see Jones 1995b). This chapter examines whether the "strong" scale effect can be exchanged with the "weak" scale effect of growth models of the second generation, i.e. whether scale effects in per capita production are sufficient to yield results comparable with the ones from the previous chapter.

3.1 Introduction

The basic stylized facts about wage inequality mentioned in the introduction of the previous chapter apply here as well and are thus not replicated. Instead the application of growth mechanisms in the literature concerning technology and wage inequality is analyzed in detail.

Acemoglu (2001) argues that nothing is lost from the results using a first generation growth model if this is replaced by a second generation model. He provides theoretical evidence by formulating a model using the growth equation of Jones (1995a), which takes on the general form

$$\dot{N}_t = \delta N_t^{\phi} S_t^{\lambda}, \tag{3.1}$$

or

$$\frac{\dot{N}_t}{N_t} = \delta N_t^{\phi-1} S_t^{\lambda}, \tag{3.2}$$

where N_t is usually interpreted as giving the degree of differentiation in a model with horizontal growth and S_t denotes scientific work, i.e. scientists. δ, ϕ and λ are exogenous parameters determining the behavior of N_t over time. Note that the basic growth equation motivated by Romer (1990) is obtained by setting $\phi = \lambda = 1$ and interpreting S_t as human capital. Denote the constant growth rate of S_t as n, then the only growth rate for N_t that is constant in the long-run is given by

$$\frac{\dot{N}_t}{N_t} = \frac{\lambda}{1-\phi} n. \tag{3.3}$$

Acemoglu (2001) uses the special case $\lambda = 1$ due to computational tractability, which is still in accordance with the basic conclusions from Jones (1995a). From equation (3.3) it can be seen why $\phi < 1$ is a reasonable assumption for positive and stable long-run growth.

Looking closer at the above equations, it turns out, that in the long-run the "strong" scale effect asymptotical vanishes as the economy gets closer to the equilibrium given by (3.3). However, this is only the case in equilibrium, i.e. on a balanced growth path. As long as the economy is off the balanced growth path, equation (3.2) is valid and this growth rate does depend on the amount of resources devoted to R&D, S_t. Thus off the balanced growth path there is still the "strong" scale effect present.

The wage inequality derived from the model in chapter one, as well as in the literature cited above, is the outcome of both, the "strong" scale effect and the adjustment towards a new equilibrium during which technology for one sector grows faster. Therefore replacing the basic Romer (1990) assumption with the growth equation of Jones (1995a) essentially changes nothing, there are still the same forces at work.

The aim of this chapter is, on the one hand, to provide a robustness check of the

Acemoglu (2001) result, but not with the just mentioned hybrid growth model of Jones (1995a), which still exhibits "strong" scale effects, but with the "pure" second generation model of Young (1998). This model was introduced in the introductory chapter where it became clear that this model is free of any "strong" scale effects. Therefore the Young (1998) model is extended to the two sector case with heterogenous labor, i.e. high and low skilled workers.

It has also been argued that second generation growth models exhibit "weak" scale effects. Per capita production depends on the size of the relevant market, given by the extent of the labor force. A priori, this "weak" scale effect seems to be a candidate for replacing the growth effect in first generation models. But there are still price effects at work which are of the form of the usual substitution effects working against wage inequality if the relative supply of e.g. the high skilled labor force rises. It is to determine which force is under what circumstances stronger.

Since the integration the idea of "weak" scale effects into the modelling of wage inequality is to some extend new, several modelling strategies are proposed and it is checked whether the results are basically consistent. As other growth models, the Young (1998) model uses factor inputs to cover R&D expenditures. In a model with two types of labor, and three types of goods, i.e. sector goods produced by high or low skilled and final output, there are many possibilities which input can be used to cover R&D expenses. On the technology side there is essentially the choice between two different production functions, first, the specification originally introduced by Romer (1987) and used afterwards in Romer (1990) and many other publications, or second, the Dixit and Stiglitz (1977) approach, first interpreted as a production technology by Ethier (1982). Several combinations for production and R&D technologies are checked for consistency with the result from the theory of directed technical change, which links a rise in relative skill supply and a corresponding rise in wage inequality. The general conclusion is that this result is in accordance with growth models of the second generation type, although as in the model of chapter two, special requirements for the parameters of the model are necessary. Interesting insights in the mechanisms gives the production technology from Ethier (1982). In this article he introduced a formulation for the Dixit and Stiglitz (1977) index, which separates the elasticity of substitution between differentiated input factors and the returns to differentiation, as mentioned in the introductory chapter.

If the discussion is about wage inequality, often comparisons between the development in the US and the UK and in some continental European countries like Germany are made. The relative supply of high skilled workers also increased here strongly, although a little less than in the US (Beaudry and Green 2000), but without being accompanied by a drastic increase in wage inequality. In fact the ratio of the upper to the lower 10 percent percentile of the wage income distribution was 2.69 in 1983 and 2.32 in 1993 (OECD 1996), thus the wage inequality even decreased. By the same time the inequality in employment dramatically increased. Whereas the unemployment rate for high skilled was about 3 percent in 1983 and 4 percent in 1997, the unemployment rate for the unskilled was about 11 percent in 1983 but rose to 24 percent in 1997 (Schimmelpfennig 2000)[1].

In the light of the labor market experience in Germany, one would immediately think about the Krugman hypothesis (Krugman 1994). In this hypothesis it is claimed that unemployment emerges whenever the economy tries to suppress a growing wage inequality. This would meet the empirical figures reported for Germany. As this hypothesis is very popular because it is an appealing story, it has not been subjected to theories involving endogenous technology decisions, possibly causing wage inequality. Therefore in one of the later sections a version of this chapter's model is extended to cover the two country case, where one country has a free labor market without any rigidities where labor supply changes possibly lead to wage inequality. The other country is characterized by some exogenous labor market policy which is able to impose relative wage rigidity and hence suppress growing relative wage inequality. Both countries can trade with each other. The aim of this exercise is to see what additional assumptions are needed to replicate the Krugman hypothesis. It turns out that very strong assumption about price setting in the different markets of the two countries are needed, which seem rather implausible.

[1]Here "skilled" means at least a university degree and "unskilled" a minimum schooling degree without vocational training (Schimmelpfennig 2000).

As another extension, the influence of skill specific exogenous technology shocks, which are often motivating wage inequality, is examined. The model is formulated in a way that allows for technology shocks to spill over to some predetermined degree from sector to sector. It will be shown under what assumptions these shocks have transitory or permanent effects on wage inequality. Furthermore two models without any scale effects are presented and it will be shown that in this case the relative supply of high skilled can not serve as an explanation for wage inequality.

3.2 A Model with Ethier/Dixit and Stiglitz Production Technology

The basic model is an extension of the growth model without any “strong” scale effects developed by Young (1998). This originally covers the situation of a one sector economy with homogeneous labor. The extended model is one with two sectors and two goods, produced by different types of labor. One sector of the economy is called the low skilled sector and produces a low skilled product. The other sector is the high skilled sector producing a high skilled good. Both goods can be combined via an aggregation technology to yield a final good which is used for consumption purposes.

3.2.1 The High and Low Skilled Sector

The low (high) skilled sector uses as inputs differentiated intermediate input factors which are produced from raw labor. These differentiated inputs are aggregated according to the following production function

$$Y_i = N_i^{\nu - \frac{1-\alpha}{\alpha}} \left[\int_0^{N_i} (\lambda_j x_j)^\alpha dj \right]^{\frac{1}{\alpha}}, \tag{3.4}$$

where all figures correspond to the current time period t if not otherwise stated, time is discrete. This production function for sector i, $i = L, H$, where L denotes the low skilled sector and H denotes the high skilled sector, follows the idea of

Ethier (1982) who first proposed separating the elasticity of substitution from the returns of differentiation in a CES production function along the lines of Dixit and Stiglitz (1977)[2]. This is done by pre-multiplying the usual CES function with the term $N_i^{\nu-\frac{1-\alpha}{\alpha}}$, where the additional parameter ν now captures the returns to differentiation, i.e. the degree of proportionality determining at what rate output grows, ceteris paribus, if the set of variants of available intermediate inputs grows. This rate is now independent of the elasticity of substitution $\sigma = \frac{1}{1-\alpha}$ between the differentiated input factors. In the usual Dixit and Stiglitz (1977) specification the returns to differentiation are fixed at $\frac{1}{\sigma-1}$.

In the above production function (3.4) x_j denotes the quantity of the jth differentiated input factor used in the production of sector output Y_i and λ_j denotes its quality level. Therefore there are two channels by which economic growth can take place in the two sectors: First, through growth in horizontal differentiation, i.e. an expansion of the available set of input factors $[0, N_i]$ in sector i, and second, through quality enhancements, i.e. increases in the quality parameter λ_j of the jth differentiated input factor in one of the two sectors. It is assumed that the aggregation defined by the production function (3.4) can be conducted by a continuum of firms and that there is perfect competition between these firms so that the sector good is priced at marginal costs.

3.2.2 Differentiated Input Factors

The differentiated input factors are produced from raw labor according to the following production technology

$$x_j = \begin{cases} l_j & \text{if} \quad i = L, \\ h_j & \text{if} \quad i = H, \end{cases} \tag{3.5}$$

where l_j and h_j denote the employed low and high skilled workers in the production of the jth differentiated input factor for sector i. Each of these differentiated input

[2] See also Benassy (1998) for an application of this kind of production function in a one sector growth model.

factors is produced by a single monopolist who has to incur a fixed investment $F_{i,j}$ at the beginning of each time period t in order to be able to produce. As in Young (1998) these fixed investments depend on the quality λ_j the individual producer wants to offer, and are modelled as[3]

$$F_{i,j} = \begin{cases} f e^{\mu \lambda_{i,j}/\bar{\lambda}_{i,t-1}} & \text{if} \quad \lambda_{i,j} \geq \bar{\lambda}_{t-1}, \\ f e^{\mu} & \text{otherwise}, \end{cases} \tag{3.6}$$

where $\bar{\lambda}_{i,t-1}$ denotes the average quality of all variants N_i of differentiated input factors in sector i in period $t-1$. Thus the fixed costs of producing an input factor are increasing in the desired product quality and decreasing in the stock of knowledge regarding product quality represented by the average quality in period $t-1$. These fixed costs apply both to goods from the range of previously produced input factors, $N_{i,t-1}$, and goods which are not already produced. Therefore entrants and incumbents in the market for differentiated input factors face the same situation in each period of time. Once the fixed costs have been paid, the individual supplier can produce with constant marginal costs as is clear from the production technology (3.5). Due to the dependence of the fixed costs $F_{i,j}$ only on the average of a continuum of past quality levels, the decisions of producers have no individual influence on the future quality environment.

3.2.3 Final Goods Production

Since there are two sectors in the economy, one low skilled and one high skilled, it still needs to be determined, how output of the two sectors is used in the economy. It is assumed that consumption in the model takes place using a combination of the two kinds of goods. The consumption good can be produced with the technology

$$Y = \left(Y_L^{\rho} + Y_H^{\rho}\right)^{\frac{1}{\rho}}, \tag{3.7}$$

[3]In the original Young (1998) model investment and production takes place in two consecutive periods to account for the dynamic structure of the innovation process. However, with this assumption, the two sector extension does not longer possess a closed form solution. Therefore this structure is simplified by assuming that R&D and production take place within one period of time.

which has the well known CES properties, i.e. the elasticity of substitution between high and low skilled products is given by $\varepsilon = \frac{1}{1-\rho}$.

3.2.4 Equilibrium Conditions

Starting with the market for differentiated input factors it is assumed that each monopolist maximizes profits by choosing the appropriate quality level of the product as well as the optimal price. The condition for a profit maximum can be obtained by maximizing the one-period profit function, since there are no individual knowledge spillover effects from one period to another. These profits are given by

$$\pi_j = (p_j - w_i)x_j^D - p_{F,i}F_{i,j} \tag{3.8}$$

where p_j denotes the price of the jth differentiated input factor, w_i are the marginal costs of producing in sector i, i.e. the wage rate for high or low skilled workers, and x_j^D is the demand for the jth differentiated input factor which can be derived from the sector aggregation function (3.4). $p_{F,i}$ is the price of the investment good in sector i which has to be purchased in order to establish production in the quantity $F_{i,j}$ as defined above.

The demand function can be computed by assuming that in the sector aggregation step production costs are minimized:

$$x_j^D = \left(\frac{p_j}{p_i}\right)^{-\frac{1}{1-\alpha}} N_i^{\frac{\nu}{1-\alpha}-\frac{1}{\alpha}} \lambda_j^{\frac{\alpha}{1-\alpha}} \left[\int_0^{N_i} (\lambda_{i,k}x_k)^\alpha \, dk\right]^{\frac{1}{\alpha}}, \tag{3.9}$$

where p_i is the price of the aggregated product in sector i. The first-order conditions for this maximization problem are

$$x_j^D + (p_j - w_i)\frac{\partial x_j^D}{\partial p_j} = 0, \tag{3.10}$$

$$(p_j - w_i)\frac{\partial x_j^D}{\partial \lambda_{i,j}} - p_{F,i}\frac{\partial F_{i,j}}{\partial \lambda_{i,j}} = 0. \tag{3.11}$$

Using the above result for the demand function one can solve these first-order conditions for the optimal price and quality level for the jth intermediate input factor supplier:

$$p_j = \frac{\sigma}{\sigma-1} w_i, \tag{3.12}$$

$$\frac{\lambda_{i,j}}{\bar{\lambda}_{i,t-1}} = \frac{\sigma-1}{\mu}. \tag{3.13}$$

Result (3.12) is the usual mark-up pricing rule for monopolistic competition and (3.13) is the rule for the optimal quality level which can also be found in Young (1998). For the results to be meaningful it is assumed that the elasticity of substitution σ in the sector production function is larger than one[4] and additionally that $\sigma - 1 > \mu$ so that quality improvements take place. The aggregate level of employed high and low skilled labor is given by

$$L = \int_0^{N_L} l_j dj, \tag{3.14}$$

$$H = \int_0^{N_H} h_j dj. \tag{3.15}$$

Additionally it is assumed that the labor market is perfectly competitive so that wages for the high and low skilled always adjust to clear the market and no unemployment occurs. This implies labor demand to always equal its supply which is given exogenously to the model. Also since all differentiated input factors enter the sector production function symmetrically and all monopolists in one sector are setting the same price, the distribution of labor between the input suppliers is symmetric, i.e. $l_j = \frac{L}{N_L}$ and $h_j = \frac{H}{N_H}$. It follows then directly, that the sector production functions can be written in reduced form as

$$Y_L = N_L^{\nu} \bar{\lambda}_H L, \tag{3.16}$$

$$Y_H = N_H^{\nu} \bar{\lambda}_L H, \tag{3.17}$$

where $\bar{\lambda}_i$ is the average quality level in sector i which is given, using the optimality condition (3.13), by $\frac{\sigma-1}{\mu}\bar{\lambda}_{i,t-1}$ since all suppliers in one sector produce with the same quality. Due to this, fixed costs are also identical for all producers of differentiated

[4] This means that no variant of the intermediate input factor is essential for production.

input factors in one sector, $F_{i,j} = F_i = fe^{\sigma-1}$. But since there is free entry in the market for differentiated input factors, the sets of active intermediate input factor producers, N_H and N_L, are also endogenous. Free entry implies zero net-profits for each intermediate input supplier, given by (3.8). Therefore, the number of potential monopolistic competitors in the markets for differentiated input factors is given by

$$N_L = \frac{1}{\sigma-1}\frac{1}{fe^{\sigma-1}}\frac{w_L}{p_{F,L}}L, \tag{3.18}$$

$$N_H = \frac{1}{\sigma-1}\frac{1}{fe^{\sigma-1}}\frac{w_H}{p_{F,H}}H. \tag{3.19}$$

The set of variants of the differentiated input factors is thus increasing in the number of employees who can work in the respective sector, and in their wage rate. The larger the market for a variant of an input factor, i.e. the larger the number of workers in that market, the higher are, ceteris paribus, the absolute profits for one input factor supplier and more individual suppliers can survive in the market. The same is true for the wage rate. The higher the wage rate is, the larger is the market for different variants of input factors. This happens to be the case because a larger set of variants reduces unit costs in sector production stronger. Thus high wage costs, implying high unit costs, strengthens the effect of differentiation. Naturally, higher investment costs of setting up the production process decrease the number of potential producers. The wage rate for the high and low skilled can be calculated from the sector specific production function using standard maximization arguments:

$$w_L = \frac{\sigma-1}{\sigma}N_L^{\nu}p_L\bar{\lambda}_L, \tag{3.20}$$

$$w_H = \frac{\sigma-1}{\sigma}N_H^{\nu}p_H\bar{\lambda}_H. \tag{3.21}$$

These wage equations are quite standard in showing the usual dependence of the wage rate on the technological level of the respective sector, given by N_i^{ν}, $i = H, L$, the price level as well as the quality level of the respective input factors. What is new, is the interrelationship between the wage rate and the level of technology as given by equations (3.20) and (3.21) and the dependence of the level of technology on the wage rate in equations (3.18) and (3.19). The next section will deal with the solution of this equation system as well as with different assumptions about the fixed investment costs which have to be incurred by the differentiated input factor producers.

3.2.5 The Price for Investment Goods and Wage Inequality

There are different possibilities for the use of the different goods in the R&D process of the economy. A priori it might be possible that the results of the above model might depend on them although the general economic reasoning behind the model stays the same. The next subsections will deal with the different combinations that can occur.

Sector Products as Investment Goods

In this first subsection it is assumed that for covering the fixed setup costs, the producer has to purchase a certain amount of the respective sector product given by equation (3.6). Thus the price for the investment good $p_{F,i}$ equals p_i. Now turning to the relevant relative figures of the model, since we are interested in the relative wage for the high and low skilled. From equations (3.18) to (3.21) it can be deducted that

$$\frac{N_H}{N_L} = \left(\frac{H}{L}\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\frac{1}{1-\nu}} \tag{3.22}$$

and that the relative wage is given by

$$\frac{w_H}{w_L} = \left(\frac{N_H}{N_L}\right)^{\nu}\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\frac{p_H}{p_L}. \tag{3.23}$$

Therefore the effect of the skill composition of the workforce, i.e. the ratio $\frac{H}{L}$, is twofold and can be decomposed as in Acemoglu (1998). First, there is a market size effect: An increase in the relative supply of high skilled raises the relative profitability for the suppliers of differentiated input factors in the high skilled sector. This is also the channel through which the "weak" scale effect operates. Therefore the relative number of variants of differentiated input factors for the high skilled increases and the relative level of technology rises in favor of the high skilled. Note however that for this market size effect to be positive it is necessary that the parameter ν representing the returns to differentiation has to be less than one. If the parameter ν were greater than one, there wouldn't be a stable equilibrium in the market for intermediate input factors. The reason for this is that new variants decrease the unit costs for sector production and therefore also their prices. Thus, if ν would be larger than one, new entrants in the market for intermediate input factors would lower the price for the investment good, inducing even more entrants. This behavior would led the set of variants to grow to infinity immediately. Therefore the case $\nu > 1$ is not economically meaningful.

Counteracting the market size effect there is a negative price effect. Given an increase in the relative supply of high skilled and the induced relative rise in the state of the sector technology just mentioned, the higher relative output of the high skilled sector depresses the relative price for high skilled products which will have an unfavorable effect on the relative wage for the high skilled. The "weak" scale effect amplifies this in a more sector environment. Because sector production is not only expanded by using more workers, also their productivity increases, yielding more output per worker. This last effect is due to the dependence of labor productivity on the extend of the market. The question which remains to be answered is, which of the two effects is stronger under what conditions.

Using the assumption that high and low skilled products can be substituted with elasticity ε according to the aggregation technology (3.7), the reduced form for the relative wage can be computed from equations (3.22) and (3.23):

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\frac{\varepsilon-1}{\varepsilon}\frac{1}{1-\nu}} \left(\frac{H}{L}\right)^{\frac{\nu}{1-\nu}-\frac{1}{\varepsilon}\frac{1}{1-\nu}} . \tag{3.24}$$

From the exponent of the term reflecting the relative supply of skills, the market size and the price effect are obvious. The former is given by $\frac{\nu}{1-\nu}$, the latter by $-\frac{1}{\varepsilon(1-\nu)}$. Thus for the market size effect to be strong enough to outweigh the negative price effect, the elasticity of substitution has to be large, i.e. $\varepsilon > \frac{1}{\nu}$. From the preceding discussion it is clear that this automatically implies that the aggregate elasticity of substitution must be greater than one. Note that if this is the case, then the quality enhancing technological change, i.e. the growth in $\bar{\lambda}_i$, is also favoring labor.

This means that, similar to the theory of directed technical change (Acemoglu 1998), there exists in the long run, after technological adjustment, the possibility that the relative (nominal) marginal product of one type of labor can increase in response to an increase in its relative supply.

If one abstracts from the generalized specification of the production function and returns to the original Dixit and Stiglitz (1977) specification, where the returns to differentiation are implicitly given by the elasticity of substitution, i.e. $\nu = \frac{1}{\sigma-1}$, the above condition is given by $\varepsilon > \sigma - 1$.

Taking all results together it is possible to rewrite the aggregate economy wide production function in reduced form taking the endogeneity of technology into account:

$$Y = f^{-\frac{\nu}{1-\nu}} e^{-\frac{\nu}{1-\nu}(\sigma-1)} \sigma^{-\frac{\nu}{1-\nu}} \left[\left(\bar{\lambda}_L L\right)^{\frac{\rho}{1-\nu}} + \left(\bar{\lambda}_H H\right)^{\frac{\rho}{1-\nu}} \right]^{\frac{1}{\rho}}. \tag{3.25}$$

The relationship (3.25) is a CES production function with elasticity of substitution given by $\frac{1}{1-\frac{\rho}{1-\nu}}$. Also, this production function exhibits increasing returns to scale, with respect to labor, if the returns to differentiation are in the interval $(0, 1)$. These increasing returns to scale arise from the returns to differentiation. If they are zero, there is no effect from the size of the workforce onto the individual productivity of each worker via the extension of the number of variants of input factors. It also can be seen that the growth rate of the economy γ_Y is given by

$$\gamma_Y = \left(\frac{\sigma - 1}{\mu} (1 + \gamma_N) \right)^{\frac{1}{1-\nu}} - 1, \tag{3.26}$$

where γ_N is the growth rate of the population applying both to the high and low skilled population. The term $\frac{\sigma-1}{\mu}$ denotes the growth factor of the quality index $\bar{\lambda}_i$ which is, due to the assumption of an equal elasticity of substitution across sectors, the same for both sectors. Note that as in the one sector model of Young (1998) there is no scale effect in the growth rate of the economy.

Final Goods as Investment Goods

The analysis of the preceding section can also be conducted after changing the assumption regarding the choice of the investment good. In this section it is assumed that the final good, aggregated according to equation (3.7), is necessary for the fixed investment. This means that the relevant price $p_{F,i}$ now equals the price of the final good which is normalized to one in the model. Economically this is equivalent to assuming, that a combination of high and low skill goods is necessary to conduct R&D. This combination equals the combination used to form consumption goods. This alters the situation compared with the assumptions in the preceding subsection. In that subsection, new entrants in the market for intermediate input factors lowered only their investment cost. Now there is a spill-over effect, because higher productivity through new variants in one sector affect the investment decision in both sectors.

Now solving for the equilibrium state of relative technology using equations (3.18) to (3.21), this gives

$$\frac{N_H}{N_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\frac{H}{L}\right)^{\frac{\varepsilon-1}{\varepsilon-\nu(\varepsilon-1)}}, \tag{3.27}$$

which obviously alters the market size effect that affects the relative wage of the high skilled workers. The existence of the negative price effects follows the same arguments as in the preceding section because equation (3.23) applies here as well.

Building the reduced form for the relative wage of the high skilled using the same calculus as in the section above it turns out that the relative wage is given by

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\frac{\varepsilon-1}{\varepsilon-(\varepsilon-1)\nu}} \left(\frac{H}{L}\right)^{\frac{(\varepsilon-1)\nu}{\varepsilon-(\varepsilon-1)\nu}-\frac{1}{\varepsilon-(\varepsilon-1)\nu}}. \tag{3.28}$$

Having established this relationship it is again interesting to find out under what conditions the market size effect outweighs the price effect with the above change in the assumptions. From (3.28) it follows that this is the case whenever the elasticity of substitution ε is greater than $1 + \frac{1}{\nu}$[5]. If this applies then the relative (nominal) marginal product of labor increases with its relative supply. Comparing this result with that from the last subsection, it becomes clear that this condition is more demanding, since the elasticity ε has to be larger, ceteris paribus. This is due to the above mentioned spill-over effect which lets both sectors profit from innovations in one sector.

If one takes the original Dixit and Stiglitz (1977) specification as a benchmark, the condition for a rising relative wage for the high skilled in response to an increase in the relative supply of the high skilled would be $\varepsilon > \sigma$. The elasticity of substitution in the aggregation of high and low skilled products must be greater than the elasticity of substitution in the sector specific production technology.

Using the above results one can again compute the reduced form economy-wide production function, taking technological adjustments into account.

$$Y = e^{-\frac{\nu}{1-\nu}(\sigma-1)} f^{-\frac{\nu}{1-\nu}} \sigma^{-\frac{\nu}{1-\nu}} \left[\left(\bar{\lambda}_L L\right)^{\frac{\varepsilon-1}{\varepsilon-\nu(\varepsilon-1)}} + \left(\bar{\lambda}_H H\right)^{\frac{\varepsilon-1}{\varepsilon-\nu(\varepsilon-1)}} \right]^{\frac{\varepsilon-\nu(\varepsilon-1)}{(\varepsilon-1)(1-\nu)}} . \quad (3.29)$$

This production function is again of the CES type and has an elasticity of substitution equal to $\frac{1}{1-\frac{\varepsilon-1}{\varepsilon-\nu(\varepsilon-1)}}$ as well as increasing returns to scale. These increasing returns to scale again stem from the returns to differentiation via the same channel as described in the preceding paragraph.

The growth rate of the economy is given by

$$\gamma_Y = \left(\frac{\sigma - 1}{\mu} (1 + \gamma_N) \right)^{\frac{1}{1-\nu}} - 1 \quad (3.30)$$

which is the same result as in the previous section where sector output was chosen to cover the fixed costs of production.

Labor as a Fixed Investment As a last possibility for the fixed investment costs in the input factor production the case of pure labor is explored. This case is a direct application of the Young's (1998) assumption in a two sector framework that a fixed

[5]As in the preceding subsection it applies that $\nu < 1$ must hold in order to obtain a stable equilibrium.

amount of labor is needed to set up the production process. Here it is assumed that a fixed amount of low skilled labor is needed in the low skill sector and analogously a fixed amount of high skilled labor is needed in the high skilled sector. The quantities are again given by specification (3.6), but the prices for the investment good are now given by w_L and w_H. The difference with respect to the two cases examined above is now, that intermediate input producers do not benefit from new variants due to decreasing investment costs. In the contrary, their investment costs increase because wage costs rise with the set of variants due to the higher labor productivity in the sector production.

Using this assumption the relative state of technology for the high and the low skilled sector is now given by

$$\frac{N_H}{N_L} = \frac{H_P}{L_P}, \tag{3.31}$$

where H_P and L_P are the quantities of high and low skilled labor used in the production process of input factors. Using the result that the quantities of labor used to cover the fixed costs of production are given for the jth firm in sector i by $fe^{\sigma-1}$, and that the number of differentiated input factors is given by equations (3.18) and (3.19) it turns out that the distribution of labor is given by

$$H_P = \frac{\sigma - 1}{\sigma} H, \tag{3.32}$$

$$H_F = \frac{1}{\sigma} H, \tag{3.33}$$

$$L_P = \frac{\sigma - 1}{\sigma} L, \tag{3.34}$$

and

$$L_F = \frac{1}{\sigma} L. \tag{3.35}$$

H_F and L_F give the number of high and low skilled workers needed to cover the fixed costs of production of all suppliers of differentiated input factors. The reduced form for the relative wage of the high skilled is given by

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\frac{\varepsilon-1}{\varepsilon}} \left(\frac{H_P}{L_P}\right)^{\frac{\varepsilon\nu-(1+\nu)}{\varepsilon}}, \tag{3.36}$$

where again the exponent of the term reflecting the relative skill distribution of the workforce is written in a way that separates the market size and the price effect, i.e. ν and $-\frac{1+\nu}{\varepsilon}$. The market size effect dominates the negative price effect if the elasticity of substitution ε is again greater than $1 + \frac{1}{\nu}$, the same result as in the preceding section where the final good was chosen to cover the fixed costs of production.

Translated in the original Dixit and Stiglitz (1977) specification the respective condition would again be $\varepsilon > \sigma$.

The economy-wide aggregated production function implied by the above assumption about the production process can be computed to be

$$Y = (\sigma-1)^{-\nu} f^{-\nu} e^{-(\sigma-1)\nu} \left(\bar{\lambda}_L^{\frac{\varepsilon-1}{\varepsilon}} L_P^{\frac{(\varepsilon-1)(1+\nu)}{\varepsilon}} + \bar{\lambda}_H^{\frac{\varepsilon-1}{\varepsilon}} H_P^{\frac{(\varepsilon-1)(1+\nu)}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3.37}$$

where H and L give the total number of high and low skilled workers in the economy. This is again a CES production function with a scale effect if the returns to differentiation are positive. The aggregate elasticity of substitution, taking technological adjustment into account, is $\frac{1}{1-\rho(1+\nu)}$ and the increasing returns to scale again stem from the existence of the returns to differentiation. Here the growth rate of the economy differs from the results of the above sections and is now given by

$$\gamma_Y = \frac{\sigma-1}{\mu}(1+\gamma_N)^{1+\nu} - 1. \tag{3.38}$$

3.3 A Model with the Romer Production Technology

The preceding analysis focused on the production technology as used by Ethier (1982) based on Dixit and Stiglitz (1977). Another popular representation using the assumption that technological progress takes the form of an increasing number of variants of input factors as well as an increase in their quality level was introduced by Romer (1987). In this specification the level of technology is given by the degree of diversification of intermediate input factors that are used in combination with labor to yield output.

3.3.1 The High and the Low Skilled Sector

It is assumed that in the high and low skilled sector high and low skilled workers are employed along with a differentiated set of intermediate input factors to yield output:

$$Y_H = H_P^{\alpha} \int_0^{N_H} (\lambda_{H,j} x_j)^{1-\alpha}\, dj, \tag{3.39}$$

and

$$Y_L = L_P^{\alpha} \int_0^{N_L} (\lambda_{L,j} x_j)^{1-\alpha}\, dj. \tag{3.40}$$

The notation is similar to the one in the preceding analysis using the Ethier (1982)/Dixit and Stiglitz (1977) specification. The difference is that the parameter α now denotes the output elasticity of labor and N_H as well as N_L now give the number of differentiated input factors which can be used in combination with labor. Their quantity and quality is denoted by x_i and $\lambda_{i,j}$ respectively and they need not necessarily to be produced from pure labor. H_P and L_P denote the number of high and low skilled workers employed in the production process. The two channels for technological progress are again a rising number of differentiated machines and a rising level of their quality.
It is also assumed that the production of the high skilled and low skilled sector goods can be conducted by many firms according to the production functions (3.39) and (3.40) so that prices equal marginal costs due to perfect competition.

3.3.2 Differentiated Intermediate Input Factors

As explained above the differentiated input factors used in the production technology (3.39) and (3.40) need not be produced by labor. For now it is assumed that they are produced by a not closer specified sector-specific production factor with price $p_{M,i}$. As in the previous analysis the assumption is made that each variant of these differentiated input factors is produced by a monopolist who has to incur a fixed cost at the beginning of the production process and faces constant marginal cost thereafter. The fixed costs are assumed to be influenced by the quality level to be chosen by the monopolist as in the preceding analysis in (3.6) which builds on Young (1998).

3.3.3 Final Good Production

As in the section using the Ethier (1982)/Dixit and Stiglitz (1977) specification it is assumed that there is a final aggregation stage in production in which sector output is used to form a good which then can be used for consumption according to

$$Y = \left(Y_L^{\rho} + Y_H^{\rho}\right)^{\frac{1}{\rho}}. \tag{3.41}$$

3.3.4 Equilibrium conditions

The suppliers of the differentiated machines used in the sector production can set prices to maximize their profits which are given similar to equation (3.8) by

$$\pi_j = (p_j - p_{M,i})x_j^D - p_{F,i}F_i, \tag{3.42}$$

where again p_j is the price for one unit of the jth variant of the differentiated input factor, $p_{M,i}$ is the price of producing one unit of it in sector i, $i = H, L$. $p_{F,i}$ is the price for the investment good covering the fixed costs of producing one variant of the differentiated machines in sector i and F_i is the necessary quantity. The demand for each variant of the differentiated machines can be computed from the usual marginal product condition:

$$x_i^D = (1-\alpha)^{\frac{1}{\alpha}} p_j^{-\frac{1}{\alpha}} p_i^{\frac{1}{\alpha}} \lambda_{H,j}^{\frac{1-\alpha}{\alpha}} H_P \tag{3.43}$$

for the high skilled sector and

$$x_j^D = (1-\alpha)^{\frac{1}{\alpha}} p_j^{-\frac{1}{\alpha}} p_i^{\frac{1}{\alpha}} \lambda_{L,j}^{\frac{1-\alpha}{\alpha}} L_P \tag{3.44}$$

for the low skilled sector.

The first order conditions for a profit maximum of the monopolist determining the optimal choice for the price p_j and the quality level λ_j are analogous to equations (3.10) and (3.11). The solutions for prices and quality are given by

$$p_j(t) = \frac{1}{1-\alpha} p_{M,i} \tag{3.45}$$

and

$$\frac{\lambda_{i,j}}{\bar{\lambda}_{i,t-1}} = \frac{1-\alpha}{\alpha}\frac{1}{\mu}, \tag{3.46}$$

and are similar to the solutions in the preceding section. Especially the development of individual quality levels guarantees that quality levels are homogenous.

The sector production functions can be written in reduced form as

$$Y_L = (1-\alpha)^{2\frac{1-\alpha}{\alpha}} \left(\frac{p_L}{p_{M,L}}\right)^{\frac{1-\alpha}{\alpha}} N_L \bar{\lambda}_L^{\frac{1-\alpha}{\alpha}} L_P \tag{3.47}$$

and

$$Y_H = (1-\alpha)^{2\frac{1-\alpha}{\alpha}} \left(\frac{p_H}{p_{M,H}}\right)^{\frac{1-\alpha}{\alpha}} N_H \bar{\lambda}_H^{\frac{1-\alpha}{\alpha}} H_P. \tag{3.48}$$

Note that here the returns to differentiation are equal to one due to the specific production function.

Since in this specification there is also free entry into the market for differentiated input factors, the set of producers, N_H and N_L , is here endogenous too. Free entry implies zero net-profits, given by (3.42) for each supplier and the number of potential producers is given by

$$N_L = (1-\alpha)\frac{1}{fe^{\frac{1-\alpha}{\alpha}}}\frac{w_L}{p_{F,L}}L_P \tag{3.49}$$

and

$$N_H = (1-\alpha)\frac{1}{fe^{\frac{1-\alpha}{\alpha}}}\frac{w_H}{p_{F,H}}H_P. \tag{3.50}$$

These results are quite similar to the results using the Ethier (1982)/Dixit and Stiglitz (1977) specification above and the economic interpretation is analogous.

The wage rate for the high and low skilled can be calculated from the sector specific production function using standard maximization arguments:

$$w_L = \alpha(1-\alpha)^{2\frac{1-\alpha}{\alpha}} p_L^{\frac{1}{\alpha}} p_{M,L}^{-\frac{1-\alpha}{\alpha}} N_L \bar{\lambda}_L^{\frac{1-\alpha}{\alpha}}, \tag{3.51}$$

and

$$w_H = \alpha(1-\alpha)^{2\frac{1-\alpha}{\alpha}} p_H^{\frac{1}{\alpha}} p_{M,H}^{-\frac{1-\alpha}{\alpha}} N_H \bar{\lambda}_H^{\frac{1-\alpha}{\alpha}}. \tag{3.52}$$

These wage equations are again quite standard in that they show as usual the (nominal) marginal product of the different types of labor which crucially depends on the level of technology given by the degree of differentiation, i.e. N_H and N_L, and the quality levels.

Not surprisingly there is again an interrelationship between the degree of differentiation and the wage rate as given by equation (3.49) and (3.50), and between the wage rate and the level of technology as given by equations (3.51) and (3.52).

In the following sections it will be analyzed how this simultaneous system of equations is solved for the relative wage for the high skilled as a function of exogenously given figures. With the Romer specification there are two prices left to be determined, the price for producing the differentiated machines $p_{M,i}$, as well as the price for the investment good covering the fixed costs of producing machines $p_{F,i}$.

Therefore there are many possible combinations to be analyzed but not all price combinations are economically meaningful since only for some of them a balanced growth path exists. The following sections deal only with cases where such a balanced growth path exists.

3.3.5 Fixed Costs, Production Costs and Wage Inequality

As mentioned above a balanced growth path cannot always be found for the model. One assumption guaranteeing a balanced growth path is that a certain amount of sector specific labor is needed to cover the fixed costs in the production of the differentiated machines. The other cases are ruled out because then the set of variants is not bounded because profits for one individual intermediate input factor producer do not decline if new entrants enter the market.

This assumption now implies that one part of the total number of high and low skilled workers, i.e. H_P and L_P is employed in the sector production stage. The remaining part of the workforce is needed to set up the production process of the differentiated machines, H_F and L_F, and, if necessary, to produce the differentiated machines.

In addition to this one more production factor is needed to cover the variable production costs of the production of differentiated input factors. In the following subsections the cases of sector specific output, final output and sector specific labor as the variable production factor will be examined.

Sector Specific Output as a Production Factor

In the first case considered it is assumed that sector specific output is introduced as the additional production factor needed to produce the differentiated input factors. This means that the price $p_{M,i}$ now equals p_i. From equations (3.49) and (3.50) it can be seen that the relative state of technology with the above assumptions is given by

$$\frac{N_H}{N_L} = \frac{H_P}{L_P}. \tag{3.53}$$

This in turn implies a relative wage for the high skilled in reduced form of

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\frac{\varepsilon-1}{\varepsilon}\frac{1-\alpha}{\alpha}} \left(\frac{H_P}{L_P}\right)^{1-\frac{2}{\varepsilon}}, \tag{3.54}$$

using the substitution effect which arises from the aggregation technology (3.7) and the wage equations (3.51) and (3.52) as well as (3.53).

From the specification of the fixed costs (3.6) and the number of variants of differentiated input factors it can be deduced that

$$H_F = (1-\alpha)H_P, \tag{3.55}$$

and

$$H_p = \frac{1}{2-\alpha}H. \tag{3.56}$$

From the above expression for the relative wage it can be seen that the relative wage for the high skilled is increasing in response to an increase in the relative supply of the high skilled if the elasticity of substitution in the final aggregation stage ε is larger than two. If this is the case, prices react by such a small amount to supply changes that the market size is the stronger economic force.

The economy wide aggregated production function in reduced form can be written as

$$Y = (1-\alpha)^{2\frac{1-\alpha}{\alpha}+1} e^{-\frac{1-\alpha}{\alpha}} f^{-1} \left[\left(\bar{\lambda}_L^{\frac{1-\alpha}{\alpha}} L_P^2 \right)^{\frac{\varepsilon-1}{\varepsilon}} + \left(\bar{\lambda}_H^{\frac{1-\alpha}{\alpha}} H_P^2 \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3.57}$$

which is again of the CES type and has increasing returns to scale for reasons mentioned in the preceding sections.

From the above production function (3.57) it can also be concluded that the growth rate of the economy is given by

$$\gamma_Y = \left(\frac{1}{\mu}\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}} (1+\gamma_N)^2 - 1, \tag{3.58}$$

which reflects the two sources of economic growth, the quality enhancement and the scale effect.

Final Output as Production Factor

In this section it is assumed that for the production of differentiated input factors, only the final good is necessary which is equivalent to setting the price $p_{M,i}$ equal to the price of the final good which is normalized to one. The equilibrium relative state of technology is again given by

$$\frac{N_H}{N_L} = \frac{H_P}{L_P}. \tag{3.59}$$

Using the wage equations (3.51) and (3.52) the relative wage and the elasticity of substitution implied by the aggregation technology (3.7) the reduced form for the relative wage for the high skilled is given by

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{(1-\alpha)\frac{\varepsilon-1}{\varepsilon\alpha+1-\alpha}} \left(\frac{H_P}{L_P}\right)^{\frac{\alpha(\varepsilon-1)-1}{\varepsilon\alpha+1-\alpha}}, \tag{3.60}$$

where again the market size and the price effect can be read off in the exponent of the relative supply of the high skilled. The market size effect dominates the price effect if the elasticity of substitution ε is greater than $1+\frac{1}{\alpha}$. If this is the case the relative marginal product of high skilled labor increases with its relative supply.

Using the above results and taking technological adjustments into account, the reduced form economy wide production function is given by

$$Y = (1-\alpha)^{2\frac{1-\alpha}{\alpha}+1} f^{-1} e^{-\frac{1-\alpha}{\alpha}} \left[\left(\bar{\lambda}_L^{\frac{1-\alpha}{\alpha}} L_P^2\right)^{\frac{\alpha(\varepsilon-1)}{\varepsilon\alpha+1-\alpha}} + \left(\bar{\lambda}_H^{\frac{1-\alpha}{\alpha}} H_P^2\right)^{\frac{\alpha(\varepsilon-1)}{\varepsilon\alpha+1-\alpha}}\right]^{\frac{\varepsilon\alpha+1-\alpha}{\alpha(\varepsilon-1)}}. \tag{3.61}$$

This production function is again of the CES type and has an elasticity of substitution equal to $\frac{1}{1-\frac{\alpha(\varepsilon-1)}{\varepsilon\alpha+1-\alpha}}$ and increasing returns to scale. These increasing returns to scale originate from the returns to differentiation which are here equal to one via the same channel as described in the preceding sections.

From the aggregate production function (3.61) it can also be seen that the growth rate of the economy is now

$$\gamma_Y = \left(\frac{1}{\mu}\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}} (1+\gamma_N)^2 - 1, \tag{3.62}$$

which reflects the two sources of technological progress, the quality improvement and the scale effect through an increasing number of variants if the population is growing.

This case can easily be modified to introduce capital into the model. The only major change that has to be made, is in the price for the production factor used to produce the differentiated input factors. Suppose that in each period the machines are produced from capital goods. Capital goods can be produced from final output alone using a simple linear production function with unit productivity. If it is assumed for simplicity that capital does not depreciate the price $p_{M,i}$ is now equal to the interest rate r.

Assume that a representative consumer has the intertemporal utility function

$$U = \sum_{t=0}^{\infty} \frac{1}{(1+\beta)^t} \ln c_t, \tag{3.63}$$

where $\beta > 0$ is the rate of time preference and c_t is consumption. If the representative consumer maximizes (3.63) subject to the intertemporal budget constraint

$$a_{t+1} = (1+r_t)a_t + w_t - c_t, \tag{3.64}$$

where w_t is income from labor supply reflecting the skill composition of the population and a_t denotes individual assets, the usual optimality condition

$$\frac{c_{t+1}}{c_t} = \frac{1+r_{t+1}}{1+\beta} \tag{3.65}$$

emerges, where r_t is the interest rate on savings.

The demand for each variant of the differentiated input factors is given by equations

(3.43) and (3.44) which specialize here to

$$x_j^D = \frac{1-\alpha}{\alpha} \frac{w_H}{r} f e^{\frac{1-\alpha}{\alpha}} \tag{3.66}$$

for the high skilled sector and

$$x_j^D = \frac{1-\alpha}{\alpha} \frac{w_L}{r} f e^{\frac{1-\alpha}{\alpha}} \tag{3.67}$$

for the low skilled sector.

Since the differentiated input factors are produced from the capital stock of the economy which is divided between the high and the low skilled sector, these two capital stocks K_H and K_L are given by

$$K_H = \frac{(1-\alpha)^2}{\alpha} H_P \frac{w_H}{r} \tag{3.68}$$

and

$$K_L = \frac{(1-\alpha)^2}{\alpha} L_P \frac{w_L}{r} \tag{3.69}$$

using the number of variants for differentiated machines in equations (3.49) and (3.50).

Using the result for the relative wage given in (3.60), it follows that

$$\frac{K_H}{K_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{(1-\alpha)\frac{\varepsilon-1}{\varepsilon\alpha+1-\alpha}} \left(\frac{H_P}{L_P}\right)^{\frac{\alpha(\varepsilon-1)-1}{\varepsilon\alpha+1-\alpha}+1} \equiv \omega_H.$$

With the constraint for the aggregate capital stock $K = K_H + K_L$, this gives

$$\begin{aligned} K_H &= \frac{\omega_H}{1+\omega_H} K, \\ K_L &= \frac{1}{1+\omega_H} K. \end{aligned}$$

It must be true that $x_j = \frac{K_i}{N_i}$ and therefore that $\int_0^{N_i} (\bar{\lambda}_i x_j)^{1-\alpha} dj = N_i^{\alpha} \bar{\lambda}_i^{1-\alpha} K_i^{1-\alpha}$. With N_i given by equations (3.49) and (3.50) and $p_{F,i} = w_i$ as well as the above rule for the two capital stocks K_H and K_L, aggregate production of the final good can be written as

$$Y = (1-\alpha)^{\alpha} f^{-\alpha} e^{-(1-\alpha)} \left[H_P^{2\alpha\rho} \bar{\lambda}_H^{(1-\alpha)\rho} \left(\frac{\omega_H}{1+\omega_H} \right)^{(1-\alpha)\rho} + \right. \\ \left. + L_P^{2\alpha\rho} \bar{\lambda}_L^{(1-\alpha)\rho} \left(\frac{1}{1+\omega_H} \right)^{(1-\alpha)\rho} \right]^{\frac{1}{\rho}} K^{1-\alpha}.$$

This is a production function with increasing returns to scale with respect to labor and capital, but with a decreasing marginal product of capital. Therefore the model extended with capital accumulation possesses the usual saddle path stability property.

The growth rate of the economy, given a constant interest rate, is still given by (3.62) so that from the optimality condition of the consumers (3.65) the interest rate on the balanced growth path can be computed as

$$r = (1+\gamma_Y)(1+\beta) - 1. \tag{3.70}$$

The result for the relative wage remains intact but the reduced form of the aggregated production function alters because capital now does not depreciate which has a level effect on production

$$Y = f^{-1} e^{-\frac{1-\alpha}{\alpha}} (1-\alpha)^{2\frac{1-\alpha}{\alpha}+1} r^{-\frac{1-\alpha}{\alpha}} \left[\left(\bar{\lambda}_L^{\frac{1-\alpha}{\alpha}} L_P^2 \right)^{\frac{\rho\alpha}{\alpha+(1-\rho)(1-\alpha)}} + \right. \\ \left. + \left(\bar{\lambda}_H^{\frac{1-\alpha}{\alpha}} H_P^2 \right)^{\frac{\rho\alpha}{\alpha+(1-\rho)(1-\alpha)}} \right]^{\frac{\alpha+(1-\rho)(1-\alpha)}{\rho\alpha}}. \tag{3.71}$$

Sector Specific Labor as Production Factor

In this last case to be considered, it is assumed that sector specific labor is used as an additional production factor to produce the differentiated input factors which are combined at the sector level with labor to produce sector goods. The production function for this takes the simple form

$$x_j = \begin{cases} l_i & \text{if} \quad i = L, \\ h_i & \text{if} \quad i = H. \end{cases} \tag{3.72}$$

Aggregation over all variants of differentiated machines gives the constraints

$$\int_0^{N_L} l_j dj = L - L_P - L_F \equiv L_X \tag{3.73}$$

and

$$\int_0^{N_H} h_j dj = H - H_P - H_F \equiv H_X. \tag{3.74}$$

This naturally implies that the price for the additional production factor $p_{M,i}$ is the sector specific wage rate w_i.

From the demand function for differentiated input factors it can be seen that the quantity of each type produced is

$$x_j = \frac{(1-\alpha)^2}{\alpha} \frac{H_P}{N_H} \tag{3.75}$$

for the high skilled sector and

$$x_j = \frac{(1-\alpha)^2}{\alpha} \frac{L_P}{N_H} \tag{3.76}$$

for the low skilled sector. Together with assumption (3.6) about the fixed setup costs of the production process of differentiated machines, this gives the demand for labor

$$H_F = N_H f e^{\frac{1-\alpha}{\alpha}} \tag{3.77}$$

and

$$L_F = N_L f e^{\frac{1-\alpha}{\alpha}}. \tag{3.78}$$

The total labor force is spread between the different production stages according to

$$\begin{aligned} H_P &= \alpha H, && (3.79)\\ H_F &= (1-\alpha)\alpha H, && (3.80)\\ H_X &= (1-\alpha)^2 H. && (3.81) \end{aligned}$$

The same rules apply for the employment of low skilled labor. Using the marginal product condition (3.51) and (3.52) for the wage rate, the assumption about the prices for the production of machines, the fixed costs as well as the elasticity of substitution in the aggregate production function (3.7) the relative wage for the high skilled turns out to be

$$\frac{w_H}{w_L} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{(1-\alpha)\frac{\varepsilon-1}{\varepsilon}} \left(\frac{H_P}{L_P}\right)^{\alpha\frac{\varepsilon-1}{\varepsilon}-\frac{1}{\varepsilon}}. \tag{3.82}$$

One again sees the forces at work of the market size and the price effect in the influence of the relative supply of skills on the relative wage. The market size effect is more powerful than the price effect if the elasticity of substitution ε in the aggregation technology (3.7) is again greater than $1 + \frac{1}{\alpha}$.

The economy-wide production function for final output in reduced form, taking account of the endogenous state of technology and using the production functions (3.39) and (3.40) together with the assumptions of this subsection, is given by

$$Y = \alpha^{-(1-\alpha)}(1-\alpha)^{2-\alpha}e^{-(1-\alpha)}f^{-\alpha}\left[\left(\bar{\lambda}_L^{1-\alpha}L^{1+\alpha}\right)^{\frac{\varepsilon-1}{\varepsilon}} + \left(\bar{\lambda}_H^{1-\alpha}H^{1+\alpha}\right)^{\frac{\varepsilon-1}{\varepsilon}}\right]^{\frac{\varepsilon}{\varepsilon-1}}. \tag{3.83}$$

This implies a growth rate for the economy as a whole of

$$\gamma_Y = \left(\frac{1}{\mu}\frac{1-\alpha}{\alpha}\right)^{1-\alpha} (1+\gamma_N)^{1+\alpha} - 1. \tag{3.84}$$

3.4 Relative Wage Rigidity and Unemployment

The last section has shown how wage inequality comes into existence in response to shocks in the relative supply of skills in the economy. A quite popular argument is the so called Krugman (1984) hypothesis. According to this hypothesis, unemployment occurs if the adjustment of relative wages, i.e. wage inequality in this case, is suppressed by some economic force. However, this hypothesis has, to the best of the knowledge of the author, not been subjected to a model in which technology, as a driving force of inequality, is endogenous. Therefore this section will show under which conditions the argument of the Krugman hypothesis works with adjusting technologies. For this purpose relative wage rigidity in a small open economy is introduced and it will be elaborated on the assumptions needed for unemployment to emerge among the low skilled workers. The setup of this extension of the model is as follows. First, the analysis is conducted in the formulation of the model of section 3.2.5, i.e., the Ethier (1982) specification where sector specific labor is used in the production of the differentiated input factors and for covering the fixed costs of their production used to improve quality. Second, it is assumed that the world economy consists only of two countries, one large and one small. Both countries engage in trade in intermediate input factors and since the number of these input factors determines the state of technology technology transfer between the two countries takes place. Production of sector goods and final output is then undertaken in each country separately. Labor is immobile between countries and the production technology (the production functions) is identical for both countries. Additionally it is assumed that a technology is available to produce very close substitutes of the intermediate input factors once they have been designed without incurring the fixed costs of research. The comparative disadvantage of this technology is that it possesses higher marginal production costs through a labor productivity equal to

α[6].

One deviation from traditional models of trade like Krugman (1980), which is most related to this section, must be noted. In a traditional trade model free trade in intermediate input factors would immediately lead to factor price equalization, a feature which is not helpful in explaining different wage behavior in different countries. Therefore assumptions have to be made guaranteeing the existence of different wages in the two economies. It is clear that these assumption are strong in the sense that they are critical for the results of the model. They should be seen as requirements of the Krugman hypothesis to work in this model. It is then left to the reader to build his opinion about the relevance of this assumptions in the real economic world. One economic story that yields the necessary assumptions could read as follows.

Assume a timing of events in the two economies: First agents involved in trading intermediate input factors have to close exclusive contracts with producers, entitling them to be the only trader for particular variants of the input factors with the obligation of serving the foreign market. These contracts can only be accepted by the trading agent and the producer if it guarantees the trader the same price as the producers of the sector goods have to pay for intermediate input factors. Second the trading agents make contracts with producers of sector goods in the export market. These producers can only accept the contract if it guarantees them the same price as for intermediate input factors in their home market because they know of the existence of the competitive technology which could serve their demand at that price[7]. Since at that point of time wages and prices in the economies are not yet determined, the trading agent does not know whether he will make profits or losses with this arrangement. Therefore to insure him from making losses, he takes exactly the same position as an exporter from the other country which frees him from any risk[8].

[6] The choice of α is arbitrary. One could introduce an additional productivity parameter smaller than one with the result that instead of monopoly pricing limit pricing would occur.

[7] This argument is true if the price difference between the exporting and the importing country is negative. If it were positive the exporter would not sign in a contract with a lower price because he then would make losses as will be clear from the arguments below.

[8] This set of assumptions is somehow similar to the set up in Engel (1996) where an exporting

After these contracts have been closed, production takes place in the usual way[9]. The result is that now the same intermediate input factors have different prices in the two countries and that within each market all variants bear the same price regardless of whether the were produced in the home or foreign country.

In order to solve the model analytically one restriction concerning the parameters of the model has to be made. The returns to differentiation, ν, have to be equal to $\frac{1-\alpha}{\alpha}$ to yield a closed solution of the model

$$Y_{i,k} = \left[\int_0^{N_i} (\lambda_j x_j)^\alpha dj \right]^{\frac{1}{\alpha}}, \tag{3.85}$$

where $i = H, L$ denotes the sector and $k = 1, 2$ the country where production takes place.

Since the elasticity of the demand schedule with respect to the quality level is the same in both countries, producers of the intermediate input factors choose the same quality regardless of whether they produce for their home market or for export purposes. The quality level then grows with rate $\frac{\alpha}{1-\alpha}\frac{1}{\mu} - 1$. Also it is assumed that in the past the quality of intermediate input factors in both economies started from the same level. Given that the demand schedules have the same price elasticity in both countries, this implies that the prices that have to be payed in both countries for the intermediate input factors are a mark-up over local wages. The demanded quantity of one particular variant of the intermediate input factor in the home and the export market is then given by:

$$\tilde{x}^D_{j,i,k} = \lambda_j^{\frac{\alpha}{1-\alpha}} \left[\left(\frac{w_{i,1}}{\alpha p_{j,1}} \right)^{-\frac{1}{1-\alpha}} Y_{i,1} + \left(\frac{w_{i,2}}{\alpha p_{i,2}(t)} \right)^{-\frac{1}{1-\alpha}} Y_{i,2} \right]. \tag{3.86}$$

In equilibrium the real profits earned by the intermediate producers equal the real

firm can insure against exchange rate uncertainty. Here the insurance is against price uncertainty

[9] Since production with the competitive technology does not yield any profits, it is assumed that no firms enter the markets with this technology.

fixed costs which have to be incurred to enhance the quality level

$$\frac{\pi_{i,j,k}}{p_k} = \frac{1-\alpha}{\alpha}\alpha^{\frac{1}{1-\alpha}}\frac{w_{i,k}}{p_k}\bar{\lambda}_i^{\frac{\alpha}{1-\alpha}}\left[\left(\frac{w_{i,1}}{p_{i,1}}\right)^{-\frac{1}{1-\alpha}}Y_{i,1} + \left(\frac{w_{i,2}}{p_{i,2}}\right)^{-\frac{1}{1-\alpha}}Y_{i,2}\right] = e^{\frac{\alpha}{1-\alpha}}f\frac{w_{i,k}}{p_k}, \quad (3.87)$$

where p_k denotes the aggregate price level in country k.

Together with the labor market clearing conditions

$$\tilde{x}^D_{j,H,k} = \frac{H_k}{N_{H,k}}, \quad (3.88)$$

$$\tilde{x}^D_{j,L,k} = \frac{L_k}{N_{L,k}}, \quad (3.89)$$

this yields the equilibrium number of producers of intermediate input factors in each market

$$N_{H,k} = H_k\frac{1-\alpha}{\alpha}e^{-\frac{\alpha}{1-\alpha}}f^{-1} \quad (3.90)$$

and

$$N_{L,k} = L_k(t)\frac{1-\alpha}{\alpha}e^{-\frac{\alpha}{1-\alpha}}f^{-1} \quad (3.91)$$

and the total number of intermediate input factors

$$N_H = (H_1 + H_2)\frac{1-\alpha}{\alpha}e^{-\frac{\alpha}{1-\alpha}}f^{-1}, \quad (3.92)$$

and

$$N_L = (L_1 + L_2)\frac{1-\alpha}{\alpha}e^{-\frac{\alpha}{1-\alpha}}f^{-1}. \quad (3.93)$$

Note that both sets of input factors are available in both countries.

The amount of high and low skilled goods produced by both countries together can be obtained by noting that the local demand for intermediate input factors equals

$$x_{j,i,k}^{D} = \frac{\left(\frac{w_{i,k}}{p_{i,k}}\right)^{-\frac{1}{1-\alpha}} Y_{i,k}}{\left(\frac{w_{i,1}}{p_{i,1}}\right)^{-\frac{1}{1-\alpha}} Y_{i,1} + \left(\frac{w_{i,2}}{p_{i,2}(t)}\right)^{-\frac{1}{1-\alpha}} Y_{i,2}} \tilde{x}_{j,i,k}^{D}. \tag{3.94}$$

Using this result together with the production function yields the following relationship for country 1

$$\left(\frac{w_{i,1}}{p_{i,1}}\right)^{-\frac{1}{1-\alpha}} Y_{i,1} + \left(\frac{w_{i,2}}{p_{j,2}}\right)^{-\frac{1}{1-\alpha}} Y_{i,2} = \frac{\alpha}{1-\alpha} f e^{\frac{\alpha}{1-\alpha}} N_i^{\frac{1}{\alpha}} \bar{\lambda}_i \left(\frac{w_{i,1}}{p_{i,1}}\right)^{-\frac{1}{1-\alpha}}, \tag{3.95}$$

where an analogous equation holds for country 2.

So far the amount of each good produced in one country has not been determined. To achieve this one has to impose a trade balance restriction. If the above assumed trading institutions are true for each sector, the amount of input factors exported must equal the amount imported in each sector:

$$\int_0^{N_{H,1}} \tilde{x}_{j,1}^{D} - x_{j,1}^{D} dj = \int_0^{N_{H,2}} \tilde{x}_{j,2}^{D} - x_{j,2}^{D} dj$$

and

$$\int_0^{N_{L,1}} \tilde{x}_{j,1}^{D} - x_{j,1}^{D} dj = \int_0^{N_{L,2}} \tilde{x}_{j,2}^{D} - x_{j,2}^{D} dj.$$

These conditions together with the above results lead to

$$\frac{Y_{H,1}}{Y_{H,2}} = \frac{H_1}{H_2} \left(\frac{w_{H,2}}{w_{H,1}} \frac{p_{H,1}}{p_{H,2}}\right)^{-\frac{1}{1-\alpha}} \tag{3.96}$$

and

$$\frac{Y_{L,1}}{Y_{L,2}} = \frac{L_1}{L_2} \left(\frac{w_{L,2}}{w_{L,1}} \frac{p_{L,1}}{p_{L,2}}\right)^{-\frac{1}{1-\alpha}}. \tag{3.97}$$

Together with equation (3.95) the following reduced form production functions can be computed:

$$\begin{aligned} Y_{H,1} &= N_H^{\frac{1-\alpha}{\alpha}} \bar{\lambda}_H H_1, \\ Y_{L,1} &= N_L^{\frac{1-\alpha}{\alpha}} \bar{\lambda}_L L_1, \\ Y_{H,2} &= N_H^{\frac{1-\alpha}{\alpha}} \bar{\lambda}_H H_2, \\ Y_{L,2} &= N_L^{\frac{1-\alpha}{\alpha}} \bar{\lambda}_L L_2. \end{aligned}$$

These reduced form production functions show that output in equilibrium is determined by the respective workforce in the particular country. Each country can through trade participate in the stock of knowledge embodied in the total number of intermediate input factors. The scale effect comes into existence through the relationship between this number and the world wide number of workers in each sector via equation (3.92).

Now from the demand schedule (3.86), the pricing rule and the labor market clearing conditions (3.88) and (3.89), the above reduced form production functions and equations (3.96) and (3.97) the wages for the high and low skilled can be deducted as

$$w_{H,k} = p_{H,k} \bar{\lambda}_H \alpha e^{-1} f^{-\frac{1-\alpha}{\alpha}} \left(\frac{\alpha}{1-\alpha} \right)^{-\frac{1-\alpha}{\alpha}} (H_1 + H_2)^{\frac{1-\alpha}{\alpha}}$$

and

$$w_{L,k} = p_{L,k} \bar{\lambda}_L \alpha e^{-1} f^{-\frac{1-\alpha}{\alpha}} \left(\frac{\alpha}{1-\alpha} \right)^{-\frac{1-\alpha}{\alpha}} (L_1 + L_2)^{\frac{1-\alpha}{\alpha}}.$$

Again the scale effect turns out in these equations through the dependence of the wage rate on the total number of workers in each sector, but workers in both countries benefit equally because of trade.

Since relative sector prices in each country are given by

$$\frac{p_{H,k}}{p_{L,k}} = \left(\frac{Y_{H,k}}{Y_{L,k}} \right)^{\rho-1},$$

the relative wage for the high and low skilled is given by

$$\frac{w_{H,k}}{w_{L,k}} = \left(\frac{\bar{\lambda}_H}{\bar{\lambda}_L}\right)^{\rho} \left(\frac{H_1 + H_2}{L_1 + L_2}\right)^{\rho\frac{1-\alpha}{\alpha}} \left(\frac{H_k}{L_k}\right)^{\rho-1}. \tag{3.98}$$

The two last terms on the right hand side of equation (3.98) reflect the scale and the substitution effect, respectively. The first operates through the number of intermediate input factors available to each sector. Since these factors are traded internationally the work force of *both* countries matters and determines the productivity in each sector. The substitution effect is given by the extent of the sector production within each country and therefore here only the number of workers in each country enter equation (3.98).

With one large and one small country the following situation can arise (let country 2 be the large country and country 1 the small one). A rise in the relative supply of high skilled in the large country induces a scale effect in favor of the high skilled through a richer set of intermediate input factors. Again the scale effect operates via the returns to differentiation as in the preceding sections. From equation (3.98) it can be seen that this carries directly over to the small country and increases wage inequality. If the small country tries to keep its relative wage fixed, i.e. relative wage rigidity, it can do so by increasing $\frac{H_1}{L_1}$ if the following inequalities hold

$$\rho\frac{1-\alpha}{\alpha}\frac{H_1}{H_1 + H_2} - (1-\rho) < 0 \tag{3.99}$$

and

$$\rho\frac{1-\alpha}{\alpha}\frac{L_1}{L_1 + L_2} - (1-\rho) < 0.$$

The mentioned increase in $\frac{H_1}{L_1}$ can be achieved by skilling up the work force or if this is not possible in an appropriate manner, through laying off some low skilled workers.

By the same time the original increase in the relative supply of the high skilled in the large country yields there increased wage inequality if

$$\begin{aligned} \rho\frac{1-\alpha}{\alpha}\frac{H_2}{H_1+H_2}-(1-\rho) &> 0 \\ \rho\frac{1-\alpha}{\alpha}\frac{L_2}{L_1+L_2}-(1-\rho) &> 0. \end{aligned} \tag{3.100}$$

It is obvious that conditions (3.99) and (3.100) can apply only in the explained situation that country 1 is small and country 2 is large. Then an increase in the relative supply of the high skilled in the large country increases the wage inequality there and an increase in the relative supply of the high skilled decreases the wage inequality there.

3.5 Exogenous Technology Shocks

In the analysis of section 3.2 it has been shown under what circumstances an increasing relative supply of high skilled workers can lead to an increase in wage inequality between high and low skilled. This is possible through a market size effect, i.e. an increasing number of variants of intermediate input factors leading to a more differentiated production process. However, this effect is transmitted through one single parameter of the model, the returns to differentiation ν. If this parameter takes the extreme value of zero there are no gains from an increasing differentiation and therefore no rising wage inequality can emerge from a change in the skill composition of the workforce. Note also that only with this parameter constellation has the model a scale effect neither in the growth rate nor in per capita production. Growth in per capita production then works only through an ever increasing level of quality of the intermediate input factors.

If one takes this as a reasonable economic setup the only way to explain a rise in the wage inequality between high and low skilled workers in this model is through exogenous shocks to the quality level. As can be seen from the derived relationships for the relative wage in the preceding sections, the relative wage for the high skilled increases with the average quality level of intermediate input factors produced by the high skilled if ε is greater than one.

In the formulation of the model used in the above sections such a shock has a permanent effect on the wage inequality due to the cost function of quality improvements (3.6). With two sectors one could also make an alternative assumption about this function. As it stands so far, the cost function exhibits total state dependence. Quality improvements in past periods in one sector decrease only the cost of further enhancements in the same sector. To allow for spillover effects between sectors one could assume the following cost functions

$$F_H = \begin{cases} f e^{\mu \frac{\lambda_{H,j}}{\bar{\lambda}^{\kappa}_{H,t-1}\bar{\lambda}^{1-\kappa}_{L,t-1}}} & \text{if} \quad \lambda_{H,j} \geq \bar{\lambda}^{\kappa}_{H,t-1}\bar{\lambda}^{1-\kappa}_{L,t-1}, \\ f e^{\mu} & \text{otherwise,} \end{cases}$$

and

$$F_L = \begin{cases} f e^{\mu \frac{\lambda_{L,j}}{\bar{\lambda}^{\kappa}_{L,t-1}\bar{\lambda}^{1-\kappa}_{H,t-1}}} & \text{if} \quad \lambda_{L,j} \geq \bar{\lambda}^{\kappa}_{L,t-1}\bar{\lambda}^{1-\kappa}_{H,t-1}, \\ f e^{\mu} & \text{otherwise.} \end{cases}$$

Obviously the parameter κ now determines the degree of state dependence. If κ equals one then there is total state dependence and past quality improvements in one sector only benefit this sector. If κ equals 0.5, there is no state dependence at all and both sectors profit equally from past quality improvements in one particular sector.

Using the Ethier (1982)/Dixit and Stiglitz (1977) specification with $\nu = \frac{1-\alpha}{\alpha}$ of section 3.2 as an example, the optimal decision of the quality level in the high skilled sector now satisfies

$$\frac{\lambda_{H,j}}{\lambda_{H,t-1}} = \frac{1}{\mu}\frac{\alpha}{1-\alpha}\left(\frac{\bar{\lambda}_{L,t-1}}{\bar{\lambda}_{H,t-1}}\right)^{1-\kappa},$$

since all producers of intermediate input factors choose the same level of quality. The condition for the low skilled sector is given by an analogous equation. Using both conditions and observing that qualities are symmetric within sectors it follows that

$$\frac{\bar{\lambda}_H}{\bar{\lambda}_L} = \left(\frac{\bar{\lambda}_{H,t-1}}{\bar{\lambda}_{L,t-1}}\right)^{2\kappa-1}.$$

With these two relationships one can solve for an analytic expression describing the adjustment path of the quality levels from given initial quality levels $\bar{\lambda}_{H,0}$ and $\bar{\lambda}_{L,0}$

$$\bar{\lambda}_H = \left(\frac{1}{\mu}\frac{\alpha}{1-\alpha}\right)^t \bar{\lambda}_{H,0}^{\frac{1}{2}[1+(2\kappa-1)^t]} \bar{\lambda}_{L,0}^{\frac{1}{2}[1-(2\kappa-1)^t]}$$

and

$$\bar{\lambda}_L = \left(\frac{1}{\mu}\frac{\alpha}{1-\alpha}\right)^t \bar{\lambda}_{L,0}^{\frac{1}{2}[1+(2\kappa-1)^t]} \bar{\lambda}_{H,0}^{\frac{1}{2}[1-(2\kappa-1)^t]}.$$

From these two equations one can read off the adjustment to the new equilibrium after a shock to $\bar{\lambda}_{H,0}$ or $\bar{\lambda}_{L,0}$. If κ is less than one the new equilibrium is characterized by an equal level of quality in both sectors, although it may take many periods to get close to this equilibrium. Only in the case of total state dependence the inequality in the level of quality after a shock persists forever. This directly carries over to wage inequality: As long as there is at least some deviation from total state dependence, wage inequality triggered by an exogenous shock to the quality level is a purely transitory phenomenon.

3.6 Neither "Strong" nor "Weak" Scale Effects

The last section presented a case for a growth model with neither "strong" nor "weak" scale effects. However it rested on a razor's edge condition, namely $\nu = 0$. This section presents jet another model without any scale effects to demonstrate that in such a case raising relative supply of one type of labor can not lead to increased wage inequality in favor of this type.

Consider an economy which produces final output Z_i in sector i through the following production technology

$$Z_i = \int_0^{N_i} (\lambda_{i,j} z_{i,j})^\delta dj,$$

where $\lambda_{i,j}$ is again the quality level and $z_{i,j}$ is the quantity used in production of an intermediate input factor. $\delta \in (0, 1)$ determines its marginal productivity. It is clear that there are now decreasing returns to scale in the use of the intermediate input factors given the set of available input factors N_i. It is assumed that profits are distributed equally among the consumers of the economy.

This model is even simpler to solve than the models in sections 3.2 and 3.3. Therefore only the results are presented here. It is assumed that the developer of an intermediate input factor has a comparative advantage in producing his particular variant and can set prices as a mark-up $\gamma > 1$ over marginal costs. The intermediate input factors are produced using sector specific labor only and the growth of the quality level is modelled as in sections 3.2 and 3.3. R&D is conducted by sector specific labor. Computing the demand function for one input factor and using the zero profit condition yields an equilibrium number of input factors

$$N_i = \frac{\gamma - 1}{\gamma} f^{-1} e^{-\frac{\delta}{1-\delta}} L_i,$$

where L_i is the labor supply to sector i. Again the number of variants of the intermediate input factors for a sector is directly proportionate to its scale of the work force.

Since the intermediate input factors are produced from labor, labor market clearing demands

$$z_{i,j} = \frac{1}{\gamma} \frac{L_i}{N_i}$$

With this results the production function for one sector i can be written as

$$Z_j = \left(\frac{\bar{\lambda}_i}{\gamma}\right)^\delta N_i^{1-\delta} L_i^\delta = \left(\frac{\gamma - 1}{\gamma}\right)^{1-\delta} \left(\frac{\bar{\lambda}_i}{\gamma}\right)^\delta f^{-(1-\delta)} e^{-\delta} L_i$$

From this equation it is clear that there are constant returns to scale in the aggregate sector production although there a diminishing returns to scale to the intermediate input factors before the adjustment of the number of variants is taken into account. Also it is obvious that per capita production is now free of any scale effect, i.e. the weak scale effect disappeared. The model still predicts productivity growth through growth in the average quality level for intermediate input factors in sector i which grows at rate $\frac{1}{\mu}\frac{\delta}{1-\delta} - 1$ from period to period. If one assumes now a simple linear aggregation technology[10] $Z = \sum_{i=1}^{I} Z_i$, where Z is the consumption good produced from output of the I sectors, it follows that labor supply does not influence wage inequality in the long run. The relative wage between workers of two sectors k and l is now given by

$$\frac{w_k}{w_l} = \left(\frac{\bar{\lambda}_k}{\bar{\lambda}_l}\right)^{\delta}.$$

Here long-run means after adjustment of technology through adjustment in N_i.
The whole discussion in this section made clear that in technology driven models the source of wage inequality is only the presence of weak or strong scale effects. If these effects are absent from a model, as in this subsection, wage inequality as a response to changes in labor supply disappears.

3.7 Conclusion

The models developed in this chapter built on the idea of growth without "strong" scale effects introduced by Young (1998) with vertical and horizontal product innovations. The original model was extended in the first part to cover a two sector economy employing two types of labor, high and low skilled. It has been shown that the remaining scale effect in the per capita terms of such growth models may serve as an explanation for the rising wage inequality between high and low skilled

[10]This assumes away any price effect, i.e. there is no substitution effect any more. This assumption does not affect the result of this section and is made only for the ease of computation. In a model with imperfect substitution there would be a pure negative relationship between relative supply and relative wages.

workers despite the rising relative supply of the high skilled.

As regards the Ethier (1982) specification of production with fixed costs in the production of the differentiated input factors several cases of the factors covering these fixed costs have been examined. The general conclusion from these exercises is that a large elasticity of substitution is needed in the aggregation of sector output to explain the observed wage pattern over time.

The second part of the chapter transferred the idea of fixed costs onto the Romer (1987) specification of the production function used in growth models. Although not for all possible cases of production a balanced growth path exists, for some it does. When a balanced growth path exists, the conclusion from the Ethier (1982) specification carries over that a high elasticity of substitution is needed to explain the widening of the wage gap between high and low skilled workers.

The third part of the chapter concentrated on unemployment in the context of an open economy model, i.e. the Krugman hypothesis. If a small country has the policy of keeping the relative wage for the high skilled fixed, it has been shown that via the development of new technology a rising high skilled supply in a large country can have negative impacts on employment of the low skilled in a small country if relative wages are rigid. However, this result depends crucially on strong assumptions about price setting and is questioning the validity of the Krugman hypothesis in a model of endogenous technical change.

The fourth section of the paper examined the behavior of the economy and in particular the relative wage in response to an exogenous shock to the quality level in one of the two sectors. It has been shown that if both sectors can profit from each other, such a shock to just one sector has a purely transitional impact on inequality and the relative wage returns to its level that prevailed before the shock occurred.

Finally, an endogenous growth model without any scale effect was developed to show that the positive relationship between rising relative supply of one type of labor and increasing wage inequality in favor of that type depends on the existence of at least "weak" scale effects.

In the next chapter the question about the existence of such scale effects is addressed empirically in a more macro-economic environment, taking account of the openness of economies. The approach is motivated by a theoretical model.

Chapter 4

Evidence for “Weak” Scale Effects

The previous chapters have shown that scale effects play an important role, not only for economic theory in general but also for certain applications, e.g. the relationship between labor supply and wage inequality. While the existence of “strong” scale effects has been rejected by the literature, evidence for or against “weak” scale effects is harder to find. This chapter tries to find such evidence on the macro-economic level, but not without a theoretical motivation of the empirical analysis.

4.1 Introduction

Studies trying to find evidence for the “weak” scale effect on the country level include Backus, Kehoe and Kehoe (1992), Sala-i-Martin (1997), Frankel and Romer (1999), Hall and Jones (1999), Alcala and Ciccone (2004) and Rose (2005). Although the studies use different methodologies, they have in common that they measure the scale of an economy by its own size, e.g. the population size or the extend of the work force.

Frankel and Romer (1999) analyze two cross sections, one of 150 countries and one of the 98 countries considered in Mankiw, Romer and Weil (1992), in 1985. They explain per capita income with the trade share, population and the country area.

Due to the possible endogeneity of trade, they use as instruments for trade the geographical characteristics of the trading partners to construct predicted values for trade. The final estimation is done by OLS and the authors find a significant positive impact of the population variable on per capita income with elasticities ranging from 0.12 to 0.35.

Hall and Jones (1999) estimate the relationship between output per worker and the social infrastructure in the particular country in 1988 for 127 countries. Social infrastructure is measured by an aggregate of an index of government anti-diversion policies and an index measuring the openness to trade. The measure of social infrastructure is instrumented by geographical characteristics. As an additional variable they add the country's population to the regression and obtain an estimated elasticity of 0.05, which is statistically insignificant at any considerable level of significance.

Backus, Kehoe and Kehoe (1992) are searching for effects of trade on growth. They find them in an extended empirical model where they regress the growth rate of production per capita in manufacturing industries and the average growth rate of GDP per capita between 1970 and 1985 on a trade index and, among other control variables, the average growth rate of the population from 1970 to 1985. Experimenting with different trade indices they estimate various elasticities of per capita production with respect to the population. They are all negative, in the case of the manufacturing sector they are not significant at the 10 percent level of significance, and range from -1.6 to -1.2.

Alcala and Ciccone (2002) estimate the effect of trade, the scale of production and institutional quality on per capita GDP using IV regression techniques separately for 1985 and 1990. As instruments they use, among others, geographical characteristics of the considered countries. They consider, like Frankel and Romer (1999), two sets of countries, one with 150 and one with 98 countries. The estimated elasticities of per capita GDP with respect to the workforce range from 0.14 to 0.46 and are all statistically significant.

Rose (2005) searches in a panel of 200 countries over 40 years for national scale effects. He controls for the economic and socio-economic environment of the countries and finds no indication of an influence of the size of a particular country on, among other things, income. None of the studies mentioned accounted for the possible role of the scale of the trading partners in the determination of per capita production.

Another perspective of looking at scale effects is the regional level of aggregation. One influential study is Ciccone and Hall (1996) who try to find empirical evidence for their theoretical model with increasing returns to scale. The implied hypothesis is that economically larger regions, i.e. regions with a higher density of production factors, ceteris paribus, have a higher labor productivity. Ciccone and Hall (1996) estimate a relationship between productivity and economic density for the US states and find strong support for their theory.

Fingleton (2001) uses a similar theoretical model as Ciccone and Hall (1996) to motivate his empirical study. He uses data for European regions and tries to explain the development of productivity in the manufacturing sector by the development, among other factors, of the population density in the different regions. He uses a spatial econometric model to account for spatial productivity effects and spatially correlated technology shocks. Fingleton (2001) finds reasonable evidence that a region's population density is a determining factor for manufacturing productivity in that region.

Ciccone (2002) is also concerned with labor productivity in Europe. He analyses the relationship between labor productivity and population density for a finer set of European regions than previously examined in Fingleton (2001). The theoretical argument for this relationship is again the model of Ciccone and Hall (1996). Ciccone (2002) obtains similar results for this relationship as previously found by Ciccone and Hall (1996) for the US.

One strand of the literature directly relevant to this chapter is concerned with convergence of income or per capita production between countries and regions. There can be found many empirical results in the literature. Maurseth (2001) reviews the theoretical and empirical literature on income disparities and convergence between countries and regions. He concludes that there are many theoretical reasons why income levels should and should not converge. While the neoclassical growth theory supports convergence, this might not take place in models of the new growth theory, depending on the spatial patterns of knowledge spillovers. More than the new growth theory does the new economic geography take account of these spatial patterns and therefore finds reasons for disparities in numerous models. A stylized fact seems to be that conditional convergence, after controlling for country or region specific effects, in income per capita took place at an annual rate of roughly 2% per annum between countries, European regions and US states up to the 1980s and slowed down afterwards[1]. There are several studies concluding that convergence failed to take place after the 1980s in the European regions (e.g. Neven and Goyette 1995, Fagerberg and Verspagen 1996; Quah 1996a and 1996b finds somehow opposite results). More recent studies as LeGallo (2004), Gardiner et al. (2004) also find evidence for low convergence in recent decades between European regions. Gianetti (2002) offers a theoretical explanation why income per capita may converge at the country level but does not so at the regional level. The mechanism behind his model is that countries consist of technologically heterogenous regions which force per capita income to be different within countries at the regional level but to converge between countries at the aggregate level. He finds empirical support for this hypothesis by looking at regional European data. Taking together this literature can be seen as supportive for the thesis that absolute convergence of per capita production is not likely to take place and this is as such one outcome of the theoretical model in this chapter. It is merely spatial heterogeneity which prevents regions from achieving the same per capita production.

One thing that all the previously mentioned studies have in common is a theoretical relationship between per capita production or some other measure of labor productivity in a specific region and the population density of that particular region. Some of the studies account for spatial effects through the specification of the empirical model. The present chapter aims to add to the literature by introducing a new perspective for looking at productivity or per capita production by directly introducing spatial effects into the theoretical model. This is done by using

[1] See Sala-i-Martin (1996).

an endogenous growth model and extending this by trade. This links regions[2] with each other and yields a relationship between per capita production and the scale of one region. Rather than measuring the scale of one economic unit simply by its population, in this context the scale of one region means something different. Productivity is determined by the available technology and technology created by one economic region is by itself determined by the extend of the work force, a general outcome of endogenous growth models. Therefore by scale of one geographical unit, the access to technology provided by itself and all other regions is meant. Through the link between technology and the extend of the work force, this gives rise to a scale variable specific to each economic region under investigation composed of the extend of the work forces of all regions, an inter-regional scale variable. This results in a spatial model linking per capita production and the effective scale of one region, i.e. an open economy analog of the "weak" scale effect known from closed economy growth models of the second generation type. This relationship serves as a starting point for the empirical analysis.

The argument of the preceding paragraph gains in importance if one thinks about the strengthening economic integration of the world. Some of the above mentioned studies try to account for this by controlling in their empirical work for trade. This might be a step in the right direction, but it seems more reasonable to account for economic interaction and integration by using the correct economic definition of the explanatory variables. Also, if the level of analysis is regional orientated, it might be hard to come by data on inter-regional trade, since there are hardly any data available. It is therefore important to correctly estimate these weak scale effects because these scale effects might play an important role in explaining productivity differences between countries and regions.

Concerning the link to the existing empirical literature it must be noted that this chapter borrows to some extent from the literature concerned with technology diffusion. Studies trying to measure knowledge or technology diffusion generally construct variables that should measure world wide available technology. This is usually

[2] In the following regions means a geographical unit like a country or a part of a country.

done by computing R&D stocks from historical investments in R&D or by historical patent behavior of sectors and countries. One influential study is Coe and Helpman (1995) who explain total factor productivity for the OECD countries and Israel with home and foreign R&D stocks. The foreign R&D stock is thereby a weighted sum of country specific R&D stocks. As weights Coe and Helpman (1995) use bilateral import shares between the home and foreign countries to compute the aggregated foreign R&D stock.

There is a number of studies building on the work of Coe and Helpman (1995) trying to refine their methodology (for a survey of the literature see Keller 2001). Most of this literature aims at finding better weights. One important publication is Keller (2002) who stresses the importance of geographical distance in the process of technology diffusion, an idea that is picked up in this chapter. This has the advantage that the weights are exogenous and therefore this approach will be applied in the empirical subsection below.

The empirical analysis in this paper is therefore to some extent related to the above cited articles and papers concerning technology diffusion due to the purpose of measuring an inter-regional scale variable. The main difference is that this chapter reduces technology to its model oriented origin, the extent of the work force. Articles dealing with technology diffusion generally do not go that far, but try to measure technology by using (accumulated) expenditures for technological purposes (R&D). In the empirical section of this paper data on production and the labor force for European and US regions as well as for a cross section of countries are used. In the regional analysis the relationship between per capita production and an inter-regional scale variable for 221 European NUTS (Nomenclature of Territorial Units for Statistics) regions in the EU 15 [3] and 3075 US counties is estimated. The EU 15 and the US were selected because previous studies focused on this group of regions and comparing the results of this chapter with their findings might be interesting. Another reason to look at the EU 15 and the US is the fact that trade of these economies is to a large extent intra-group trade. In 2003 more than 60% of imports

[3] These are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Ireland, Luxemburg, the Netherlands, Portugal, Spain, Sweden and the UK.

and exports of the EU 15 countries took place within the EU 15[4]. For the US also intra-regional trade is important due to the fact that usual measures of openness of the US economy are very low, i.e. the ratio of exports and imports to GDP was only 0.2620 in 2000[5].As will be clear from the theoretical model below, trade is the key to the mentioned inter-regional scale variable. This scale variable is a weighted sum of the work forces of all regions. In construction of this scale variable for each of the 221 regions the inverse great circle distance from the region under consideration to all other regions is used. Finally, since the regional work force is probably an endogenous variable, it is instrumented for by regional geographical characteristics in the case of the EU 15 and historical population for the US. The result of the estimation is that the inter-regional scale variable is a highly significant determinant of per capita production with an estimated elasticity of 0.45 in the EU 15 and 0.48 in the US.

The country level analysis of this chapter uses data on 88 countries for the year 2000. It will be shown that per capita GDP in these countries can be explained by the scale of technologically important partner countries, i.e. the G7 countries. A spatial scale variable will be constructed using also insights from the literature on technology diffusion (see Keller 2001) which serves to uncover the weak scale effect in an open economy context. The results indicate that this scale measure is significant in explaining variation in per capita GDP.

The results of the empirical analysis give further support on Jones' (2005) conclusion that the weak scale effect in second generation growth models is more a feature than a bug. Furthermore, as will be shown below, this has important implications for existence or non-existence of absolute convergence in per capita production.

The chapter is structured as follows. Section two develops a multi-regional growth model of the second generation type using the basic idea of Young (1998). The model accounts for trade between regions which are to some extent subject to trade frictions. The model accounts for perfect capital mobility between regions. The empirical part of this chapter is concentrated in sections three and four where the

[4]Published by Eurostat on its internet page.

[5]Penn World Tables 6.1.

data are described, estimation issues are discussed and results are presented. Finally section five concludes.

4.2 The Theoretical Model

This section develops a multi regional endogenous growth model with trade friction to highlight the importance of scale effects in explaining labor productivity. Production takes place in several stages: One sector is engaged in producing final output using labor and intermediate input factors. The second sector is producing these intermediate input factors with an increasing returns to scale technology. Before producing the intermediate input factors, firms have to incur quasi fixed R&D costs. Production then takes place at constant marginal costs which are caused by rented capital goods required to produce intermediate inputs.

The model employs the production technology familiar from Romer (1987, 1990) and combines it with the growth mechanism of Young (1998) to obtain a multi regional growth model. At the first sight the model seems to be similar to the model in Spolaore and Wacziarg (2005) but there are important differences. First Spolaore and Wacziarg (2005) do not account for steady state growth in there model. This is due to their assumption that technology is only given by the horizontal differentiation of production as in the first generation growth models (Romer 1987 or Grossman and Helpman 1991a). Second, and more important, they assume in a multi country and multi region setup capital immobility between countries besides trade in goods between regions and countries. This assumption merely serves as a capacity constraint to obtain a result for level of technology. In the model to be presented below, capital is allowed to move freely between regions; the necessary restriction to yield a solution for the level of technology is instead taken from the endogenous growth mechanism of the Young (1998) model which adds another dimension of growth through vertical innovations to the model. This gives a set of more economic plausible assumptions for a multi regional growth model.

Regions in this economic environment are assumed to be heterogenous with respect to several factors. First, it is assumed that every region is endowed with a given labor supply. Second, in every region firms producing intermediate input factors can enter the market. Finally, what is a distinct feature of the model with respect to the set-up in Ciccone and Hall (1996), the intermediate input factors can be traded between regions. Therefore every region can potentially access all variants of intermediate input factors. Nevertheless there exist transport costs in intermediate input factors modelled as in Samuelson (1954) as "iceberg" costs. For one unit of a particular intermediate input factor originating from region i to reach final good producers in region j $\tau_{i,j} = \tau_{j,i} > 1$ units have to be produced and shipped.

Time in the model is discrete. To simplify the notation the time subscript is suppressed where no confusion can occur, variables without time subscript correspond to the current time period t.

4.2.1 Households

The economic environment is assumed to admit a representative agent i who maximizes lifetime utility given by

$$U_t = \sum_{\tau=t}^{\infty} \frac{\ln c_{i,\tau}}{(1+\rho)^{\tau-t}}, \tag{4.1}$$

where $c_{i,t}$ is consumption in period t and ρ is the rate of time preference. Maximizing (4.1) with respect to an intertemporal budget constraint gives the well known optimality condition

$$\frac{c_{i,t+1}}{c_{i,t}} = \frac{1+r_{t+1}}{1+\rho}, \tag{4.2}$$

where r_t is the net interest rate of the economy. Households own the total capital stock. Capital goods can be linearly produced from final output with unit productivity and are traded freely between regions. The rate of depreciation on capital goods is denoted by δ. Full financial integration is assumed with an identical interest rate for all regions.

4.2.2 Production

Production in this multi regional context takes explicitly account of spatial interaction between regions. The general M region case is considered and production of final output in region i in time period t is given by

$$Y_i = L_{p,i}^{\alpha} \int_0^N (\lambda_j x_j)^{1-\alpha} dj. \tag{4.3}$$

$L_{p,i}$ is labor employed in production in region i and x_j denotes the quantity of the jth variant of an intermediate input factor used in the production of the final good Y_i and λ_j is its quality level. The total labor supply L_i to region i is given exogenously and it will become obvious below how $L_{p,i}$ is related to L_i. With this production function it is clear that productivity is determined by the available set of intermediate input factors N and their quality levels, i.e. economic growth can take place through vertical and horizontal technical innovations.

4.2.3 Growth

In order to solve the model one has to compute the set of available intermediate input factors. The assumptions concerning these are very similar as in Young (1998) and are as follows. Before entering the market for intermediate input factors a potential producer of the jth variant has first to decide every time period on the quality level. The chosen quality level determines the quasi fixed R&D costs in terms of labor according to the following real cost function

$$F_j = \begin{cases} f e^{\mu \lambda_j / \bar{\lambda}_{t-1}} & \text{if } \lambda_{j,t} \geq \bar{\lambda}_{t-1}, \\ f e^{\mu} & \text{otherwise,} \end{cases} \tag{4.4}$$

with $\bar{\lambda}_{t-1} = \frac{1}{N_{t-1}} \int_0^{N_{t-1}} \lambda_{j,t-1} dj$ as the average quality level in period $t-1$.

The optimal choice of λ_j is the quality level that maximizes the profits for the producer of one particular variant of the intermediate input factors. Once the quasi fixed costs for R&D have been incurred, the units of the particular intermediate input factors can be produced from capital goods with a linear production technology with unit productivity.

The particular intermediate input factor producer is faced with demand from all M regions including region i where his production is located. Given the production function (4.3), the demand function the producer j in region i is faced with is given by

$$x_j^d = \sum_{k=1}^{M} \left(\frac{\chi_{ik}}{p_k} \right)^{-\frac{1}{\alpha}} (1-\alpha)^{\frac{1}{\alpha}} \lambda_{j,t}^{\frac{1-\alpha}{\alpha}} L_{p,k}, \tag{4.5}$$

where χ_{ik} is the nominal price a producer from region i charges in region k, p_k is the price of the final good in region k. This demand function can be obtained by aggregating the single demand functions derived from marginal product conditions in the M different regions. Since capital goods have the same price in all regions and are produced from final output linearly, $p_k = p$ for all k, p is normalized to one in the following.

The individual intermediate input factor producer is assumed to possess some market power which allows him to set a price as a mark-up γ on marginal costs. Therefore for one unit of his product he charges the price $\chi_{ik} = \gamma \tau_{ik}(r + \delta)$.

The remaining problem of the producer in i of the jth variant of the intermediate input factors is to choose λ_j in order to maximize

$$\pi_{i,j} = (\gamma - 1)\gamma^{-\frac{1}{\alpha}} \sum_{k=1}^{M} (\tau_{ik}(r+\delta))^{-\frac{1-\alpha}{\alpha}} (1-\alpha)^{\frac{1}{\alpha}} \lambda_j^{\frac{1-\alpha}{\alpha}} L_{p,k} - w_i F_{i,j}. \tag{4.6}$$

Setting the derivative of (4.6) with respect to λ_j equal to zero and noting that entry into the market of intermediate input factors occurs until profits are driven down to zero, gives as the optimality condition

$$\frac{\lambda_j}{\bar{\lambda}_{t-1}} = \frac{1}{\mu} \frac{1-\alpha}{\alpha}, \tag{4.7}$$

which is very similar to the result in Young (1998). The optimality condition shows that all intermediate input factor producers chose the same quality level in period t given the average quality level in time period $t-1$, i.e. $\lambda_j = \bar{\lambda}$, and that the average quality level grows with a constant rate from period to period.

As mentioned before producers enter the market for intermediate input factors as long as there are profits to be earned. Thus equilibrium requires the profits to equal zero in all of the M regions of the economy[6]. This exactly gives M equations that can be solved for the M unknowns N_i, $i = 1, ..., M$, which give the number of intermediate input factors produced in region i. To find the solution, one first has to elaborate a little bit more on the R&D costs. These are costs in terms of labor and in this model labor earns the same wage rate regardless whether it is employed in production of the final good or R&D. This means that the wage rate, the intermediate input producer has to pay for workers employed in R&D, is equal to the marginal product of workers employed in production of the final good, i.e. $w_i = \alpha \frac{Y_i}{L_{p,i}}$.

Using the marginal product condition for the demand of intermediate input factors and integrating over all available variants in the production function (4.3) gives the reduced form

$$Y_i = (1-\alpha)^{\frac{1-\alpha}{\alpha}} \gamma^{-\frac{1-\alpha}{\alpha}} \bar{\lambda}^{\frac{1-\alpha}{\alpha}} (r+\delta)^{-\frac{1-\alpha}{\alpha}} L_{p,i} \left(\sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}} \right), \tag{4.8}$$

and therefore the real wage rate in region i as

$$w_i = \alpha(1-\alpha)^{\frac{1-\alpha}{\alpha}} \gamma^{-\frac{1-\alpha}{\alpha}} \bar{\lambda}^{\frac{1-\alpha}{\alpha}} (r_t+\delta)^{-\frac{1-\alpha}{\alpha}} \left(\sum_{j=1}^{M} N_j \tau_{ij}^{-\frac{1-\alpha}{\alpha}} \right), \tag{4.9}$$

Together with this last result the zero profit conditions for all M region imply

$$N_i = \frac{1-\alpha}{\alpha} \frac{\gamma-1}{\gamma} f^{-1} e^{-\frac{1-\alpha}{\alpha}} L_{p,i}, \tag{4.10}$$

Labor is divided into production and R&D. With (4.10) the number of workers

[6] In Spolaore and Wacziarg (2005) this assumption is replaced by a capital shortage constraint due to capital immobility. In an open economy context this is however unrealistic.

engaged in R&D, $L_{r,i}$ in region i is simply

$$L_{r,i} = N_i f e^{\frac{1-\alpha}{\alpha}} = \frac{1-\alpha}{\alpha}\frac{\gamma-1}{\gamma}L_{p,i}, \tag{4.11}$$

and therefore with $L_{p,i} + L_{r,i} = L_i$

$$L_{p,i} = \frac{\alpha\gamma}{\gamma+\alpha-1}L_i, \tag{4.12}$$

$$L_{r,i} = \frac{(1-\alpha)(\gamma-1)}{\gamma+\alpha-1}L_i. \tag{4.13}$$

With the results in (4.3), (4.8), (4.10) and (4.12) it is now easy to compute per capita production in reduced form

$$\frac{Y_i}{L_i} = c_1\bar{\lambda}^{\frac{1-\alpha}{\alpha}}(r+\delta)^{-\frac{1-\alpha}{\alpha}}\left(\sum_{j=1}^{M}\tau_{ij}^{-\frac{1-\alpha}{\alpha}}L_j\right), \tag{4.14}$$

$$c_1 = (1-\alpha)^{\frac{1-\alpha}{\alpha}}\gamma^{-\frac{1-\alpha}{\alpha}}f^{-1}e^{-\frac{1-\alpha}{\alpha}}\left(\frac{\alpha\gamma}{\alpha+\gamma-1}\right)^2\frac{\gamma-1}{\gamma}\frac{1-\alpha}{\alpha}.$$

This equation shows the relationship between per capita production and scale. The relevant figure determining scale is an inter-regional scale variable given by a weighted sum of work forces of all participating regions. The weights are given by functions in the transport costs. The mechanism behind this is that every region contributes to the available level of technology by providing intermediate input factors with a specific level of quality. Although the level of quality is identical in all regions, the other determinant of technology, the available set of intermediate input factors, is more heterogeneous. Each region is able to produce a set of these factors whose extent is directly proportional to its work force. Because of trade frictions the effective available set of intermediates is different for every region. Therefore per capita production or labor productivity is determined besides the quality level $\bar{\lambda}$ by the scale of a region given by its access to other regions. The relevant scale variable for one regional unit is thus not only its own size but a weighted sum of population sizes of regions with which trade takes place. This is also an open economy analogy to the terminology "weak" scale effect introduced in Jones (2005), the implication of second generation growth models that larger economies have a higher per capita production than smaller.

The strong result in (4.14) is that the elasticity of per capita production with respect to the inter-regional scale variable is equal to one. In the empirical section below a more general specification with an elasticity to be estimated will be employed.

Balanced Growth Path: Since the populations of the different regions are assumed to be stationary, the growth rate of production of final goods in every region is determined by growth of the quality level of intermediate input factors. The reduced form of the production function (4.3) is given by, using (4.3), (4.8), (4.10) and (4.12),

$$\begin{aligned} Y_i &= c_1 \bar{\lambda}^{\frac{1-\alpha}{\alpha}} (r_t+\delta)^{-\frac{1-\alpha}{\alpha}} L_i \left(\sum_{j=1}^{M} \tau_{ij}^{-\frac{1-\alpha}{\alpha}} L_j \right). \qquad (4.15) \\ c_1 &= (1-\alpha)^{\frac{1-\alpha}{\alpha}} \gamma^{-\frac{1-\alpha}{\alpha}} f^{-1} e^{-\frac{1-\alpha}{\alpha}} \left(\frac{\alpha\gamma}{\alpha+\gamma-1} \right)^2 \frac{\gamma-1}{\gamma} \frac{1-\alpha}{\alpha}. \end{aligned}$$

On the balanced growth path final output and consumption grow at the same rate determined by the optimality condition of the households

$$\frac{c_{t+1}}{c_t} = \frac{Y_{i,t+1}}{Y_t} = \frac{1+r_{t+1}}{1+\rho} = \left(\frac{\bar{\lambda}_{t+1}}{\bar{\lambda}_t} \right)^{\frac{1-\alpha}{\alpha}} \left(\frac{r_{t+1}+\delta}{r_t+\delta} \right)^{-\frac{1-\alpha}{\alpha}}. \qquad (4.16)$$

With a constant interest rate and the optimality condition (4.7) this gives

$$\frac{c_{t+1}}{c_t} = \left(\frac{1}{\mu} \frac{1-\alpha}{\alpha} \right)^{\frac{1-\alpha}{\alpha}} \qquad (4.17)$$

and (4.18)

$$r = \left(\frac{1}{\mu} \frac{1-\alpha}{\alpha} \right)^{\frac{1-\alpha}{\alpha}} (1+\rho) - 1. \qquad (4.19)$$

The zero profit condition for producers of intermediate input factors also implies that trade in intermediate input factors between regions is always balanced. Equilibrium in the regional markets for the final good implies that trade in final and capital goods is balanced as well. It can be shown that the model has the usual saddle path properties[7].

4.3 Empirical Analysis on the Country level

This section tries to test the result of the theoretical model of the preceding section using country data. A cross section of 88 countries in the year 2000 is analyzed in order to find support for the inter-regional analog of the "weak" scale effect. Special emphasis is on the source of this scale effect.

4.3.1 Data

For testing equation (4.14) empirically, data on per capita production, the scale of the technological important trading partners of the considered economies as well as on the transport costs are needed. For the cross section of countries the sample in Hall and Jones (1999) serves as a starting point.

The data used for per capita production is per capita GDP for the year 2000 taken from the Penn World Tables 6.1. The variable used is RGDPCH which is measured at purchasing power parity in 1996 US Dollars using a chain index. This makes the per capita GDP comparable across countries (see Summers and Heston 1991).

Finally, data on transport costs are needed. Since there are no data available for the considered cross section of countries, a proxy is used. It is well known that trade patterns follow geographical patterns, i.e. trade between neighboring countries is stronger than between countries that are separated by large distances (see e.g. Frankel and Romer 1999). It is therefore natural to assume that trading costs are tied to the distance between trading partners. As a proxy for transportation costs in the subsection below, functions of the great circle distances between the capital cities of the countries considered in the analysis and the G7 countries are used.

Data availability on GDP per capita in the Penn World tables restricts the original 120 country sample from Hall and Jones (1999). Furthermore city states like Hong Kong or Singapore were deleted from the sample. This results in a cross section of

[7] For a proof of local saddle path stability and additional calculations see the appendix at the end of the chapter.

88 countries listed in table 4.1.

4.3.2 Methodology and Results

The computation of the scale variable is of great importance for the empirical analysis in this section. From the theoretical point of view the scale of the economies is given by their own work force and the work force of the trading partners. Inspection of trade statistics reveals that almost every country in the world trades to some extent with every other country. Therefore it might seem reasonable to include the scales of all countries in some way in the scale variable for one particular economy under consideration. However, there are good reasons to deviate here a little bit from theory. 94% of all business enterprise R&D expenditure in the OECD countries is conducted by the G7 countries Canada, France, Germany, Italy, Japan, UK and USA (see e.g. Keller 2001). From the theory of the last section it became clear that the scale effect operates via technology which is determined by the work force of the countries performing R&D. A plausible way of calculating the necessary scale variable is therefore to proxy for scale by the extent of the populations in the G7 countries.

Another important point is the weighting scheme in the scale variable. From equation (4.14) it can be seen that the scale variable is a weighted sum of population sizes; functions in the transport costs determine the weights. From the above cited literature on technology diffusion two approaches can be adopted, the parametric and the non-parametric way of calculating such a scale variable. E.g. Coe and Helpman (1995) use import trade shares as weights while Keller (2002) uses exponential functions in the geographical distance as weights.

In this section both the non-parametric and the parametric approach will be explored to yields estimates of the "weak" scale effect in per capita production. The general model to be estimated is

$$\ln y_i = \alpha_0 + \alpha_1 \ln s_i + \beta x_i + \varepsilon_i, \tag{4.20}$$

where y_i is per capita GDP of country i, s_i denotes the scale variable to be defined

Table 4.1: List of Countries

1	Argentina	31	Guatemala	61	Panama
2	Australia	32	Guinea	62	Paraguay
3	Austria	33	Honduras	63	Peru
4	Bangladesh	34	Iceland	64	Philippines
5	Belgium	35	India	65	Portugal
6	Benin	36	Indonesia	66	Rwanda
7	Bolivia	37	Iran	67	Senegal
8	Brazil	38	Ireland	68	South Africa
9	Burkina Faso	39	Israel	69	Spain
10	Burundi	40	Italy	70	Sri Lanka
11	Cameroon	41	Jamaica	71	Swaziland
12	Canada	42	Japan	72	Sweden
13	Chad	43	Jordan	73	Switzerland
14	Chile	44	Kenya	74	Syria
15	Colombia	45	Korea, South	75	Tanzania
16	Congo	46	Lesotho	76	Thailand
17	Costa rica	47	Madagascar	77	Togo
18	Cote d'Ivoire	48	Malawi	78	Trinidad and Tobago
19	Denmark	49	Malaysia	79	Tunisia
20	Dominican Republic	50	Mali	80	Turkey
21	Ecuador	51	Mexico	81	Uganda
22	Egypt	52	Morocco	82	UK
23	El salvador	53	Mozambique	83	USA
24	Ethiopia	54	Nepal	84	Uruguay
25	Finland	55	Netherlands	85	Venezuela
26	France	56	New Zealand	86	Yemen
27	Gambia	57	Niger	87	Zambia
28	Germany	58	Nigeria	88	Zimbabwe
29	Ghana	59	Norway		
30	Greece	60	Pakistan		

below and x_i is a vector of controls. α_0, α_1 are parameters and β is a parameter vector to be estimated. ε_i is a usual error term. The scale variable is defined as

$$s_i = \begin{cases} \sum_{l=1}^{7} d_{il}^{-1} pop_l & \text{non-parametric,} \\ \sum_{l=1}^{7} e^{-\alpha_2 d_{il}} pop_l & \text{parametric,} \end{cases} \tag{4.21}$$

where l indicates the G7 countries, d_{il} is the great circle distance between country i measured in kilometers and the G7 country, pop_l is population in the G7 country and α_2 is a parameter to be estimated in the parametric case. In the non-parametric case the inverse of the distance is used as a weight as is often done in spatial econometrics (see e.g. Anselin (1988)). As the G7 countries are themselves part of the cross section the distances d_{ll} is set equal to one half of the square root of the land area of country l to approximate for transport costs within the country.

In the control vector x_i distance from the equator[8] and regional dummies are included for: Africa, Asia, Australia/New Zealand, Central America, the EU, Near East, South America and the Indian subcontinent; North America is the control group. The geographical controls mainly serve to account for spatial effects in per capita production not caused by spatial scale effects. Additional control variables were omitted because of several reasons. First equation (4.14) is a reduced form of the production function per capita. Thus it accounts for scale effect after all other variables like the physical or human capital have adjusted, in the latter case via the knowledge incorporated in the set of available intermediate input factors determined by the extent of the population. This reduced form is exactly what is to be estimated. There might be other factors not included in the theoretical model influencing per capita production like social or economic infrastructure. These variables are likely to be endogenous and useful instruments might be hard to come by. However, the regional controls might be good proxies for these variables and last but not least it is very unlikely that the scale variable defined in (4.21) is correlated with them.

Estimation of (4.20) is done by non-linear least squares in the parametric and OLS in

[8]The data for the distance from equator are taken from Hall and Jones (1999) provided through http://elsa.berkeley.edu/∼chad/datasets.html.

the non-parametric case, table 4.2 provides the results. Heteroskedastic consistent standard errors were computed using the White covariance estimator in its non-linear and linear version[9]. As this is a spatial analysis with geographical units, the residuals of the model might be subject to spatial dependence in which case the usual estimators for the covariance matrix of the parameters is invalid. To account for this, the robust covariance matrix proposed by Conley (1999) was computed as well and corresponding standard errors are reported[10].

In the parametric case the parameters of interest are clearly α_1 and α_2. The estimates for both coefficients in the first column of table 1 show the expected signs and are of magnitude 0.249 and -0.000268. However, looking at the estimated standard errors, both coefficients seem to be insignificant. Inspecting the data this seems to be merely a problem of collinearity. The correlation of the gradients of the regression function with respect to α_1 and α_2 at the estimated parameter values is 0.94, thus it is likely that the scale variable as defined above is nevertheless a significant determinant of per capita production. To explore this issue further two additional regression were estimated (column 2 and 3 in table 4.2). The first is a conditional estimation based on the point estimate of -0.000268 for α_2. This gives a statistical conditional significant estimate of the parameter α_1 showing that using a weight of $e^{-0.000268 d_{il}}$ the scale variable explains a significant part of the variation of per capita GDP in the cross section. The second estimate is the non-parametric model with inverse distances as weights. The coefficient of estimate 0.294 for α_1 is slightly higher than in the parametric specification but is again highly significant. Together these results strongly indicate that per capita GDP in this 88 country sample is influenced by a scale variable determined by the scale of the G7 countries as predicted by the theoretical model in section 2. Besides the studies of Frankel and Romer (1999) and Alcala and Ciccone (2002) this gives further support to the existence of

[9]In case of the non-linear estimation robust standard errors could also be computed by bootstrapping. This gives similar results.

[10]For this the Stata files scale.ado and x_ols.ado provided by T.G. Conley through http://www.gsb.uchicago.edu/fac/timothy.conley/research/ were used. scale.ado is a multidimensional scaling procedure using the approach proposed by Mardia et al. (1979). x_ols.ado computes the robust covariance matrix. Four spatial lags have been used for the computations, experimenting with more lags did not change the results.

Table 4.2: Estimation Results Country Level

Dependent variable: Log of GDP per capita			
Model:	parametric[a]	conditional[b]	non-parametric[c]
Log Scale	0.249	0.249	0.294
	(0.971)	(0.114)	(0.103)
		[0.107]	[0.096]
Distance	-0.000268	-0.000268	-
	(0.000969)		
Dist. equator	2.960	2.960	2.742
	(0.489)	(0.477)	(0.487)
		[0.446]	[0.455]
Africa	-1.957	-1.957	-1.930
	(0.284)	(0.283)	(0.305)
		[0.265]	[0.285]
Asia	-0.290	-0.290	-0.331
	(0.448)	(0.317)	(0.309)
		[0.292]	[0.289]
Australia/New Zealand	0.484	0.484	0.395
	(0.519)	(0.395)	(0.352)
		[0.369]	[0.329]
Central America	-0.791	-0.791	-0.722
	(0.329)	(0.324)	(0.346)
		[0.303]	[0.324]
EU	-0.620	-0.620	-0.620
	(0.228)	(0.225)	(0.255)
		[0.211]	[0.239]
Near East	-1.793	-1.793	-1.691
	(0.294)	(0.289)	(0.324)
		[0.270]	[0.303]
South America	-0.673	-0.673	-0.630
	(0.279)	(0.277)	(0.297)
		[0.259]	[0.278]
Sub Indian cont.	-1.727	-1.727	-1.676
	(0.393)	(0.392)	(0.405)
		[0.367]	[0.379]
Constant	5.900	5.900	7.396
	(12.671)	(1.409)	(0.615)
		[1.318]	[0.576]
Observations	88	88	88
R^2	0.889	0.889	0.890

[a] Scale variable with parametric weights (exponential functions) for the population sizes of the G7 countries. Estimation by non-linear least squares, heteroskedasticity consistent standard errors (in parentheses) computed using the non-linear version of the White covariance estimator.

[b] Scale variable with exponential functions in the distance as weights for the populations of the G7 countries. Distance parameter fixed at the value from the parametric model. Estimation by OLS, heteroskedasticity consistent standard errors in parentheses computed by the White covariance estimator, spatial dependence robust standard errors (Conley 1999) in braces.

[c] Scale variable with inverse distance as weights for the populations of the G7 countries. Estimation by OLS, heteroskedasticity consistent standard errors in parentheses computed by the White covariance estimator, spatial dependence robust standard errors (Conley 1999) in braces.

scale effects in per capita production, but this time using a different scale variable founded by a reasonable multi-regional endogenous growth model.

The results also have implications concerning convergence in per capita production. Maurseth (2001) reviews the theoretical and empirical literature on income disparities and convergence between countries and regions. He concludes that there are many theoretical reasons why income levels should and should not converge. In the model presented above the reason why convergence in per capita production does not take place, is the heterogeneity in transport costs, or transferred to the empirical results, the heterogeneity in geographical location relative to the G7 countries. Figure 4.1 tries to visualize this result by showing the effect of the scale variable. The displayed value is the exponential of the influence of the non-parametric scale variable, $\exp(\log(s_i))$ relative to the maximum value of all countries. The calculations were done for all countries in the world and not just for the cross section of the 88 countries considered for estimation. As expected, the map shows a clustering of high GDP per capita producing economies around the G7 countries. None of the countries has a value below 69% of the maximum value which is actually obtained for the Isle of Man due to the very central location. But nevertheless this clearly shows an impediment to absolute convergence in per capita GDP.

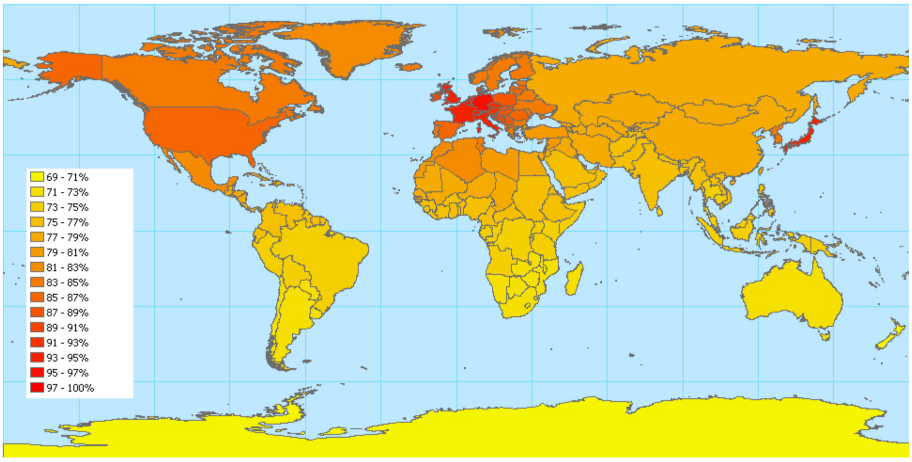

Figure 4.1: Scale effect of the G7 Countries

Influence of the non-parametric scale variable ($\exp(\alpha_1 \log s_i)$) as percentage of the maximum value in the sample.

4.4 Empirical Analysis on the Regional Level

This section empirically tests the theoretical results of the theoretical model, in particular equation (4.14) which gives a relationship between regional per capita production and an inter-regional scale variable. The regions under consideration in this section are European regions of the EU 15 (NUTS regions) and counties for the US.

4.4.1 European Data

The regional units under consideration are the NUTS2 regions of the EU 15 in the year 2002. For the EU 15 there are 214 NUTS2 regions. One exception is Denmark where NUTS2 regions are not defined. Therefore the 15 NUTS3 regions were used in the case of Denmark. The analysis below applies to the core regions of the EU 15, for France the 4 overseas Departments were excluded as well as the regions Azores and Madeira in the case of Portugal and the exclaves of Spain on the African continent Ceuta and Melilla. These peripheral regions are often subject to

Table 4.3: Descriptive statistics: GDP and population EU 15

	GDP per capita	Population in working age
Mean	35,565	1,115,814
Median	34,235	825,891
Max.	101,523	7,343,137
Min.	19,157	16,896
Std. dev.	10,886	1,045,513

Descriptive statistics for GDP per capita and population in working age for the EU 15 NUTS2 regions (NUTS3 for Denmark). GDP per capita in 2002 Euros at purchasing power parity. Data source: Eurostat REGIO database.

special economic conditions like tax exemptions which are not part of the theoretical analysis of section 2. It seems therefore reasonable to work with the remaining 221 regions.

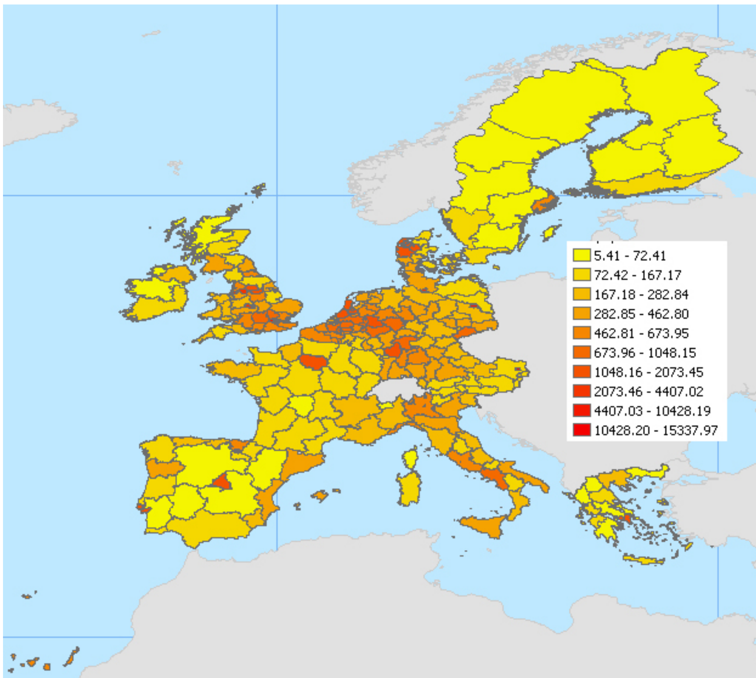

Figure 4.2: Population density EU 15

Population per square mile in European NUTS2 regions (NUTS3 for Denmark).

Per capita production is obtained from GDP measured at purchasing power parity divided by the regional population with the age from 16 to 64 years. These data were obtained from the Eurostat REGIO database. In the analysis below trade costs are proxied by geographical distance between regions. For this the great circle distances between the geographical centroids of the NUTS regions were computed.

Figure 4.2 shows the distribution of population density in the EU 15 regions as measured by regional population in working age per square mile. Figure 4.3 displays the distribution of GDP per capita in 2002 Euros at purchasing power parity. Some descriptive statistics about these figures are reported in table 4.3. Mean GDP per capita is 35,565 Euros, the maximum value is 101,523 Euros in Outer London (UK) and the minimum value is 19,157 Euros obtained by the NUTS2 region Norte in Portugal. The dispersion is quite high with a standard deviation of 10,886 Euros. Population in working age is distributed quite similar disperse. The region with the highest population is Ille de France containing Paris, the lowest population can be found in Åland, Finland.

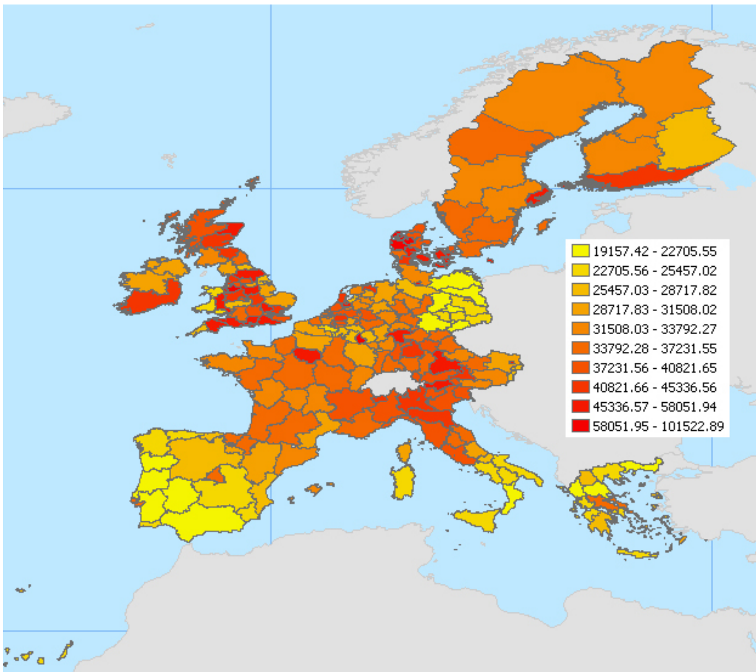

Figure 4.3: GDP per capita EU 15

Distribution of actual GDP per capita in the EU 15 NUTS2 regions (NUTS3 regions for Denmark). All figures correspond to 2002 Euros at purchasing power parity.

4.4.2 US Data

The regions under consideration for the US are the 3141 counties. The empirical analysis focuses on mainland USA due to geographical localization, thus the states Alaska and Hawaii are excluded from the sample. There are some statistical problems for a couple of counties in Virginia where larger cities within a rural area are defined as a separate county within a rural county. However these areas are not separated in the official statistics where they are treated as one unit. Therefore the counties with these characteristics were geographically merged and were treated as single geographical units. For reasons of data availability for some historical variables 5 more counties have to be excluded from the final estimation leaving in total 3075 observations.

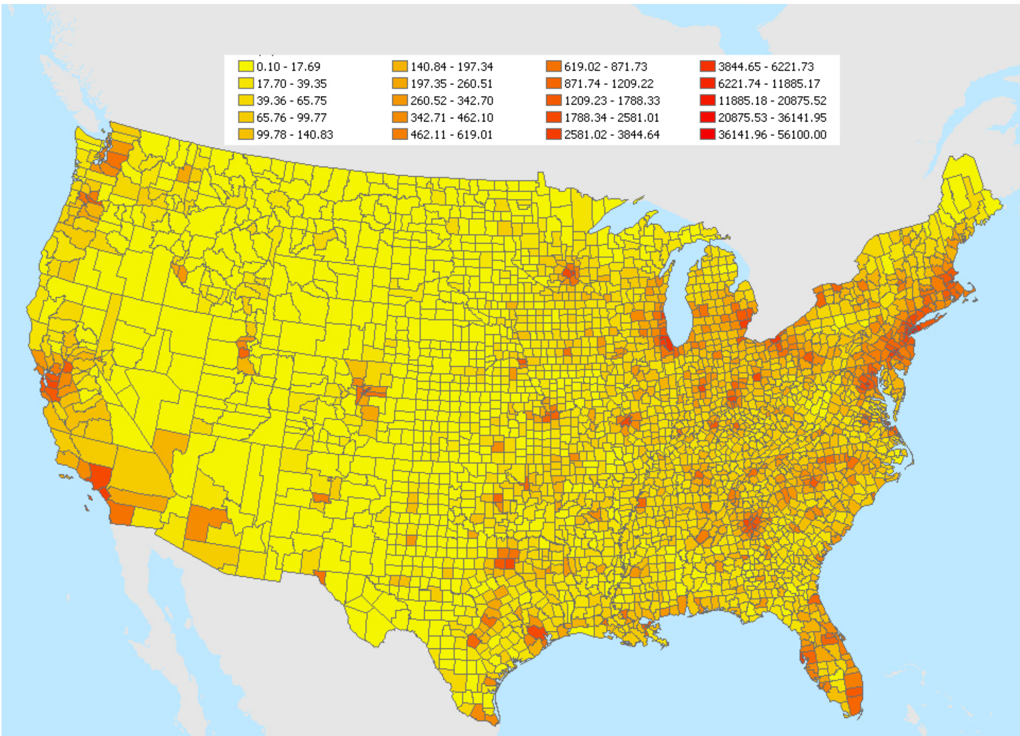

Figure 4.4: Population density US

Population per square mile in the US counties.

Table 4.4: Descriptive statistics: Personal income and population US

	Personal income per capita	Population
Mean	23,951	92,892
Median	23,055	25,262
Max.	82,486	9,763,815
Min.	5,813	66
Std. dev.	5,811	301,981

Descriptive statistics for personal income per capita and population for the US counties. Personal income per capita in 2002 US Dollars. Data source: US Bureau of Economic Analysis.

GDP or GDP per capita data are not available for the US at the county level. However the US Bureau of Economic Analysis provides data on estimated per capita personal income on the county level. Per capita personal income is defined as the sum of wages, salary disbursements, supplements to wages and salaries proprietors' income, dividends, interest, rents and personal current transfer receipts[11]. County population used in the empirical analysis below is measured on the 1st of July of the respective year.

Figure 4.4 displays the distribution of population measured by population per square mile in 2002. From this picture it can be seen clearly that there is a concentration of population on the east and southern west cost of the US with a large area in between with a low density of population. The highest density is obtained by New York City with about 55,500 inhabitants per square mile. Figure 4.5 shows the distribution of personal income per capita. The picture shows a large variation in the magnitude of personal income across states. Clustering of high income regions is present in the area of New York and the west coast of California, however other high income regions exist as well. Per capita personal income is again highest in New York City. Descriptive statistics for the US counties are reported in table 4.4.

[11]For this definition see also http://www.bea.gov/bea/regional/articles/lapi2003/sources.pdf.

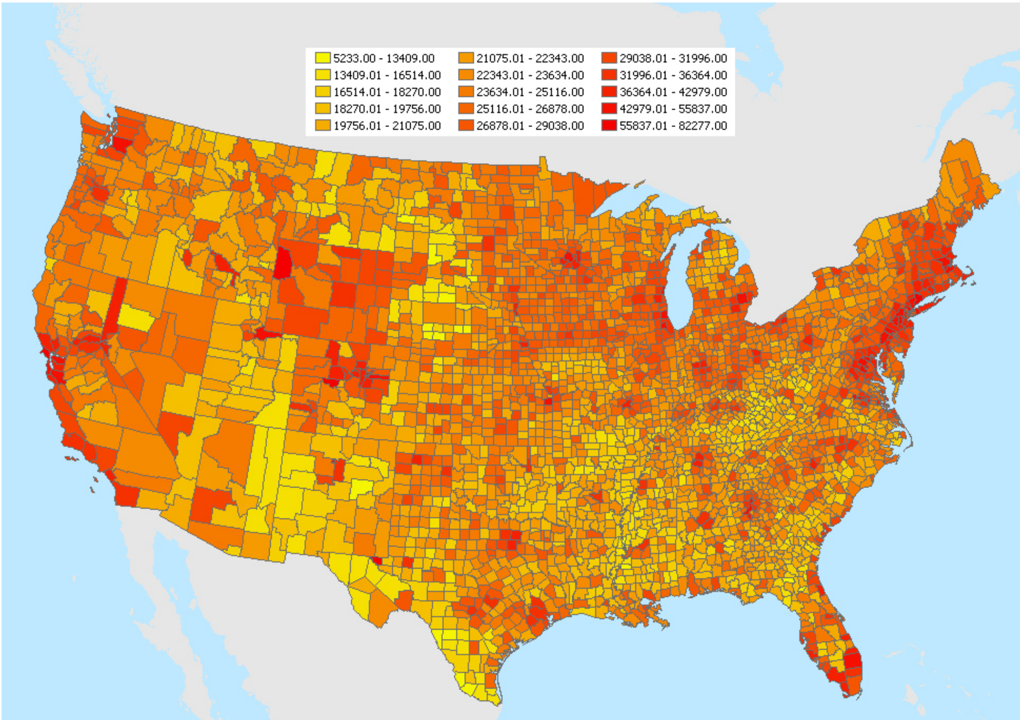

Figure 4.5: Personal income per capita US

Distribution of actual personal income per capita in the US counties. All figures correspond to 2002 US Dollars.

4.4.3 Estimation Issues

The empirical model is motivated by equation (4.14) of the theoretical section and is specified as

$$\ln\left(\frac{Y_i}{L_i}\right) = \alpha_0 + \alpha_1 \ln\left(\sum_{j=1}^{M} \omega_{ij} L_j\right) + \sum_h \delta_h d_h + \varepsilon_i. \tag{4.22}$$

d_h are country or state dummies defined for all countries or states. For the EU 15 Luxemburg serves as the base country. ε_i denotes an error term. In (4.20) the inter-regional scale variable is a weighted sum of regional work forces. The weights ω_{ij} are proxied by the inverse great circle distance between the regional centroids as is often done in spatial econometrics (see e.g. Anselin 1988). For the weights ω_{ii} one half of the square root of the regions land area is used to proxy for the average distance within a region. There is a growing literature on technology diffusion focusing on finding weights determining the diffusion process, a topic relevant at this point of the analysis. Some authors use trade shares (Coe and Helpman 1995), data on FDI (Lichtenberg and van Pottelsberghe de la Potterie 1996) or data on multinational enterprises (Xu 2000). One caveat by doing so is that all these weights are very likely to be endogenous and therefore might bias the estimate. Therefore the analysis below follows the findings of Keller (2002) by using geographical distance in constructing weights.

It is very likely that regional population is an endogenous variable. Regions with a high GDP or income per capita and therefore high wages might attract workers from other regions. There might be also externalities that affect the population of neighboring regions. To circumvent this problem the regional population is instrumented for the EU 15 by geographical characteristics and additionally by historical population in the US. In a first step regression for the EU 15 the regional population in working age is regressed on the regional land area, the squared area, area to the power of three and country dummies (table 4.5), the predicted values from this regression were then used to compute the scale variable in (4.22). For the US counties log population in 2002 is regressed on state dummies, the population growth rate from 1970 to 1971 and a polynomial of order three in the county land area, the slope coefficients for population growth are allowed to vary between states. The predicted values from this regressions were then used to compute the scale variable in (4.22).

As this is essentially a spatial economic analysis with geographical units, the residuals of the model (4.20) might be spatially autocorrelated. Standard OLS estimates of the parameters are in this case still consistent but the usual estimator for the covariance matrix of the coefficient is not. To make valid inferences about the parameters, standard errors are computed by the estimator proposed by Conley (1999) which accounts for spatial autocorrelation[12]. Additional standard errors using the

[12]For this the Stata files scale.ado and x_ols.ado provided by T.G. Conley were used. In the case of the EU 15 the coordinates of the centroids of the regions were projected on a two dimensional plane by the eigenvectors of the scaled distance matrix (see Mardia et al. 1979) and were then used to correct the covariance matrix of the estimated coefficients. In the case of the US counties

Table 4.5: First step regression for regional population EU 15

Dependent variable: Log population Variable	OLS est.	White std. errors
Area	$6.24\ 10^{-5}$	$1.97\ 10-5$
Area^2	$-8.50\ 10-10$	$3.62\ 10^{-10}$
Area^3	$3.10\ 10^{-15}$	$1.71\ 10^{-15}$
AT	12.65	0.261
BE	13.01	0.192
DE	13.49	0.172
DK	12.01	0.183
ES	13.18	0.317
FI	11.80	0.768
FR	13.09	0.359
GR	12.26	0.298
IE	12.70	0.526
IT	13.24	0.306
LU	12.45	0.049
NL	13.26	0.226
PT	12.71	0.413
SE	12.31	0.469
UK	13.25	0.145
Observations	221	
R^2	0.406	

First step regression results of population aged 16 to 64 in the European NUTS regions.

White estimator are reported as well.

Finally it must be noted that there might be variables other than those in (4.22) that influence per capita GDP or personal income. To justify the approach it can be argued that, first, equation (4.14) which serves the motivation for the empirical analysis is a reduced form and that the empirical model tries to estimate this reduced form after other endogenous factors have adjusted to the scale variable. Second, there are other factors not accounted for in the theoretical model, so that the last argument might not apply for these, but these variables might of course be endogenous and adding them to the empirical model without using suitable instruments is problematic. And third, it is very unlikely that the inter-regional scale variable is correlated with other economic variables, so that omitting them might not be that problematic.

the projection method which needs the calculation of the eigenvectors of the scaled distance matrix between 3075 counties was not applied because of the difficulties in computing eigenvectors of such a large matrix. Instead the latitude and longitude were projected on a two dimensional plane using the Azimuthal equidistant projection, a standard projection method in geography. The projection was done using the software package ArcGIS.

4.4.4 Results

Table 4.6 gives the results of the estimation of model (4.20). Of great importance is the estimate of the scale elasticity which takes a value of 0.45 for the EU 15 and 0.48 for the US and is highly significant in both cases meaning that an inter-regional defined scale variable is an important determinant of per capita production. This result can also be seen as support for the theoretical result of endogenous growth models to create weak scale effects.

Concerning the literature on convergence in per capita production or income, this result gives an argument why total convergence might not take place. Since geographical location and thereby access to the scale of other regions is heterogeneous, it is not to be expected that per capita production can converge as long as transport cost do not vanish. Figure 4.6 displays the influence of the scale variable at the estimated value for α_1, $\hat{\alpha}_1$, i.e. the geographical distribution of $\hat{\alpha}_1 \ln \sum w_{ij} L_j$. The influence is measured in percent of the highest value obtained in the sample which happens to occur for Inner London (UK). There is a clear tendency of clustering, as expected, in the central regions of the EU 15 countries. This pattern can also be observed in the distribution of actual GDP per capita in figure 1, although there some outliers. For the US the figures 4.4 and 4.7, which are constructed analogously, give the same impression. There is a clear tendency for the scale variable to cluster in the eastern and south western part which are populated more densely. On average these are the higher income regions. Thus the theoretical model of section 2 captures an important aspect of the distribution of per capita production in the European and US regions.

Table 4.6: Estimation Results

Dependent variable:	Log of GDP per capita (EU)	Log personal income per capita (US)
Log Scale	0.455	0.474
	(0.114)	(0.049)
	[0.088]	[0.049]
State dum.	-	yes
		$(\chi^2(48) = 6,358)^a$
		$[\chi^2(48) = 5,076]$
AT	-0.445	
	(0.071)	
	[0.063]	
BE	-0.668	
	(0.085)	
	[0.082]	
DE	-0.698	
	(0.039)	
	[0.036]	
DK	-0.096	
	(0.083)	
	[0.071]	
ES	-0.542	
	(0.088)	
	[0.070]	
FI	0.005	
	(0.180)	
	[0.148]	
FR	-0.532	
	(0.049)	
	[0.042]	
GR	-0.464	
	(0.142)	
	[0.113]	
IE	-0.263	
	(0.183)	
	[0.169]	
IT	-0.488	
	(0.064)	
	[0.054]	
NL	-0.609	
	(0.050)	
	[0.048]	
PT	-0.546	
	(0.157)	
	[0.133]	
SE	-0.143	
	(0.135)	
	[0.110]	
UK	-0.427	
	(0.057)	
	[0.046]	
const.	-6.728	
	(1.511)	
	[1.169]	
Observations	221	3,075
R^2	0.450	0.292

OLS estimation results of model (4.20). Column 2 contains estimation results for the EU 15 regions, Luxemburg is the base country. Column 3 contains estimation results for the US. White standard errors in parenthesis, Conley (1999) standard errors, corrected for spatial dependence, in braces.

[a]Null hypothesis: Identical coefficients for the state dummies.

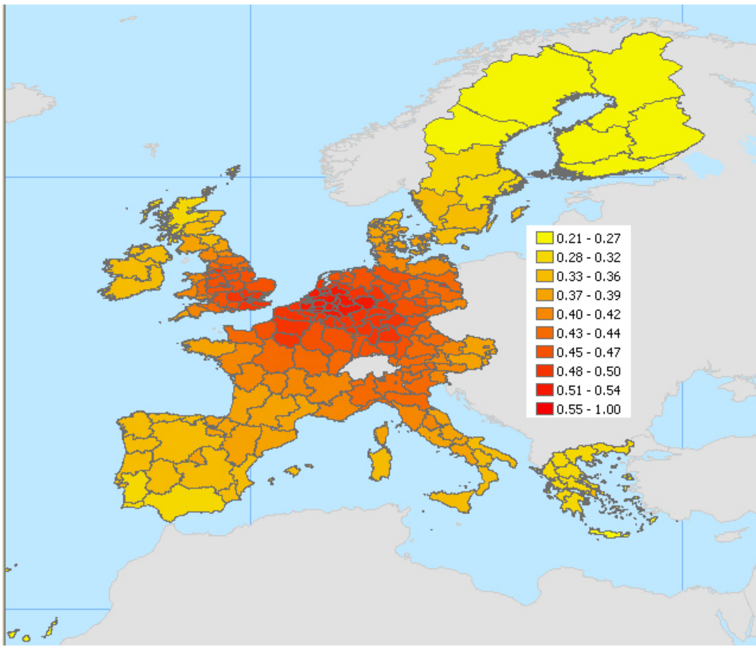

Figure 4.6: Influence of the scale variable EU 15

Influence of the scale variable on GDP per capita $(\exp(\hat{\alpha}_1 \log(\sum w_{ij} L_j)))$ at the estimated parameter value $\hat{\alpha}_1$ from table 4.6 as percentage of Inner London (UK).

Some comments are also in order with respect to the estimates in the literature. Fingleton (2001) uses total manufacturing output in one particular region as the scale variable. His estimates for the elasticity with respect to scale are somewhat higher ranging from 0.59 to 0.80. Ciccone (2002) uses population density in one particular region as the scale variable to explain production per capita. The elasticities he obtains for the European regions are about 0.05. However these results can not be directly compared with the ones in this paper since the definition of scale is here a different one, i.e. it takes trade into account whereas this is not included in the studies of Fingleton (2001) and Ciccone (2002). Ciccone and Hall (1996) estimated returns to scale in per capita production on the state level for the US using aggregated county data for economic density. They find similar elasticities as Ciccone (2002) for the European case.

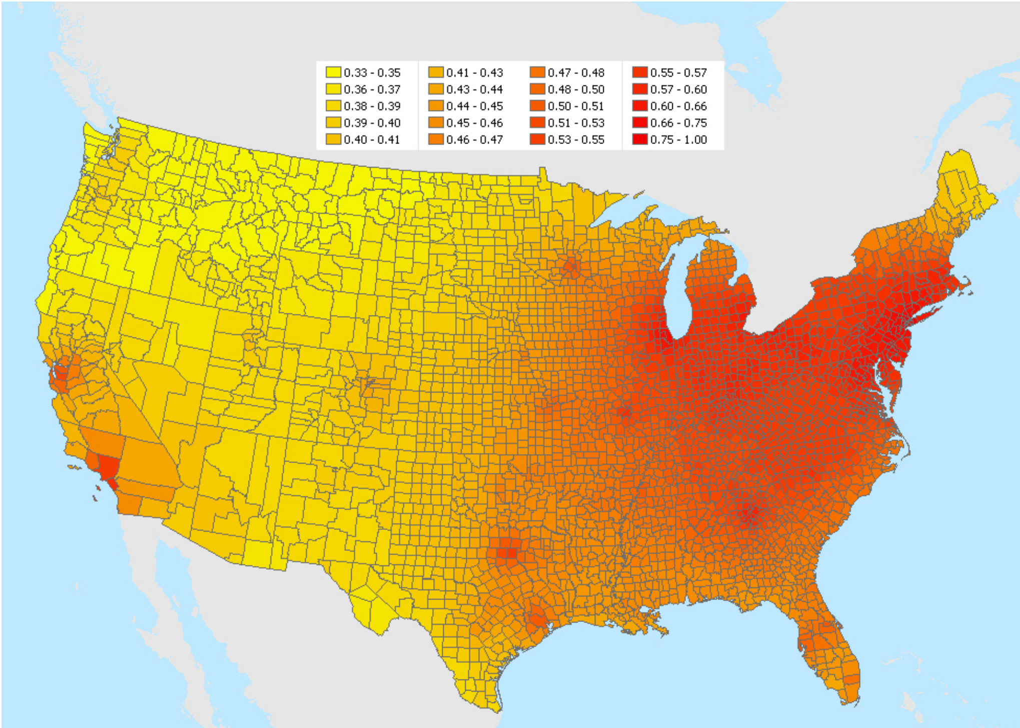

Figure 4.7: Influence of the scale variable US

Influence of the scale variable on GDP per capita ($\exp(\hat{\alpha}_1 \log(\sum w_{ij} L_j))$) at the estimated parameter value $\hat{\alpha}_1$ from table 4.6 as percentage of New York City.

4.5 Conclusion

The theoretical part of this paper has presented a multi-regional second generation endogenous growth model with inter-regional trade in intermediate or technology goods. This model shares the feature of second generation growth models that create "weak" scale effects. In the regional context this means that a region has a higher per capita production as it is economically more closely related to other regions giving it a better access to the economic size of partner regions. In reduced form this size is given by an inter-regional scale variable defined as weighted sum of population sizes of trading partner regions. The weights are given by functions of the trade frictions. This is a open economy analog of the "weak" scale effect of second generation growth models, i.e. a region can participate in the scale of other regions through trade.

This theoretical result is tested empirically by employing a new approach in defining the scale variable for the geographical units under consideration, a weighted sum of working age populations. As weight, as is often done in spatial econometrics, the inverse distance between units has been used. The results for a cross section of 88 countries in the year 2000 indicates that a scale variable composed of the scales of the G7 countries, the origin of most of the available technology, is a significant variable in explaining GDP per capita. The estimated elasticity ranges from 0.25 to 0.29, depending on the estimation technique.

On the regional level the geographical units considered were the NUTS regions of the EU 15 and the US counties. The results show a highly significant positive relationship between per capita GDP in the EU 15 regions and the inter-regional scale variable with an estimated elasticity of about 0.46. For the US using county level data on per capita personal income a comparable result with an elasticity of 0.47 is found. As has been shown this result has important implications for absolute convergence in per capita production since the regions considered are of course geographically heterogenous. This translates as shown by the theoretical model and the empirical results into heterogenous levels of per capita production.

These results give further support to the existence of weak scale effects in per capita production. and the corresponding assumptions in the second generation growth models seem reasonable. As such the same conclusion as in Jones (2005) can be drawn. i.e. the “weak” scale effects are more a feature than a bug to second generation growth models.

4.6 Appendix: Proof of local Saddle Path Stability and additional Calculations

Final and capital goods can be traded freely, consumer preferences and the interest rate are identical across regions. Therefore capital and final output as well as the development of consumption can be aggregated across all M regions.

Denote the capital stock necessary to provide region i with intermediate input factors as K_i[13]

$$K_i = \gamma^{-\frac{1}{\alpha}} \frac{\alpha\gamma}{\alpha+\gamma-1}(1-\alpha)^{\frac{1}{\alpha}} \bar{\lambda}^{\frac{1-\alpha}{\alpha}} (r+\delta)^{-\frac{1}{\alpha}} L_i \sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}}, \tag{4.23}$$

with N_j given by equation (4.10). Then

$$\int_0^N (\bar{\lambda} x_j)^{1-\alpha} dj = \gamma^{-\frac{1-\alpha}{\alpha}} \left(\frac{\alpha\gamma}{\alpha+\gamma-1}\right)^{1-\alpha} (1-\alpha)^{\frac{1-\alpha}{\alpha}} (r+\delta)^{-\frac{1-\alpha}{\alpha}} \bar{\lambda}^{\frac{1-\alpha}{\alpha}} L_i^{1-\alpha} \sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}},$$

with (4.23) this gives

$$\int_0^N (\bar{\lambda} x_j)^{1-\alpha} dj = K_i^{1-\alpha} \bar{\lambda}^{1-\alpha} \left(\sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}} \right)^{\alpha}$$

and

$$Y_i = \left(\frac{\alpha\gamma}{\alpha+\gamma-1} L_i \sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}} \right)^{\alpha} \left(\bar{\lambda} K_i\right)^{1-\alpha}.$$

The inter-regional capital stock K is then given by

$$K = (1-\alpha)^{\frac{1}{\alpha}} \bar{\lambda}^{\frac{1-\alpha}{\alpha}} (r+\delta)^{-\frac{1}{\alpha}} \frac{\alpha\gamma}{\alpha+\gamma-1} \sum_{i=1}^{M} \left(L_i \sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}} \right)$$

and therefore

$$K_i = \omega_i K$$

[13]As will be clear later, the capital required to produce the intermediates for one region is always identical to the amount of capital used by that region in order to provide the rest of the world with its own variants.

with

$$\omega_i = \frac{L_i \sum_{k=1}^{M} N_k \tau_{ik}^{-\frac{1-\alpha}{\alpha}}}{\sum_{h=1}^{M} \left(L_h \sum_{k=1}^{M} N_k \tau_{hk}^{-\frac{1-\alpha}{\alpha}} \right)}.$$

Aggregated production $Y = \sum_{i=1}^{M} Y_i$ can be written as

$$Y = \sum_{i=1}^{M} \left(\frac{\alpha\gamma}{\alpha+\gamma-1} L_i \sum_{k=1}^{M} N_k \tau_{hk}^{-\frac{1-\alpha}{\alpha}} \right)^{\alpha} \left(\bar{\lambda} w_i\right)^{1-\alpha} K^{1-\alpha}.$$

Therefore the interest rate is given by

$$r + \delta = (1-\alpha)\frac{Y}{K}.$$

The aggregate economy behaves according to

$$\begin{aligned}
K_{t+1} &= (1-\delta)K_t + Y_t - C_t, \\
C_{t+1} &= \frac{1 + d_1 \bar{\lambda}_{t+1}^{1-\alpha} K_{t+1}^{-\alpha} - \delta}{1+\rho} C_t, \\
C_t &= \sum_{i=1}^{M} L_i c_{i,t}, \\
d_1 &= (1-\alpha) \sum_{i=1}^{M} \left(\frac{\alpha\gamma}{\alpha+\gamma-1} L_i \sum_{j=1}^{M} N_j \tau_{hj}^{-\frac{1-\alpha}{\alpha}} \right)^{\alpha} (\omega_i)^{1-\alpha}.
\end{aligned}$$

Dividing the above system by $\bar{\lambda}_{t+1}^{\frac{1-\alpha}{\alpha}}$ and using (4.7) gives

$$\begin{aligned}
\tilde{k}_{t+1} &= (1-\delta)\frac{1}{d_2}\tilde{k}_t + \frac{d_1}{1-\alpha}\frac{1}{d_2}\tilde{k}_t^{1-\alpha} - \frac{1}{d_2}\tilde{c}_t, \\
\tilde{c}_{t+1} &= \frac{1 + d_1 \tilde{k}_{t+1}^{-\alpha}}{1+\rho}\frac{1}{d_2}\tilde{c}_t, \\
\tilde{c}_t &= \frac{C_t}{\bar{\lambda}_t^{\frac{1-\alpha}{\alpha}}}, \\
\tilde{k}_t &= \frac{K_t}{\bar{\lambda}_t^{\frac{1-\alpha}{\alpha}}}, \\
d_2 &= \left(\frac{1}{\mu}\frac{1-\alpha}{\alpha} \right)^{\frac{1-\alpha}{\alpha}}.
\end{aligned}$$

Linearizing around the steady state gives the Jacobian

$$
\begin{aligned}
J &= \begin{pmatrix} 1+\rho & -\frac{1}{d_2} \\ -\alpha \frac{r^{g*}}{d_2}\frac{\tilde{c}^*}{\tilde{k}^*} & 1+\alpha \frac{r^{g*}}{1+\rho}\frac{\tilde{c}^*}{\tilde{k}^*}\frac{1}{d_2^2} \end{pmatrix},
\end{aligned}
$$

with

$$
\begin{aligned}
r^{g*} &= (1+\rho)\left(\frac{1}{\mu}\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}} - 1 + \delta, \\
\frac{\tilde{c}^*}{\tilde{k}^*} &= 1 - \delta - d_2 + \frac{r^{g*}}{1-\alpha},
\end{aligned}
$$

which has two real eigenvalues where one is positive and smaller than one and one is positive and larger than one. Denote these eigenvalues by μ_1 and μ_2, it holds that[14]

$$
\begin{aligned}
\mu_1\mu_2 &= 1+\rho, \\
\mu_1+\mu_2 &= 1+\rho+1+\alpha\frac{r^{g*}}{1+\rho}\frac{c^*}{k^*}\frac{1}{d_2^2}.
\end{aligned}
$$

Therefore the eigenvalues solve

$$
\phi(\mu) = \mu + \frac{1+\rho}{\mu} = 1+\rho+1+\alpha\frac{r^{g*}}{1+\rho}\frac{c^*}{k^*}\frac{1}{d_2^2}
$$

But $\phi(1)$ which equals $1+\rho$ is smaller than $1+\rho+1+\alpha\frac{r^{g*}}{1+\rho}\frac{c^*}{k^*}\frac{1}{d_2^2}$. Therefore one eigenvalue is positive and smaller than one while the other is positive and larger than one.

It can also be shown that regional imports of intermediate input factors are always equal exports. To show this compute imports given by the demand for intermediate inputs coming from final good producers in region i using (4.23) as

$$
\gamma K_i(t) = \gamma^{-\frac{1-\alpha}{\alpha}}\frac{\alpha\gamma}{\gamma+\alpha-1}(1-\alpha)^{\frac{1}{\alpha}}(r^g)^{-\frac{1}{\alpha}}L_i(t)\left(\sum_{k=1}^{M}\tau_{ik}^{-\frac{1-\alpha}{\alpha}}N_k\right).
$$

The demand for intermediate input factors produced in region i from all regions

[14]See Heer and Maußner (2005).

equals

$$\gamma^{-\frac{1-\alpha}{\alpha}}\frac{\alpha\gamma}{\gamma+\alpha-1}(1-\alpha)^{\frac{1}{\alpha}}(r^g)^{-\frac{1}{\alpha}}N_i\left(\sum_{k=1}^{M}\tau_{ik}^{-\frac{1-\alpha}{\alpha}}L_k\right).$$

Which can be obtained by calculating the marginal product of one variant produced in region i and integrating the implied demand function over all variants produced in region i and using (4.12). Clearly the two figures are identical if (4.10), i.e. the zero profit condition, is satisfied. That means that the capital stock used in final good production, i.e. intermediate input factors coming from all regions in the economy, is always identical to the capital stock used in production of intermediate input factors, i.e. intermediate input factors going to all regions in the economy.

The behavior of capital imports by a particular region i can be deducted from the budget constraint and consumption behavior. Denote aggregate consumption as $C_{i,t} = L_i c_{i,t}$, assets at the begin of period t as $A_{i,t}$ and capital imports as $K^I_{i,t}$, then

$$A_{i,t+1} = A_{i,t} + Y_{i,t} - C_{i,t} - (r_t+\delta)K^I_{i,t} - \delta(K_{i,t} - K^I_{i,t}). \tag{4.24}$$

The aggregate consumption function can obtained from the optimality condition (4.2) and the intertemporal budget constraint

$$\sum_{t=0}^{\infty}\frac{C_{i,t}}{\prod_{s=0}^{t}(1+r_s)} = A_{i,0} + \sum_{t=0}^{\infty}\frac{L_i w_{i,t}}{\prod_{s=0}^{t}(1+r_s)},$$

where the transversality condition has already been imposed. With this it follows that

$$\frac{C_{i,t}}{1+r_t} = \frac{\rho}{1+\rho}(A_{i,t} + PV_{i,t}), \tag{4.25}$$

where $PV_{i,t}$ is the present value of income at the beginning of period t. Now observing that $A_{i,t} = K_{i,t} - K^I_{i,t}$ and using equations (4.24) and (4.25) aggregated over all regions, it follows that

$$\frac{K^I_{i,t+1}}{K^I_{i,t}} = \frac{1+r_t}{1+\rho}.$$

The inconsistency in the time subscripts compared with the consumption behavior is due the fact that capital stocks are measured at the beginning of the period and consumption at the end of the period. Defining $\tilde{K}_{i,t} = K_{i,t+1}$ as the capital stock measured at the end of period t it is obvious that capital imports always grow with the same rate as consumption.

Chapter 5

Conclusions

This chapter draws some conclusions from the results of the different chapters of this thesis. First the contents are summarized with a focus on the new results. In the second part of this chapter some comments on future research prospects are given, i.e. some extensions and generalizations of the theoretical models and empirical applications that can be found in the preceding chapters.

5.1 Summary

As has been laid out in the introductory chapter, the main motivation for this thesis comes from the importance scale effects play in modern economic theory. Scale effects are the corner stones of the new trade theory, the new economic geography and the new growth theory. As the word "new" in all categories indicates, these theories are state of the art in economic writing.

It is exactly the presence of scale effects, which arise in all of the models used in these theories from the existence of increasing returns to scale at some level of the economy, that drives the results and shapes the thinking of economists. As such, scale effects link these important fields of economics. The literature cited and summarized in the introductory chapter works as the fundamental economic theory used to understand effects in the economy at different levels of aggregation. Therefore these theories can be and are used in many applications to explain more detailed aspects of economics.

That is the point where the contributions of this thesis set in.

The first two chapters take a look at the more disaggregated level of the economy, comprised of two different sectors using heterogenous labor. The distinction is made between high and low skilled workers which both can work for a “high-tech” or a “low-tech” sector in the economy. Technology, which determines productivity and therefore to a large part also wages, is endogenous and is driven by market forces. The economic environment can be summarized by two decisions: First, each firm in the market has to decide how much technology it devotes to high skilled and how much to low skilled workers. Second, providers of technology have to decide whether they develop new technologies for the “high-tech” or the “low-tech” sector. In equilibrium all actions taken by firms in the market have to yield equal returns. This leads to certain directions of technological change in order to reach the general equilibrium.

The important result is that changes in the skill structure of the economy, i.e. the relative supply of high skilled workers, biases technological change, i.e. skill and sector specific technological change in favor of the factor that is getting more abundant. This in turn affects productivity positively which, under certain conditions can lead wages to rise. The new contribution of the model in chapter two is that both sectors can employ high and low skilled workers, although with different intensities. This opens the possibility to distinguish between the elasticity of substitution between sectors and skill groups. The result is, that the elasticity of substitution between sectors determines the result, a conclusion which could not be drawn from earlier contributions to the literature. This elasticity has to be large in order to lead relative wages of high skilled to rise in response to a rise in their relative supply. This phenomenon is observed until the mid 1990ies in economies as the US or the UK which a generally accepted to have a certain affinity to a free labor market. Still the elasticity of substitution must be very high for this result. But if one thinks in longer periods of time this elasticity might be high in the long-run. It is easy to find examples in economic history where jobs or products which have been produced by low skilled workers have been very good replaced by new services and goods now associated with "high-tech". The results also depend on the existence of so called "strong" scale effects in the growth mechanism governing technological innovations. It is present in the model because a first generation growth model has been used to explain the innovation process in the economy. A market that gets larger, i.e. through an increase of the relevant labor force, experiences temporarily higher growth rates until the new equilibrium is reached. As an additional result the model presents an explanation why high skilled employment in the "high-tech" sector is growing faster than in the "low-tech" sector if high skilled employment overall increases. This result has been shown to be consistent with empirical observations for the US manufacturing industries.

The model of chapter two has been extended to cover some open economy considerations. It has been shown that the results stay intact in the open economy and, furthermore, that wage inequality can spread from one economy to another, if both are linked through trade.

However, the model suffers from the dependence of the result on "strong" scale effects, as just mentioned. It has been demonstrated by the literature, that such scale effects are very unlikely to be consistent with empirical observations.

This serves as the motivation for the third chapter which focuses on the role of "weak" scale effects in explaining the observation of the coincidence of a rising relative supply of high skilled workers and their relative wage in some economies. The "weak" scale effect is an effect that is in common to all second generation growth models and leads in equilibrium to the result that a larger economy, on the very aggregated level, should have a higher per capita production than a smaller. Thus labor productivity, measured by the ratio of production to the extent of the labor force, should increase with the size of the economy. This effect can serve as a substitute for the "strong" scale effect used in the model of chapter two if the idea of second generation growth models is transferred to the more sector case. If relative supply of the high skilled increases, in equilibrium more technologies are available for them to use and therefore their productivity increases. This can lead to a higher relative wage, provided that the elasticity of substitution between high and low skilled sector products is large enough. The arguments of the preceding paragraph apply here as well.

The analysis of the third chapter has been undertaken using several theoretical modelling strategies regarding the production technology and the use of high and low skilled products and labor in the economy. This is true for the general production technique as well as for the innovation sector of the economy. By doing so, it has been shown that the possibility of, at the same time, rising relative supply and relative wages is always present as long as the stability conditions of the models are met.

A quite popular view of the increased unemployment problem in some continental European countries is the Krugman hypothesis. In brief, it states that unemployment is the result of suppressed wage inequality between high and low skilled workers. The model has been extended by an open economy set-up with on large country, allowing for wage inequality, e.g. the US, and one small economy that suppresses this, e.g. Germany. This is an attempt to confront the Krugman hypothesis with an endogenous technology environment which has not been done in the literature so far. The result is, that for the Krugman hypothesis to stay intact, strong assumptions about price setting in the two economies for the same goods have to be made. As such this is to be interpreted as critical for this hypothesis.

The chapter also deals with exogenous technology shocks to the two different sectors and their implications for wage inequality. This is done by introducing state dependence in the R&D process, i.e. technological possibilities for both sectors depend on the state of technology in both sectors. The result is that as long both sectors can learn from each other, exogenous technology shocks have only a temporarily impact on wage inequality, although adjustment to the originally equilibrium for the relative wage might take a long time

To emphasize the role of "weak" scale effects in this model, two models have been constructed showing no scale effect at all, i.e. neither a "strong" nor a "weak" scale effect. The first model is nested within the general models developed in the chapter, but rests on a razor's edge condition for one particular parameter. The second is non-nested but is somewhat more general. In both models, changes in the relative supply of high skilled workers can not cause wage inequality. Both models possess constant returns to scale. The second, however, only in the long-run after technology has adjusted to labor supply changes. This is in contrast with the models used in the first sections of the chapter which all have increasing returns to scale in the long-run after technology has adjusted to labor supply levels.

To summarize up to this point, scale effects play an important role in explaining the coincidence of rising supply and rising relative wages for the high skilled in some economies. While "strong" scale effects have been rejected by the literature, "weak" scale effects are not rejected. But it must be noted that empirical evidence in favor of the existence of these scale effects is also thin. This is mainly due to the fact that second generation growth models, creating the "weak" scale effects are relatively new to the literature and little attention has been paid to conduct research directly on these scale effects. This is exactly the motivation for the fourth chapter which searches for "weak" scale effects on different levels of geographic aggregation.

Chapter four start with a theoretical model bringing several ideas from different economic areas together. A second generation growth model with many economic regions is built which accounts for free capital movements between regions and trade which is subject to friction. Therefore there are elements from the new trade theory, the new economic geography and the new growth theory. As the growth section of this model is of the second generation type, the model possesses "weak" scale effects. But now the "weak" scale effect is governed by open economic regions. As such not longer the scale of one particular economic region is relevant, but the size of all regions linked through trade. Heterogenous trade frictions between regions gives heterogenous conglomerates of economic sizes of all regions for each region, i.e. an inter-regional scale variable, unique for each economic region but with common roots.

This theoretical result gives, first, an explanation for the fact that both small and large economies can have similar labor productivities and, second, serves as a starting point for the empirical analysis.

To test the theoretical result empirically country and regional data are used. A cross section of countries in the year 2000 is used to test the impact of a scale variable on GDP per capita in these countries. The scale variable is defined over the population of the G7 countries as a weighted sum. As weights, the distances to the G7 countries are used. This is done because of two observations. First, the "weak" scale effect operates via technology. The literature on technology diffusion has shown that the major source of world wide available technologies are to the most extent the G7 countries. The inter-country "weak" scale effect is caused through trade linkages. Geographic distance is a commonly used instrument for trade intensity, the inverse distance is often used in spatial econometrics.

On the regional level data on NUTS2 regions for the 15 "old" EU countries and 3075 counties of the US main land are used. An inter-regional scale variable analogous to the approach in the country level analysis has been computed. But now the scale variable is defined over all regions in the data sample. Labor productivity is measured for the EU 15 by GDP per regional inhabitant in working age and for the US counties by personal income per capita. The last measure serves as a proxy for per capita production since no other data are available.

A cross sectional regression, taking account of spatial effects, has been undertaken using the inter-country/regional scale variable as an explanatory variable for labor productivity. It turned out that, on all aggregation levels, this variable is a statistically significant determinant for labor productivity. The elasticity is about 0.25 to 0.29 on the country level and 0.45 to 0.48 on the regional level. The reason for the cross country elasticity to be lower might be due to the fact that many not fully developed countries are in the data sample, for which the world technology frontier is less relevant. To sum up the empirical analysis, there is strong evidence for the "weak" scale effect on the mentioned level of aggregation. This can be interpreted as a justification for the use of these effects in economic models.

The following section will deal with some possible extensions of the chapters of this thesis leading to possibly new results.

5.2 Prospects for Future Research

One interesting application of the theoretical models in chapters two and three would be a synthesis. The set-up of two sectors, one "high-tech" and one "low-tech", which both can use high and low skilled labor with different intensities, could be combined with a second generation model. In such a model it could be tested whether the result from chapter two, that the "high-tech" sector gains more in high skilled employment than the "low-tech" sector if overall high skilled employment increases, is robust with respect to the growth mechanism.

Also the results from the models in chapters two and three should be tested empirically in more detail. It would be interesting to see whether there is a statistically significant relationship between relative supply of high skilled and the relative wage as predicted by the models. Furthermore this analysis should take as many control variables as possible into account to make sure that the true relationship is measured without possible biases. One more problem that one has to deal with, if undertaking such an analysis, is the endogeneity of the skill structure at different levels of aggregation. On the economy level, the skill structure response to the wage structure through education decisions of the population. On the sector and firm level, skilled employment is determined by profitability of the sectors and firms. Successful firms and sectors with high labor productivities attract high skilled workers. Before one can empirically estimate relationships between relative wages and relative supply, if suitable instruments have been found.

As has been demonstrated, both theoretically and empirically, the open economy plays an important role. Any empirical analysis has to control for open economy effects and has to correctly compute the relevant scale which might influenced by the global economy.

As concerns the empirical analysis in chapter four, there are many aspects which could be included. A prominent one is approach recently taken by the literature on technology diffusion (see the literature cited in chapter four). This literature tries to explain productivity in several definitions by international available stocks of knowledge. The later have been approximated by weighted sums of R&D expenditure in different countries. There are many recent approaches which could be adopted in the computation of an inter-regional scale variable mentioned in the preceding section. There might be other channels through which scale effects "diffuse" than pure geographic distance. Possible candidates are direct trade channels, technological or social similarities and linkages due to multinational enterprises. Also progress has been made regarding the computation of weight function by parametrizing these. On attempt has been made in the country analysis in chapter four and the difficulties by doing so have been discussed.

It might also be possible to link both approaches, diffusion of technology and scale effects. While in the long-run the scale measured by the population or work force is the determining factor, knowledge might by influenced by other factors in the short run. This leads one to immediately think about cointegration techniques where the long-run relationship is modelled by using scale variables and the short-run dynamics are governed by R&D related figures.

After all extensions, there might be new applications where scale effects play an important role. One example is the depletion of nonrenewable resources. The pure existence of nonrenewable resources which are essential for production implies an ever decreasing usage of this resource. This means the market for technologies directed to this resource is, ceteris paribus, shrinking. If scale effects matter, this might be a further obstacle for long-run sustainable growth. Furthermore the population might rise over time resulting in increased needs for the resource. On the other hand the "weak" scale effect also implies positive effects of a growing population, counteracting the just mentioned arguments. A rigorous theoretical treatment of this economic question is needed in order to judge under what conditions long-run non-negative growth is feasible. A possible modelling strategy would be a two sector model with one labor and one resource using sector with endogenous technology governed by a second generation growth model.

References

Acemoglu, D. (1998): Why Do New Technologies Complement Skills? Quarterly Journal of Economics, 111, 779-804.

Acemoglu, D. (1999a): Patterns of Skill Premia. NBER Working Paper No. 7018.

Acemoglu, D. (1999b): Labor- And Capital-Augmenting Technical Change. NBER Working Paper No. 7544.

Acemoglu, D. (2001): Directed Technical Change. NBER Working Paper No. 8287.

Acemoglu, D. (2002): Technical Change, Inequality, and the Labour Market. Journal of Economic Literature, 40, 7-72.

Acemoglu, D. and Zilibotti, F. (2001): Productivity Differences. Quarterly Journal of Economics, 116, 563-606.

Aghion, P. and P. Howitt (1992): A Model of Growth through Creative Destruction. Econometrica, 60, 323-351.

Aghion, P. and P. Howitt (1998): Endogenous Growth Theory, MIT Press, Cambridge.

Alcala, F. and A. Ciccone (2004): Trade and Productivity. Quarterly Journal of Economics, 119, 613-646.

Anselin, L. (1988): Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

Autor, D and L. Katz (2000): Changes in the Wage Structure and Earnings Inequality. In Handbool of Labor Economics Vol. III. O. Ashenfelter and D. Card, eds. Amsterdam, Elsevier.

Autor, D, A. L. Katz and A. Krueger (1998): Computing Inequality: Have Computers changed the Labor Market? Quarterly Journal of Economics, 113, 1169-1213.

Backus, D. K, P. J. Kehoe and T. J. Kehoe (1992): In Search for Scale Effects in Trade and Growth. Journal of Economic Theory, 58, 377-409.

Bhagwati, J. N., A. Panagariya and T. N. Srinivasan (1998): Lectures on International Trade. Second edition, MIT Press

Beaudry, P. and D. Green (2000): The Changing Structure of Wages in the US and Germany: What explains the Difference? NBER Working Paper No. 7697.

Benassy, J.-P. (1998): Is there too little research in endogenous growth with expanding product variety? European Economic Review, 42, 61-69.

Berman, E., J. Bound and S. Machin (1997): Implications of Skill-Biased Technological Change: International Evidence. NBER Working Paper 6166.

Card, D. and J. E. DiNardo (2002): Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. NBER Working Paper No. 8769.

Caselli, F. (2005): Accounting for Cross-Country Income Differences. In: Handbook of Economic Growth Vol. 1, North-Holland.

Ciccone, A. (2002): Agglomeration Effects in Europe. European Economic Review, 46, 213-227.

Ciccone, A. und R. E. Hall (1996): Productivity and the Density of Economic Activity. American Economic Review, 86, 54-70.

Coe, D. T. und E. Helpman (1995): International R&D Spillovers. European Economic Review, 39, 859-887.

Conley, T. G. (1999): GMM Estimation with Cross Sectional Dependence. Journal of Econometrics, 92, 1-45.

DiNardo, J., N. Fortin and T. Lemieux (1995): Labor Market Institutions and the Distribution of Wages , 1973-1992: A Semiparametric Approach. Econometrica, 64, 1001-1044.

Dinopoulos, E. and P. Thompson (1998): Schumpetrian Growth without Scale Effects. Journal of Economic Growth, 3, 313-335

Dixit, A. and J. E. Stiglitz (1977): Monopolistic Competition and Optimal Product Diversity. American Economic Review, 67, 297-308.

Edwards, S. (1993): Openness, Trade Liberalization, and Growth in Developing Countries. Journal of Economic Literature, 31, 1358-1393.

Engel, C. M. (1996): A Model of Foreign Exchange Rate Indetermination. NBER Working Paper 5766.

Ethier, W. J. (1982): National and International Returns to Scale in the Modern Theory of International Trade. American Economic Review, 72, 389-405.

Fagerberg, J. and B. Verspagen (1996): Heading for Divergence - Regional Growth in Europe Reconsidered. Journal of Common Market Studies, 34, 431-448.

Feenstra, R. and G. Hanson (2001): Global Production Sharing and Rising Inequality: A survey of Trade and Wages. NBER Working Paper No. 8372.

Feenstra, R. C. and H. L. Kee (2004): Export Variety and Country Productivity. NBER Working Paper No. 10830.

Fingleton, B. (2001): Theoretical Economic Geography and Spatial Econometrics: Dynamic Perspectives. Journal of Economic Geography, 1, 201-225.

Frankel, J. A. and P. Romer (1999): Does Trade cause Growth? American Economic Review, 89, 379-399.

Freeman, R. (1991): How much has De-unionization Contributed to the Rise of Male Earnings Inequality? NBER Working Paper No. 3826.

Futjita, M, P. Krugman und A. J. Venables (1999): The Spatial Economy: Cities, Regions, and International Trade. MIT Press, Cambridge.

Gardiner, B., R. Martin and P. Tyler (2004): Competitiveness, Productivity and Economic Growth across the European Regions. Regional Studies, 38, 1045-1067.

Gianetti, M. (2002): The Effects of Integration on Regional Disparities: Convergence, Divergence, or both? European Economic Review, 46, 539-567.

Grossman, G. M. and E. Helpman (1990): Comparative Advantage and Long-Run Growth. American Economic Review, 80, 796-815.

Grossman, G. M. and E. Helpman (1991a): Innovation and Growth in the Global Economy, MIT Press.

Grossman, G. M. and E. Helpman (1991b): Growth and Welfare in a Small Open Economy. In: Helpman, E. and A. Razin (1991): International Trade and Trade Policy. MIT Press.

Hall, R. E. and C. I. Jones (1999): Why Do Some Countries Produce So Much More Output per Worker than Others? Quarterly Journal of Economics, 114, 83-116.

Haskel, J. E. and M. J. Slaughter (2002): Does the sector bias of skill-biased technical change explain changing skill premia? European Economic Review, 46, 1757-1783.

Heer, B. and A. Maußner (2005): Dynamic General Equilibrium Models. Computational Methods and Applications. Springer, Berlin.

Helpman, E. (1981): International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach. Journal of International Economics, 11, 305-340.

Helpman, E. (1988): Growth, Technological Progress, and Trade. Empirica, 1, 88, 5-25.

Helpman, E. and P. Krugman (1985): Market Structure and Foreign Trade. Increasing Returns, Imperfect Competition, and the International Economy. MIT Press, Cambridge.

Howitt, P. (1999): Steady Endogenous Growth with Population and R&D Inputs growing. Journal of Political Economy, 107, 715-730.

Hummels, D. and V. Lugovskyy (2005): Trade in Ideal Varieties: Theory and Evidence. NBER Working Paper No. 11828.

Jones, C. I. (1995a): R&D-Based Models of Economic Growth. Journal of Political Economy, 103, 759-784.

Jones, C. I. (1995b): Time Series Tests of Endogenous Growth Models. Quarterly Journal of Economics, 110, 495-525.

Jones, C. I. (1999): Growth: With or Without Scale Effects? American Economic Review, Papers and Proceedings, 89, 139-144.

Jones, C. I. (2005): Growth and Ideas. In: Aghion, P. and S. Durlauf (2005): Handbook of Economic Growth. North-Holland

Keller, W. (2001): Knowledge Spillovers at the World's Technology Frontier. CEPR Working Paper No. 2815.

Keller, W. (2002): Geographic Localization and International Technology Diffusion. American Economic Review, 92, 120-142.

Kiley, M. (1999): The Supply of Skilled Labor and Skill-Biased Technological Progress. The Economic Journal, 109, 708-724.

Kortum, S. S. (1997): Research, Patenting, and Technological Change. Econometrica, 65, 1389-1419.

Krugman P. (1979): Increasing Returns, Monopolistic Competition, and International Trade. Journal of International Economics, 9, 469-479.

Krugman, P (1980): Scale Economies, Product Differentiation, and the Pattern of Trade, American Economic Review, 70, 950-959

Krugman, P. (1981): Intraindustry Specialization and the Gains from Trade. Journal of Political Economy, 89, 959-973.

Krugman, P. (1991): Increasing Returns and Economic Geography. Journal of Political Economy, 99, 483-499.

Krugman, P. (1994): Past and Prospective Causes of High Unemployment. Economic Review, S. 23-43.

Krugman, P. (1998): What's new about the New Economic Geography. Oxford Review of Economic Policy, 14, 7-17.

Krugmann, P. (2000): Technology, Trade and Factor Prices. Journal of International Economics, 50, 51-71.

Krugman, P. and A. Venables (1995): Globalization and the Inequality of Nations. Quarterly Journal of Economics, 110, 857-880.

Lancaster, K. (1979): Variety, Equity, and Efficiency. Columbia University Press, New York.

Lee, D. S. (1999): Wage Inequality in the U.S. During the 1980s: Rising Dispersion or Falling Minimum Wage? Quarterly Journal of Economics, 114, 941-1024.

Lee, H. Y., L. A. Ricci and R. Rigobon (2004): Once again, is Openness good for Growth? NBER Working Paper No. 10749.

LeGallo, J. (2004): Space-Time Analysis of GDP Disparities among European Regions: A Markov Chain Approach. International Regional Science Review, 27, 138-163.

Lichtenberg, F. und Bruno van Pottelsberghe de la Potterie (1996): International R&D Spillovers: A Re-Examination. NBER Working Paper No. 5668.

Lucas, R. E. Jr. (1988): On the Mechanics of Development Planning. Journal of Monetary Economics, 22, 3-42.

Mankiw, N. G., D. Romer and D. N. Weil (1992): A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics, 107, 407-437.

Mardia, K. V., J. T. Kent and J. M. Bibby (1979): Multivariate Analysis. Academic Press.

Marshall, A. (1890): Principles of Economics. London, Macmillan Press.

Marshall, A. (1919): Industry and Trade. London, Macmillan Press.

Marx, K. (1867-94): Capital. Harmadworth.

Maurseth, P. B. (2001): Convergence, Geography and Technology. Structural Change and Economic Development, 12, 247-276.

Mill, J. S. (1848): Principles of Political Economy. London, J. W. Parker.

Neven, D. and C. Goyette (1995): Regional Convergence in the European Community. Journal of Common Market Studies, 33, 47-65.

OECD (1996): Employment Outlook. OECD Paris.

Peretto, P. (1998): Technological Change and Population Growth. Journal of Economic Growth, 3, 283-311.

Puga, D. and A. Venables (1997): Preferential Trading Arrangements and Industrial Location. Journal of International Economics, 43, 347-368.

Quah, D. (1996a): Regional Convergence Clusters in Europe. European Economic Review, 40, 951-958.

Quah, D. (1996b): Empirics for Economic Growth and Convergence. European Economic Review, 40, 1353-1375.

Rivera-Batiz, L. A. and P. Romer (1991): Economic Integration and Endogenous Growth. Quarterly Journal of Economics, 106, 531-555.

Rodriguez, F. and D. Rodrik (2000): Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence. In: Bernanke, B. and K. Rogoff (2000): NBER Macroeconomics Annual 2000. MIT Press.

Romer, P. M. (1986): Increasing Returns and Long-Run Growth. Journal of Political Economy, 94, 1002-1037.

Romer, P. M. (1987): Growth Based on Increasing Returns Due to Specialization. American Economic Review, 77, 56-62.

Romer, P. M. (1990): Endogenous Technical Change. Journal of Political Economy, 98, S71-S102.

Romer, P. M. (1991): Increasing Returns and New Developments in the Theory of Growth. In: Barnett, W. A., B. Cornet, C. d'Aspermont, J.J. Gabszewicz and A. Mas-Colell (1991): Equilibrium Theory and Applications: Proceedings of the Sixth International Symposium in Economic Theory and Econometrics. Cambridge University Press.

Rose, A. K. (2005): Size Really Doesn't Matter: In Search of a National Scale Effect. CEPR Discussion Paper No. 5350.

Sala-i-Martin, X. (1996): The Classical Approach to Convergence Analysis. The Economic Journal, 106, 1019-1036.

Samuelson, P. (1954): The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments. The Economic Journal, 64, 264-289.

Schimmelpfennig, A. (2000): Structural Change of the Production Process and Unemployment in Germany. Mohr Siebeck, Tübingen.

Segerstrom, P. (1998): Endogenous Growth without Scale Effects. American Economic Review, 88, 1290-1310.

Smith (1776): An Inquiry into the Nature and the Courses of the Wealth of Nations. Ed. E. Cannan, London, 1961.

Spolaore, E. and R. Wacziarg (2005): Borders and Growth. Journal of Economic Growth, 10, 331-386.

Stiglitz, J. E. (1969): Allocation of Heterogeneous Capital Goods in a two Sector Economy. International Economic Review, 10, 373-390.

Summers, R. and A. Heston (1991): The Penn World Table (Mark 5): An expanded Set of International Comparisons, 1950-1988. Quarterly Journal of Economics, 106, 327-368.

Uzawa, H. (1965): Optimal Technical Change in an Aggregative Model of Economic Growth. International Economic Review, 6, 18-31.

Vassilakis, S. (1987): Increasing Returns to Scale. In: The New Palgrav, a Dictionary in Economics, Macmillan Press, London.

Ventura, J. (2005): A Global View of Economic Growth. In: Aghion, P. and S. Durlauf (2005): Handbook of Economic Growth. North-Holland

Xu, B. (2000): Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth. Journal of Development Economics, 62, 477-493.

Xu, B. (2001): Factor bias, sector bias, and the effects of technical progress on relative wages. Journal of International Economics, 54, 526.

Young, A. (1998): Growth without Scale Effects. Journal of Political Economy, 106, 41-63.