European University Institute

DEPARTMENT OF ECONOMICS

# Technology and the Economy: the Two-Way Causality

Jan Witajewski-Baltvilks

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 27 February 2015

European University Institute
**Department of Economics**

# Technology and the Economy: the Two-Way Causality

Jan Witajewski-Baltvilks

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

**Examining Board**

Prof. Árpád Ábrahám, EUI, Supervisor
Prof. Miklós Koren, Central European University
Prof. Ramon Marimon, EUI
Prof. José Vicente Rodríguez Mora, University of Edinburgh

# abstract

The thesis explores the role of technology in some of the most important economic phenomena of the last decades and examines how changes in the state of the economy could influence the nature of technology. In the first chapter I study the relation between supply of skilled labor, firms' choice of optimal technology and wage inequality. Researchers have acknowledged that one of the key causes of the increase in inequality across OECD countries was the introduction of skill-biased production methods, which generated a higher demand for skilled workers. In the chapter, I explore whether the shift to skill-biased production method was a consequence of changing nature of new global technological paradigm (specifically, the arrival of the information technology) or a consequence of firms' choice to exploit the new technological paradigm in a way that favors skilled workers. Such choice could be motivated by a rapid increase in availability of college graduates in 70s and 80s. To study these questions, I first observe that while the source of the latter cause is global, the source of the former rests in labor market conditions at the country level. Hence panel data estimators could be used to disentangle the two effects. I find that endogenous technology choice at the local level can explain 30% of the increase of the college premium in the OECD countries. The second chapter studies how the rate and direction of technological change is influenced by the parameters of consumers' preferences. I demonstrate that the elasticity of substitution between goods in the Dixit-Stiglitz framework can be represented as a simple linear function of a taste heterogeneity measure. I combine this result with Young's model of endogenous growth, which predicts that the speed of technological progress depends positively on the elasticity of substitution between goods. The purpose of the third chapter is to summarize the convergence of Central and Eastern European to Western European economies in the period between 1995 and 2007.

I decomposes growth of relative output into growth of capital, labor input, human capital and TFP. I find the evidence for the massive contribution of TFP convergence in the GDP convergence.

## Acknowledgement

I would like to thank my primary supervisor, Professor Arpad Abraham for his valuable suggestions, long and inspiring conversations and the great atmosphere of work, which he created. I am very grateful to my three advisors: professors Nicola Pavoni, Francesco Caselli and Ramon Marimon for the fruitful discussions and their insightful comments.

I would also like to thank Fabio Canova, Russel Cooper, Avinash Dixit, Andrea Ichino, Peter Hansen, Evi Papa, Alwyn Young, Michal Dziedziniewicz and participants in seminars at EUI and FEEM and conferences at Queen Marry University and Paris School of Economics for all helpful comments.

Finally, I would like to thank my parents, my sister, my god parents and all my friends for their unconditional support during these five years. Above all, I would like to thank my wife for her presence and love.

# Preface

In my thesis I explore the role of technology in some of the most important economic phenomena of the last decades in the OECD countries such as sharp increase in wage inequality and catching up of Central and Eastern European economies with their Western neighbors. I examine also how changes in the state of the economy could influence the nature of technology.

Initially, in economic theory, technology was understood as a set of parameters determining the productivity of factors of production. In the Solowian tradition, the assumption on Cobb-Douglas production function implies that there could be only one parameter governing the productivity of all factors (the Total Factor Productivity, TFP). In the tradition of endogenous growth theory, the productivity parameter was replaced with a variable that represents the stock of knowledge. Finally, several authors studied technological change using the constant elasticity of substitution production function with several productivity parameters (or knowledge stocks), one for each factor of production (Samuelson (1960), Acemoglu (1998, 2000, 2002, 2007), Caselli and Coleman (2006)). Each of the chapters in my thesis refers to one of these traditions.

In the first chapter I examine the relation between supply of skilled labour, firms' choice of optimal technology and wage inequality. Researchers have acknowledged that one of the key causes of the increase in inequality across OECD countries was the introduction of skill-biased production methods, which generated a higher demand for skilled workers. However, is a production method skill-bias ingrained in the skill-biased nature of new global technological paradigm (i.e. information technology). Should we instead trace its genesis to the choice of firms to exploit the new technological paradigm in a way that favours skilled workers. This choice could have been motivated by a rapid increase in availability of college graduates? To study this question, I first observe that while the source of the latter cause is global, the source of the former rests in labour market conditions at the country level. Hence panel data estimators could be used to disentangle the two effects. Through econometric analysis, I find that countries which experienced higher college graduate growth than other countries also witnessed higher growth of college wage premiums a decade later. A number of arguments suggest that endogenous technology choices on a national level is the most plausible explanation for this finding. One implication is that any policy that affects the supply of skilled workforce will have an impact on the skill-bias of equilibrium technology and wage inequality dynamics. At a theoretical level, I set a microfoundation for the model by showing how research in the R&D sector might generate a tradeoff between technologies that are skill-biased and those that are not.

The second chapter studies how the rate and direction of technological change is influenced by the parameters of consumers' preferences. First, the chapter formalizes Edward Chamberlin's idea that monopoly power depends on the heterogeneity of consumers' tastes. This is done by demonstrating that the elasticity of substitution between goods in the Dixit-Stiglitz framework can be represented as a simple linear function of a taste heterogeneity measure. The result can enrich interpretation of a broad range of models using the Dixit-Stiglitz framework: if predictions of these models depend on the elasticity of substitution, they might depend also on the heterogeneity of consumers' tastes. Subsequently, I combine this result with Young's model of endogenous growth, which predicts that the speed of technological progress depends positively on the elasticity of substitution between goods. Thus, combined with the result, the model predicts a negative dependance of growth on the diversity of tastes in population. The reason is that if consumers in the population have heterogeneous valuations of a product, the quality improvement of this product will bring only a relatively small increase in sales. This disincentivises firms to invest in the improvement of quality.

The purpose of the third chapter is to summarize the convergence of Central and Eastern European to Western European economies in the period between 1995 and 2007. I focus on two decompositions of convergence. The first one decomposes growth of relative output into growth of capital, labour input, human capital and TFP. I propose a new simple method that takes advantage of the availability of the data on relative factor prices to separate the effect of increased shares of well-educated workers and the effect of higher productivity of a more abundant educational group. Furthermore, if workers with different education levels are not perfect substitutes, the method allows isolation and quantification of the negative effect of one type of workers becoming very scarce. The second decomposition is the sectoral decomposition that allows distinguishing whether growth comes from growth of productivities of industries or from moving labour towards more productive industries. I find the evidence for the massive contribution of TFP convergence in the GDP convergence. In turn the accumulation of physical capital per unit of labour was surprisingly sluggish and did not keep pace with the rapid productivity improvement. The sectoral decomposition reveals that the primary source of convergence was the growth of within-sector productivity. Interestingly, the allocation of labour into a more productive sector turned out to be very sluggish and had no contribution in convergence.

# Contents

# 1 Chapter I: Can endogenous technology choices explain wage inequality dynamics? Empirical and theoretical evidence

## 1.1 Introduction

The second half of the 20th century has brought a notable increase in skill premium - i.e. the relative pay of well educated to low-educated workers - in almost all developed countries. The change has attracted wide interest among researchers and motivated a number of studies aimed to explore its roots. The explanation that wins a growing popularity among researchers is a skill-biased nature of new information and communication technology[1]. In the argument, production methods are, to a large extent, shaped by the technology platform[2] (also referred to as general purpose technology or GPT, in the text). A technology platform is a technological paradigm, a basis for further secondary innovations and invention of production methods. Examples of the technological platforms that we have witnessed since the outbreak of the industrial revolution are the steam engine, electric dynamo and now the new ICT technology. A technological platform is also characterized by its world-wide presence (at least in the developed world). The belief is that the new ICT technology platform is by nature ideally matched with highly-skilled workers, thus it has dramatically increased their productivity (relative to the low-skilled workers) and hence led to higher relative wages.

However, by focusing on the change in the global technology paradigm, economists overlooked the presence of the LeChatelier principle within one tech-

[1]The other prominent explanations were the institutional changes (weakening of trade unions - e.g. DiNardo, Fortin and Lemieux (1995)) and globalization (shifting of low-skill-intensive production to less developed countries - e.g. Wood (1995) and Leamer (1995)). The institutional change explanation is not convincing because skill premium in the US started to increase before deunionization (Acemoglu (2000)). Although globalization seems to play important role in the rise (Van Reenen (2011)) it cannot fully explain its pattern (Acemoglu (2000)).

[2]The technology platform is sometimes called a general purpose technology, see for instance Aghion (2002).

nological paradigm: since every global technology platform offers a variety of possibilities for the production process, individual production firms have an option to switch to more skill-intensive production methods upon increase in supply of skilled workers. One implication is that the skill-intensity of production processes may change, even in the absence of technology paradigm change. Another, perhaps more interesting, implication is that if the paradigm does change, higher skill-intensity can arise not necessarily because the new paradigm is more skill-biased by nature, but because firms have optimally chosen to exploit it in a skill-biased way - simply by picking only the skill-intensive processes from the menu of new possibilities it has offered.

The difference between the change in the nature of the technology platform (henforth, 'technological change') and choices by individual firms (henceforth 'endogenous technology choice') can be better understood with an illustration: imagine a good that can be produced in two types of plants, one being an automatized factory with robots that need to be programmed. Its operation rests heavily on skilled labour. The number of tasks that can be performed by unskilled workers is limited, perhaps to cleaning and guarding. The 'automatized plant' is therefore skill inensive and unskilled labour saving. The second type of plant can be referred to as a 'hand and eye' plant: robots are replaced by unskilled workers, and the tasks that need to be performed by white collars are reduced to supervision, training or organizing logistics. The hand and eye plant is therefore skill saving.

There are two reasons why firms could wish to switch from a hand and eye plant to an automotized plant: the first reason is that the arrival of an ICT global technological platform leads to a greater improvement in an automatized plant than in the a hand and eye plant. The second reason would be that the number of unskilled workers has reduced, while skills became abundant -

generating an interest in an unskilled labour saving, skill intensive production method. This second reason is perhaps even more obvious than the first. It is also not a novel economic theory, rather a simple application of the LeChatelier principle. However it has not been yet explored as a potential cause for the 20th century skill-premium outburst[3]. How can we decompose the two effects? How much of the skill premium increase was driven by new technology choices at the local level, and how much due to the global skill-biased technological change? Answering these questions is one of the main purposes of this paper.

It is important to distinguish between the endogenous technology choice hypothesis and the hypothesis in Acemoglu's (1998, 2000, 2007, 2014), often called the directed technology change hypothesis. If we were to classify the two to strands of literature, the latter would be assigned to Hick's (1932) and Samuelson's (1965) induced technology change strand[4], while the former is contained in the LeChatelier strand of literature initiated with Samuelson (1960) paper. Acemoglu's papers assume that there is only one global technology (a single production method) developed in a profit-maximizing, world scale R&D firm and whose skill bias may be influenced by the relative number of skilled and unskilled workers. Instead, the endogenous technology choice hypothesis assumes that, if there is a world-scale R&D centre, it develops a technology platform that is only a basis for a development of a range of production methods - some

[3]The endogenous technology choice hypothesis appeared in labour economic literature: Peri (2009) uses the hypothesis to explain why immigration has a negative impact on skill-bias of technology and subsequently a very modest effect on low-skilled labour wages in the United States. Caselli and Coleman (2006), who found a positive correlation between the level of country GDP and skill bias of technology, argue that it can be driven by the fact that less developed countries have a higher share of low-skilled labour and thus firms in these countries choose the non skilll-biased technologies (they propose a formal model to describe this argument - the model is used as a basis for the dynamic model presented in section 1.2). However, to the knowledge of the author, no study has used the hypothesis to explain the dynamics in college wage premium in the second half of 20th century.

[4]The idea that there might be a tradeoff between developing a technology that is augmenting one factor of production and developing technology that is augmenting other factor dates back to Hicks (1932) and Samuelson (1965). Acemoglu's papers are probably the most complete and the most well known continuation of this tradition. An interesting alternative to Acemoglu's setup is the paper by Ruiz-Arraz (2003), that explores the impacts of directed technology changes using a translog production function.

being more and some being less skill-biased. The production methods that are chosen depends on individual final-good producers. As a result, while directed technological change predicts that the supply of skills affects technology at the global level, the endogenous technology choice predicts that the skill-bias of the production methods in any country depends on the choices of the firms that are influenced by the local (country-level) skills supply.

Is the distinction between the changing nature of the technology platform and the shift of technology choices important? In fact, the two concepts are very closely related - both imply the change in the production methods. Yet the differences might turn out to be crucial. Firstly, the shift in technology choices does not have to happen during a major technological change. Thus in the future we might observe rapidly a changing skill premium structure, even if we will not observe any change in general purpose technology. Secondly, and probably more importantly the technology choices of firms appear to be much more predictable and influenced by the policy than the changes in the nature of general purpose technology. The direction of GPT is highly random, depending more on the wild nature of discoveries rather than any government policies. In turn, the technology choice hypothesis implies that the firms' skill bias will respond to any policy that governs the supply of skilled and unskilled labour. For instance, a policy improving early age education or extending the retirement age for less skilled workers can incentivise firms to pay more attention to production methods that favours this kind of workers. Through the effect on technology choices, these policies will have a negative effect on wage inequality. Policy makers may wish to take this effect into account. Finally, even if we assume that the direction of the GPT nature change might be as responsive to policies as firms' technology choices, the control of GPT direction would need to involve coordinated world-scale actions. Instead, the technology choices of firms depend on the local labour market conditions. This has two implications: the first is that each government can influence technology choices and further, wage inequality independently of the other governments. The second implication is that the model that incorporates the endogenous technology choice argument might predict a variety of wage inequality dynamics across countries.

There are four parts in the paper: The first step is to present a model which can frame both, the technology change and the technology choice hypothesis. The model is found in section 1.2. In essence, it is a dynamic extention of the model by Caselli and Coleman (2006).

The second step is to propose an empirical identification strategy to calibrate the theoretical model. The key challenge is determining how to econometrically distinguish the effect of technology choices from the effect of directed technological change. The idea to isolate the two is based on the presumption that the change of the nature of the technology platform must have global consequences (at least if we focus on developed countries). In turn, the technology choice hypothesis involves changes in skill premium as a response to changes in the conditions of local labour market (i.e. the local relative supply of of skilled labour). As a result, the technology choice hypothesis predicts an increase in the skill wage premium above the average international increase, in countries that have also experienced an above average increase in the skilled labour supply. I devise an empirical model that exploits the cross-section and time-series variation in the data to calibrate the theoretical model. The results implies that approximately one third of the skill premium increase across OECD countries can be explained with the endogenous technology choice hypothesis, while the remaining two third can be exaplined by the skill-biased change in the nature of GPT.

The identification strategy, designed originally for calibration purposes, has in fact uncovered an interesting regularity in the data. The regression results imply that countries that did experience increases in skilled-labour supply higher than other countries have also witnessed larger increases in skill premium. The coefficient is statistically significant and predicts a 0.22% increase in skill premium after a 1% increase in relative skill supply. This result is interesting on its own and certainly deserves a further investigation. In section 1.4 I pursue an exploratory econometric analysis. In the empirical section I find that the result is unlikely to be driven by trends in globalization, the fall of trade unions or institutional differences between countries. Finally, I show that reverse causality, if present, should bias the coefficient downward. The endogenous technology choice at a national level is therefore left as the most plausible explanation for the result.

However, there might be other mechanisms that offer similar predictions. I present and formally describe two such mechanisms: the spillover effect (higher density of skilled labour helps each skilled worker to utilize the technology and thus increase her productivity) and the incentive for adoption of ICT effect (more skilled workers implies firms have more incentive to adopt the ICT technology, which is skill biased by nature). I argue that each of these hypothesis is not plausible separately, however, they might compliment well with the endogenous technology choice effect.

Interestingly the effect of lagged skill supply on skill premium starts to vanish at the beginning of the 21st century. I present three possible explanations for this observation: first, as explained in the theoretical section, a delayed response of skill premium to skill supply will take place only if the shock to supply of skills was unanticipated. This could have been true in the 70s, but it becomes less likely in the 90s. The second possible reason is that over time, the labour market between countries became more integrated. Under the mobility of skills, we would expect the skill premium to become more exogenous, less responsive to local demographics and more dependent on global factors. The last reason could be that adjustment of technological choices is easiest to be performed duirng a time of a major technological change - such as a diffusion of information and communication technology in the 80s. Putting this last hypothesis in the context of previous discussion, possibly global technological change is necessary to grease endogenous technological choices.

In the appendix, I return to discussing the theoretical foundations of the endogenous technology choice. The heart of the hypothesis is the presence (at any point in time) of a tradeoff: firms might choose between technologies that assign higher productivity to skilled workers and those that assign higher productivity to unskilled workers. The derivation of this tradeoff is therefore vital for the entire model. In the appendix, I demonstrate how the R&D process in which researchers invent a finite number of production processes might generate the trade-off between two types of technologies.

## 1.2 Endogenous Technology Choice Model.

In this section, I present a simple dynamic model [5] which illustrates how the labour supply might affect the endogenous technological choice and further, the skill premium. The model sets the basis for the empirical framework introduced in the next section.

Consider an economy with one final product. Suppose that a given technology platform offers a menu of production methods for generating this product, each of them utilizing two inputs - skilled and unskilled labour - but each of them characterized by different productivity parameters. In particular, suppose that production methods $i$ in the menu offered by the platform is characterized

[5] The model presented here is a dynamic extention of the model by Caselli and Coleman (2006)

with the following production function:

$$F_i = [(A_{is}L_s)^{\sigma} + (A_{iu}L_u)^{\sigma}]^{\frac{1}{\sigma}} \tag{1}$$

where $L_s$ and $L_u$ stand for skilled and unskilled labour inputs, and $A_{is}$ and $A_{iu}$ are the productivity parameters for the two types of labour associated with production method $i$. Apart from choosing the quantities of labour inputs, the firm can also choose a technology from the menu. The menu of production methods is determined by the current GPT (technology platform) and is described by the set of pairs $(A_{is}, A_{iu})$ that satisfies

$$\frac{1}{\gamma}A_{is}^{\omega} + A_{iu}^{\omega} \leq B \tag{2}$$

Since every production method is fully characterized by the $(A_{is}, A_{iu})$ pair, it can be represented as a point in the $A_s$, $A_u$ space. Further, the menu of technologies offered by the technology platform may be represented by the set of points satisfying (2). Figure 1 gives two examples of such sets differing in the values of $\gamma$.

The key point to be noticed in the figure is that given technology platforms, the firms face a tradeoff between technologies that give highly productive roles to skilled workers and those that assign a highly productive role to unskilled workers. This is indeed the central assumption of the model and the entire technology choice hypothesis. Is it justifieble?

### 1.2.1 Potential sources of the trade-off between productivities

For simplicity of the argument in the above mode,l the trade-off between the two productivity parameters at the frontier is explicitly imposed although there are various models in which it will come up naturally. One way to generate the

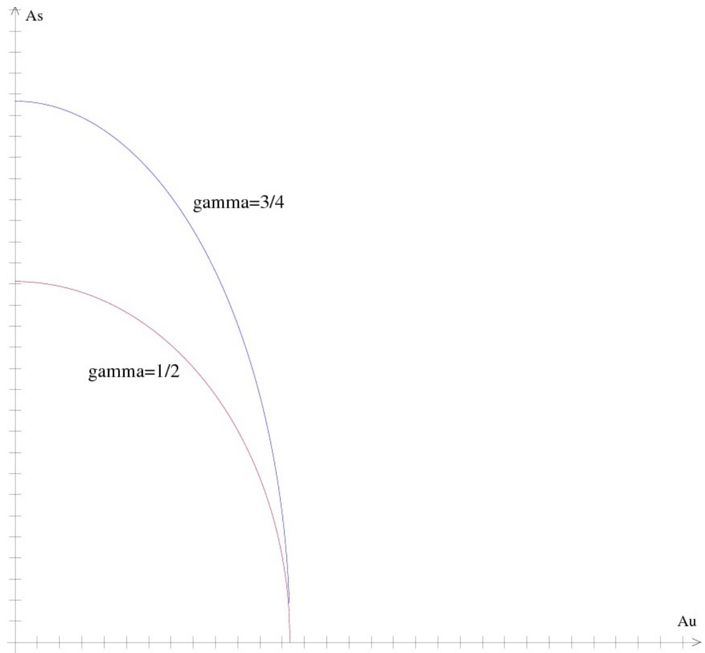



Figure 1: Production methods menu for two different values of gamma.

trade-off is to introduce in the model the costs of the adoption of technologies for the firms (in terms of units of their final output). The more advanced the technology it aims to adopt, the higher is the cost of adoption. Suppose there are two types of machines, each assisting different type of labour. Adopting advancements in the machines that assist skilled workers and in the machines that assist unskilled workers has different cost: $\frac{c}{\gamma}$ and $c$ respectively. The firm optimization problem can then be stated as:

$$\max_{L_s, L_u, A_s, A_u} [(A_s L_s)^\sigma + (A_u L_u)^\sigma]^{\frac{1}{\sigma}} - w_s L_s - w_u L_u - \frac{c}{\gamma} A_s^\omega - c A_u^\omega$$

The firm will therefore face the trade-off - it might spend less on the unskilled dimension of the technology but advance more on the skilled dimension of the technology, or vice versa. The trade-off will be rulled by the relative cost of adoption parameter $\gamma$. The model in this version is elaborated further in the appendix A in the subsection A1.1.

Another model would be one that treats production methods as randomly generated objects. Imagine a Science University that has just devised a new civilizational milestone (such as steam power , semi-conductors, or radioactive decay). The finding has been passed to the Engineering Institute that will try to determine how to combine the new scientific discovery and two types of labour inputs to generate a final good. In fact, they might have various ideas on how to do it, and each idea will involve some degree to which the newly discovered law of nature can compliment the work of skilled and unskilled workers. Thus each idea can be represented with the production function (1) with parameters $(A_{is}, A_{iu})$.

The ideas (the pairs $(A_{is}, A_{iu})$ that engineers could determine) depend partly on chance and partly on the nature of the scientific discovery made in the Science University. Therefore we might think about each idea, or rather a pair $(A_{is}, A_{iu})$ that characterize it, as a draw from the bivariate distribution whose parameters depends on the nature of the discovery (some discoveries might be skill-biased by nature in the sense that the explored law of nature compliments ideally with the effort of educated workers - then engineers have much higher chances of finding out the production methods with very high $A_s$). Engineers might have $n$ ideas and thus $n$ production methods (with $n$ associated $(A_{is}, A_{iu})$ pairs) will appear as possibilities to be picked up by firms around the globe.

Now consider figure 2 (either right or left panel) that illustrates $n$ random draws from the bivariate distribution. For a moment, let's focus on the draw that assigns the highest value to $A_s$ - i.e. the skilled workers productivity parameter. We would expect the probability that this draw happens also to

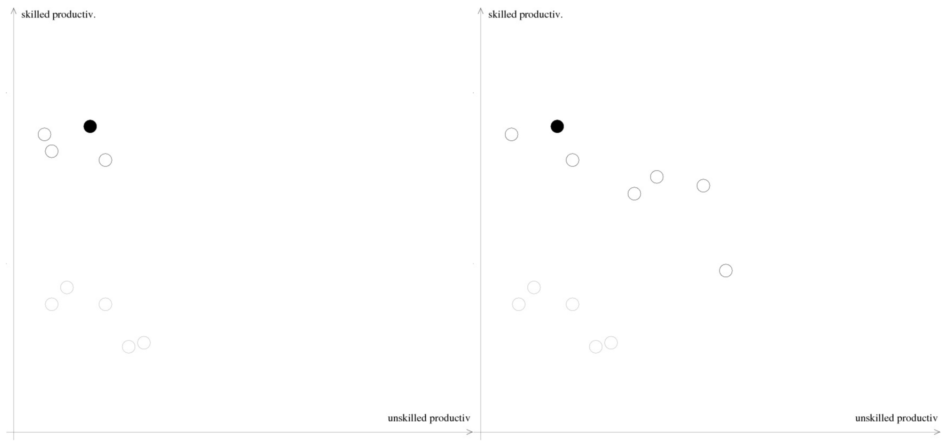



Figure 2: Production methods possibilities in space of the productivity of unskilled workers vs. the productivity of skilled workers. The panel on the left shows the case of the skill-biased technological platform. The panel on the right presents the case of the skill-neutral technological platform.

assign the highest value of $A_u$, the productivity parameter of unskilled workers among all the $n$ draws (i.e. that the quarter east-south to that point is empty) is rather low. The existence of the other point that would assign a higher value of $A_u$ and lower value of $A_s$ then implies a trade-off between the two. Intuitively, although researchers at the Engineering Institute working on utilization of ICT technology have high chances of designing a production method in which the skilled workers play the key role and unskilled workers play very modest role, it is likely that they came up with a different production method in which the role of unskilled workers is more significant (and the role of skilled workers is less significant).

To illustrate this idea with the example, suppose there are two ways of producing output: one is a fully automatized factory where most of the task are performed by skilled labour (e.g. programming robots) while unskilled workers perform less important tasks (such as cleaning, guarding etc.). The other way is the production side on which unskilled workers perform the key production tasks while technology and skilled labour assist them by analyzing mistakes, supervising, organizing logistics and training to maximize unskilled workers performance. Comparing to the former production method the latter production process generates higher demand for unskilled workers.

More generally the model captures the idea that, no matter what is the nature of the current state of science, what were the milestone discoveries, if engineers are able to devise a production method in which the role of unskilled workers and technology is only to assist skilled workers it is very unlikely that they cannot come up with the idea to utilize the milestone discovery in the production method in which the role of skilled workers and technology is limited only to assisting unskilled. The availability of such two production methods implies then a trade-off between productivity of skilled and unskilled workers. Subsection A1.2. in the appendix discusses this idea in more detail and illustrates it with a formal model. In proposition 2 in that subsection I demonstrate that the prediction of that model is the same as the predictions of the model in section 1.2, in which the tradeoff is assumed.

### 1.2.2 Characterization of the equilibrium

The introduction of new GPT (or technology platform) will involve the changes in $\gamma$ and $B$ parameters and thus the change of the menu of available production methods. If the technology platform becomes more skill-biased, it will offer opportunities of production methods that make very good use of a skilled worker. In the framework presented above, this will involve appearance of possibilities to choose production functions with very high productivity parameters for high-skilled workers. We can capture it in the model as an increase in the $\gamma$ parameter. Figure 2 illustrates how the menu of available $(A_{is}, A_{iu})$ pairs changes when the platform becomes more skill-biased (i.e. $\gamma$ rises).

Since dynamics play important role in the empirical analysis, we shall incorporate them in the theoretical model. I assume that firms cannot immediately switch technology in response to changes in labour market conditions. This reflects the fact that firms first have to spot the change in the labour market, then they have to develop a new strategy, replace the technology (perhaps by replacing capital goods) and train workers until the new production method operates at its full potential. Therefore, I assume that firms can choose technology only for the next period - the current technology of the firm was determined one period before.

The firm's value function is then:

$$V\left(A_s, A_u, L_s, L_u\right) =$$

$$\max_{A'_s, A'_u, L_s, L_u} \left\{ \left[ \left(A_s L_s\right)^\sigma + \left(A_u L_u\right)^\sigma \right]^{\frac{1}{\sigma}} - w_u L_u - w_s L_s + \beta E \left[ V\left(A'_s, A'_u, L'_s, L'_u\right) \right] \right\}$$

subject to $\frac{1}{\gamma} A'^{\omega}_{is} + A'^{\omega}_{u} \leq B$ . $x'$ denotes the value of the variable $x$ next period. The first-order conditions for technology choices are

$$\frac{dE\left[V\left(A'_s, A'_u, L'_s, L'_u\right)\right]}{dA'_s} = \lambda \frac{1}{\gamma} \omega A'^{\omega-1}_{s}$$

$$\frac{dE\left[V\left(A'_s, A'_u, L'_s, L'_u\right)\right]}{dA'_u} = \lambda \omega A'^{\omega-1}_{u}$$

and the envelope conditions are

$$\frac{dV\left(A_s, A_u, L_s, L_u\right)}{dA_s} = \beta \left[ \left(A_s L_s\right)^\sigma + \left(A_u L_u\right)^\sigma \right]^{\frac{1}{\sigma}-1} \left(A_s L_s\right)^{\sigma-1} L_s$$

$$\frac{dV\left(A_s, A_u, L_s, L_u\right)}{dA_s} = \beta \left[ \left(A_s L_s\right)^\sigma + \left(A_u L_u\right)^\sigma \right]^{\frac{1}{\sigma}-1} \left(A_u L_u\right)^{\sigma-1} L_u$$

Combining all the above conditions:

$$\frac{E\left[\left[\left(A'_sL'_s\right)^{\sigma}+\left(A'_uL'_u\right)^{\sigma}\right]^{\frac{1}{\sigma}-1}\left(A'_sL'_s\right)^{\sigma-1}L'_s\right]}{E\left[\left[\left(A'_sL'_s\right)^{\sigma}+\left(A'_uL'_u\right)^{\sigma}\right]^{\frac{1}{\sigma}-1}\left(A'_uL'_u\right)^{\sigma-1}L'_u\right]}=\frac{A'^{\omega-1}_s}{\gamma A'^{\omega-1}_u}$$

Log-linearizing and applying the approximation[6] $\log\left(E\left[x\right]\right)=E\left[\log\left(x\right)\right]$:

$$\log\left(\frac{A'_s}{A'_u}\right)=\frac{1}{\omega-\sigma}\log\left(\gamma\right)+\frac{\sigma}{\omega-\sigma}E\left[\log\left(\frac{L'_s}{L'_u}\right)\right]$$

This condition already reflects the fact that the higher the (expected) number of skilled workers is in the economy (relative to number of unskilled), the more skilled-biased technology will be chosen by the firm.

If we combine this result with the first-order conditions for labour choices, we find that

$$\log\left(\frac{w_s}{w_u}\right)\bigg|_t=-\left(1-\sigma\right)\log\left(\frac{L_s}{L_u}\right)\bigg|_t+\sigma\log\left(\frac{A_s}{A_u}\right)\bigg|_t=$$

$$-\left(1-\sigma\right)\log\left(\frac{L_s}{L_u}\right)\bigg|_t+\frac{\sigma}{\omega-\sigma}\log\left(\gamma\right)\bigg|_{t-1}+\frac{\sigma^2}{\omega-\sigma}E_{t-1}\left[\log\left(\frac{L_s}{L_u}\right)\bigg|_t\right]$$

The first term is the standard effect associated with diminishing returns to each type of labour: if we increase the number of skilled workers (relative to unskilled), the skilled workers will become (relatively) less productive and earn smaller skill premiums.

[6]This approximation is correct if the variance of the relative supply of skilled labour, $\log\left(\frac{L_s}{L_u}\right)$, is small and firms before time $t$ do not expect any rapid changes in relative skilled labour supply. If we focus on the optimization problem of firms in the early 70s, this is exactly what would be expected: until then the number of skilled workers grew steadily in a constant trend, and the deviations from this trend were marginal. At the beginning of the 70ss firms could have believed the variance of relative skilled labour is very low. Later, it turned out that they were wrong, since the supply growth increased. The unexpected change in the variation is important for the identification, as explained later, in section 1.3.1.

The second effect is associated with exogenous change in the nature of the technology platform: If the technology platform becomes more skill-biased ($\gamma$ increases), the menu of available production methods will now include numerous production processes that involve a high productivity of skilled workers. The firms will respond to this change in opportunities with a shift of opimal production method choices towards the ones that favour skilled workers. This will in turn increase their relative productivity and skill premium.

Finally, the third term captures the key mechanism of the endogenous technology choice hypothesis: a higher (expected) number of skilled workers gives an incentive for firms to choose the production method that fits skilled workers better. As a result, their relative productivity increases and so does the skill premium.

To close the model we should model the supply side of the labour market. For simplicity of the analysis, I assume that the relative supply of skilled workers is exogenous and follows a random walk with drift process:

$$\log\left(\frac{L_s}{L_u}\right)\bigg|_t = \log\left(\frac{L_s}{L_u}\right)\bigg|_{t-1} + \mu + \xi_t$$

where $\xi_t$ is an iid disturbance term. The assumption on the exogeneity of the skills supply is discussed in the empirical section.

We could also model firms' expectations about next period's relative supply of skilled labour. Given that relative supply of skills follows a random walk with drift, firms base their expectations on the current relative supply of skills:

$$E_{t-1}\left[\log\left(\frac{L_s}{L_u}\right)\bigg|_t\right] = \log\left(\frac{L_s}{L_u}\right)\bigg|_{t-1} + \mu$$

Collecting all these conditions, we find that the equilibrium skill wage premium is determined as:

$$\log\left(\frac{w_s}{w_u}\right)\bigg|_t = -(1-\sigma)\log\left(\frac{L_s}{L_u}\right)\bigg|_t + \frac{\sigma}{\omega-\sigma}\log(\gamma)\bigg|_{t-1} +$$

$$+\frac{\sigma^2}{\omega-\sigma}\left(\log\left(\frac{L_s}{L_u}\right)\bigg|_{t-1} + \mu\right) \tag{3}$$

The skill wage premium depends therefore on the exogenous changes in current relative supply of skilled labour, the skill-bias of the global technological platform and last period relative supply of skills, since the latter was used by the firms' last period to form predictions about the current relative supply of skills. The relative supply of skills at time $t$, $\log\left(\frac{L_s}{L_u}\right)\Big|_t$ is going to be correlated with firms' prediction, $\left[\log\left(\frac{L_s}{L_u}\right)\Big|_{t-1} + \mu\right]$. The amount of the correlation depends on the size of the unexpected shock to the relative supply of skills, $\xi_t$.

## 1.3 Calibration of the Model

In this section, I design and estimate the empirical model to determine the extent to which new technology choices (motivated by a skilled labour supply increase) could have contributed to the overall increase in skill premium. In section 1.4 I investigate whether there is evidence for the causal impact of the relative skill supply on national-level technology choices and thus on the skill premium.

The empirical model can be directly derived from equation (3). The equation is restated below:

$$\log\left(\frac{w_s}{w_u}\right)\bigg|_t = -(1-\sigma)\log\left(\frac{L_s}{L_u}\right)\bigg|_t + \frac{\sigma}{\omega-\sigma}\log(\gamma)\bigg|_{t-1} +$$

$$+\frac{\sigma^2}{\omega-\sigma}\left(\log\left(\frac{L_s}{L_u}\right)\bigg|_{t-1} + \mu\right) \tag{4}$$

### 1.3.1 Identification

The callibration of this model involves two identification problems: first, we have to isolate the effect of the actual increase in the relative skills supply (the first term in the equation above will decrease skill premium due to diminishing returns to skilled labour) and the effect of the expected increase in the relative skills supply (the last term in the equation above will increase college wage premium as firms wish to adjust their technology choices based on a higher number of skilled workers). If the expectation is exactly the same as the actual change, the identification of the two effects would not be possible. The firms, however, cannot perfectly forecast the future supply of skills and we can exploit this fact for the identification.

The second identification problem is to isolate changes in the global technological platform from new choices of technologies driven by the increasing number of skilled workers. For this purpose, we are going utilize the fact that the growth in the number of skilled workers varied across countries. Thus we can use a cross-section of the data to isolate the role of technology choices from the role of global technological change. The assumption that is required for identification is that, within OECD, all countries face the same technology platform – i.e. that access to all available production methods is free among all developed countries. Therefore, the parameter $\gamma$ will be considered as global and will be indexed by the time but not by the country index. [7]

The model should take into account that some countries might have traditionally different productivities of skilled and unskilled labour, perhaps due

[7] In fact the assumption might be much less restrictive: countries technology platforms might be characterized (see equation (2)) by different $B$ parameter (thus we allow some countries to have higher overall productivity). Furthermore given that the estimation uses a first difference regression (as described later) at any moment of time the countries might face different $\gamma$ parameter (that captures the skill-bias of technological platform, the menu of available production methods, see equation (2)). In fact the only restriction needed is that the change of $\gamma$ parameter should be uncorrelated with the country growth of relative skills supply.

to the differences in educational systems and the skilled workers' productivity, relative to unskilled workers' productivity in some countries is lower than in others. To account for this fact, I include country fixed effects in the empirical model.

The above observations and assumptions help to form the following empirical model:

$$w_{it} = \alpha_1 l_{it} + \alpha_2 l_{it-5} + \alpha_3 l_{it-10} + d_t + c_i + \varepsilon_{it}$$

where $w_{it} = \log\left(\frac{w_s}{w_u}\right)$, $l_{it} = \log\left(\frac{L_s}{L_u}\right)$ in country $i$ at time $t$ and $d_t$ and $c_i$ are time and country fixed effects. [8].

Because the country fixed effect might be potentially correlated with skills supply (e.g. a more egalitarian education system that decreases the skill premium might also discourage higher education, affecting the skills supply), it might potentially bias the estimates. One way to remove this problem is to look at the above equation through first differences:

$$\triangle w_{it} = \alpha_1 \triangle l_{it} + \alpha_2 \triangle l_{it-5} + \alpha_3 \triangle l_{it-10} + \hat{d}_t + \triangle \varepsilon_{it} \tag{5}$$

Perhaps we can gain an additional clarity if the above equation is rearranged into:

$$(\triangle w_{it} - \triangle w_t) = \alpha_1 (\triangle l_{it} - \triangle l_t) + \alpha_2 (\triangle l_{it-5} - \triangle l_{t-5})$$

$$+\alpha_3 (\triangle l_{it-10} - \triangle l_{t-10}) + \triangle \epsilon_{it}$$

where $b_t$ is the cross-country average of variable $b$ at time $t$ and $\epsilon_{it} = \varepsilon_{it} - \varepsilon_t$.

Therefore, the effect that we actually meassure with the coefficients $\alpha_2$ and

[8] I use both, five and ten years lags since it is difficult to assume apriori how long is the adjustment time for technology

$\alpha_3$ is the impact of the deviation of skill supply growth from the average international growth on the deviation of the growth in college wage premium from its globally-observed growth. Putting it differently, we can examine if countries that experienced growth in the number of college graduates higher than that of other countries also experienced higher growth of college wage premiums a decade later. If this did occur, there must be some country-level mechanism that generates this dependence. In this section, I will attribute this dependence to endogenous technology choices. The estimation of the parameters will therefore serve to determine a possibility result on how large the role of the adjustment of technology choices in shaping the dynamics of wage inequality was.

In section 1.4, I will discuss other potential local-level mechanisms that could explain the dependence of skilled premium growth on past skill supply growth. However, it turns out that it is difficult to explain the result with any explanation other than the endogenous technology choice hypothesis.

### 1.3.2 Data

The source of the data is the EU KLEMS dataset, 2008 release, covering annual data between 1970 and 2005 for 23 countries (although the panel is not balanced). The data contains information on total hours worked and total compensation for three groups of employees: highly-skilled (those with at least a tertiary education), medium-skilled (those with a secondary education) and low-skilled (those with at most a primary education).

To simplify the analysis and maintain clarity I merge the low-skilled and medium skilled worker groups into one group of “unskilled” workers. Because the hours of medium-skilled work might be worth more than the hours of low-skilled work in the computation of unskilled labour supply, I use the standard approach to weight the medium-skilled workers’ hours by their productivity, relative to the low-skilled workers’ productivity. Hence unskilled labour supply is computed as $L_u = L_l + ({}^{w_m}/_{w_l})L_m$. As a result, the labour supply of unskilled work is measured in terms of low-skilled hours equivalents.

To avoid potential problems with cyclicality, I use only the datapoints in 1970, 1975, 1980, 1985, 1990, 1995 and 2005 and meassure the differences over 5-year periods. Obviously the 5-year difference is effectively an average of the annual differences in a 5-year period.

### 1.3.3 Regression Results

The results from the random effect regression are presented in table 1, column 2. As predicted by the model, the coefficient on the current change in skills supply is negative; this reflects the diminishing returns to skilled labour. The effect is substantial (the 10% increase in relative skill supply is associated with the 8% drop in skill premium), though very close to the estimates obtained by Katz and Murphy (1992) in the similar regression of the skill premium series on the skill supply series using US data. Nevertheless, as mentioned above, this result should not be taken as a causal effect due to the likely reverse causality.

The results also show a significant positive effect of past increases in the relative supply of skilled labour. The effect is also substantial in economic terms: a 10% increase in the relative skill supply involves a 2.2% increase in the skill premium. Interestingly, the time needed for the change in relative labour supply to be reflected in the change in skill premium is rather long: the coefficient is positive and significant only for the relative skill supply lagged by 10 years. The coefficient on the five year lag is not significantly different from zero and in fact negative.

An important remark on the regression result is a warning that the positive coefficient on the lagged skill supply *does not* imply that the increased supply of skills brings in the long run an increase in skill premium. If we take the Katz and Murphy estimates of the slope of demand (reconfirmed later by the more

| | (1) | (2) | (3) |
|---|---|---|---|
| skills supply growth {t} | -0.804*** | -0.825*** | -0.803*** |
| | (0.125) | (0.133) | (0.126) |
| skills supply growth {t-5} | -0.255* | -0.224 | -0.253* |
| | (0.144) | (0.154) | (0.145) |
| skills supply growth {t-10} | 0.218** | 0.217** | 0.225** |
| | (0.09) | (0.097) | (0.092) |
| d85 | | -0.006 | |
| | | (0.05) | |
| d90 | | -0.03 | |
| | | (0.048) | |
| d95 | | 0.013 | |
| | | (0.038) | |
| d00 | | 0.013 | |
| | | (0.037) | |
| year | | | 0.001 |
| | | | (0.002) |
| constant | 0.167*** | 0.163*** | -1.466 |
| | (0.026) | (0.033) | (4.181) |
| Rsquare: | 0.5803 | 0.5887 | 0.5816 |

Table 1: The dependent variable is five years change of college wage premium (in logs). The independent variables are the 5 year change in the ratio of college graduates to remaining part of labour force (in logs), its 5 and 10 years lags and dummy variables for each year (or linear trend). All estimations comes from Random Effect regressions.

careful instrumental variable estimates in Ciccone and Peri (2005) and by the results of regression in Table 1), an increase in relative skill supply leads first to an approximately 7% drop in skill premium. If we use estimates from the regression results, we will expect the skill premium to rebound in a decade and increase by approximately 2%. This means that the initial level of skill premium will not be restored, and the long run effect of skill supply on skill premium will be a 5% drop.

At this stage we can calibrate the model from section 1.2 and calculate the contribution of an endogenous technology choice in the total increase in skill premium. Since a number of countries do not have observations before 1980, I will consider the period of 1990-2005. Over these 15 years, the relative skill supply in 12 OECD countries (that have data for entire period) increased by 32%. Using Katz and Murphy estimates of demand curve, that should translate into an 18% drop in skill premium. Instead, the skill premium in this period increased by 28%. This leaves a 56% of unexplained wage increase (0.193 log points). In the period of 1980-1995, the relative skills supply increased by 90%. This, according to the model and the estimates above, should lead to a 15% (0.062 log points) increase in skill premium due to the endogenous technology choice between 1990 and 2005. The residual increase left is then 35% (0.131 log points), which can be probably attributed to skill-biased technology changes. This leads to the conclusion that endogenous technology choice can explain 32% (0.062 out of 0.193 log points) of the increase in skill premium that could not be explained in the standard demand-supply (Katz and Murphy model) framework.

## 1.4 Empirical Investigation

The empirical results presented above indicate that skill premium dynamics are not explained solely by global factors (such as the nature of general purpose technology), but are also shaped by mechanisms operating in local labour markets. For some readers, the observation that countries which experienced a higher growth of skill premium have also witnessed a higher growth of skill supply a decade earlier could be the most interesting result of this paper. In section 1.2, I have shown one possible explanation for this correlation: changes in the labour market might change the optimal production method choice from the set of available technologies. In subsections 1.4.1 -1.4.3, I present several alternative explanations, and I discuss whether they could be supported by the regression results. In subsection 1.4.4, I investigate whether the effect is stable over time. Interestingly, the correlation between the growth of skill premium and the lagged growth of skills supply weakens over time. I discuss several explanations for this finding.

### 1.4.1 Endogeneity

Operating in the context of labour market equilibrium, it is important to keep in mind that the relative productivity and relative wage will impact the relative skills supply. Although this should bias the estimated effect of the current skills supply on the current skill premium, it is not obvious why it would drive a correlation between *past* skill supply and *current* skill premium. It is not likely that workers can predict that the growth of college wage premiums in their country will be higher than in other countries in a decade. Second, even if they could predict this, it is not obvious why the *change* in relative skill supply should depend on the growth of the college wage premium a decade later. If workers who consider attending college in 1990 predict that in their countries the college premiums will grow substantially, they may be more keen to attend college. However, the same is true for workers in 1985 and earlier.

However, the endogeneity could bias the results if the error terms in the regression are serially correlated. Suppose that one country experienced an exogenous positive shock that increased the skill premium both in the 70s and 80s. Since the change in skill supply may depend on the current changes in skill premium, a country will witness an increase in skills in the 70s. Hence the same exogenous shock could lead to skills accumulation in the 70s and skill-premium growth in the 80s.

To illustrate this logic more formally, assume that the innovation for the *level* of skill wage premium (in deviation from the international level) follows the IMA(1,1) process:

$$w_{it} - w_t = X\alpha + \epsilon_{it}$$

where $X$ is the set of controls in the regresssion and

$$\epsilon_{it} = \eta_{it} + \beta\eta_{it-10} + \epsilon_{it-10}$$

It follows that the change in the innovation can be expressed as:

$$\Delta\epsilon_{it} = \eta_{it} + \beta\eta_{it-10} \tag{6}$$

The source of the upward bias of the results might be the MA component in the innovation in the skill premium. The MA components are associated with the factors that impact the increase (or decrease) of the skill premium (relative to an international increase) over the 5 year period (the time period of the observation). This might be due to a change in the labour or tax policies or a change in the education system that leaves a permament impact on the skill premium. Most of these factors are unlikely to lead to a further increase in skill premium a decade later (it is difficult to imagine a tax policy reform that would lead to an increase of the skill premium in the 80s and another increase in the 90s). The exceptions might be globalization and a decreasing importance of trade unions (we could imagine globalization or trade unions collapse will lead to an increase in the college wage premium over 20-30 years).

The differences in exposure to globalization across countries could indeed be a factor that explains the result: countries that quickly becomes exposed to globalization might experience a quicker growth of demand for educated workers. This will encourage more workers to become skilled. If in the next decade the same country continues to become exposed to globalization more than other countries, the demand for educated workers might shift further, increasing the skill premium. This would generate a spurious correlation between an increase of the college premium today and growth of the college workforce ten years ago. To control for the change in exposure to globalization, I include in the regression the change of the ratio of export to GDP. The results are reported in the second column (regression number (4)) in Table 2. The positive coefficient on the past growth of the skill supply remains significant. Furthermore, it appears that once the past growths of the skill supply are controlled for, the increase in export to GDP ratio has no effect on the change in the skill premium. Although not reported in the table, I have also included the level (rather than growth) of the ratio of export to GDP. Again, this does not affect the results.

Another possibility is that the results are driven by the collapse of trade-unions: if the number of unskilled workers drops significantly, this might undermine the bargaining power of trade-unions and lead to an increase in wage inequality. Moreover, this effect is likely to be delayed. Nevertheless, inclusion of the change in trade union density does not change the results, as reported

| | (2) | (4) | (5) | (6) |
|---|---|---|---|---|
| skills supply growth {t} | -0.825*** | -0.822*** | -0.793*** | -0.786*** |
| | (0.133) | (0.138) | (0.132) | (0.137) |
| skills supply growth {t-5} | -0.224 | -0.222 | -0.242 | -0.242 |
| | (0.154) | (0.158) | (0.151) | (0.155) |
| skills supply growth {t-10} | 0.217** | 0.213** | 0.191** | 0.183* |
| | (0.097) | (0.101) | (0.096) | (0.1) |
| d85 | -0.006 | -0.007 | -0.018 | -0.021 |
| | (0.05) | (0.052) | (0.05) | (0.051) |
| d90 | -0.03 | -0.028 | -0.034 | -0.031 |
| | (0.048) | (0.052) | (0.048) | (0.051) |
| d95 | 0.013 | 0.015 | -0.001 | 0.000 |
| | (0.038) | (0.042) | (0.038) | (0.042) |
| d00 | 0.013 | 0.013 | 0.006 | 0.003 |
| | (0.037) | (0.043) | (0.036) | (0.042) |
| supply of skills (level) | | | -0.041* | -0.042* |
| | | | (0.024) | (0.025) |
| change in union density | | -0.034 | | -0.053 |
| | | (0.145) | | (0.142) |
| change in export to gdp ratio | | -0.006 | | 0.004 |
| | | (0.089) | | (0.087) |
| constant | 0.163*** | 0.161*** | 0.274*** | 0.272*** |
| | (0.033) | (0.036) | (0.072) | (0.075) |
| Rsquare: | 0.5887 | 0.5894 | 0.6143 | 0.6156 |

Table 2: The dependent variable is the five-year change of the college wage premium (in logs). The independent variables are the 5-year change in the ratio of college graduates to the remainder of the labour force (in logs), its 5- and 10- year lags, dummy variables for each year, the proportion of skilled labour in total labour (lagged 10 years), 5- year change in ratio of exports to GDP (in logs), and 5- year change in trade union density (in logs). All estimations come from random effect regressions.

in the second column (regression number (4)) of Table 2. It seems that trade unions do not have a significant impact on the college wage premium in OECD countries.

The argument that the results are unlikely to be biased by the autocorrelation of changes in the college wage premium also comes from the the inspection of the data. In almost all countries, the period of raising the relative skill supply comes first, and only after certain time can the beginning of an upward trend in skill premium be seen.

Another possibility is that the result is driven by the differences in the potency to adopt ICT across countries. Specifically, countries which experienced high growth of skilled labour a decade earlier are able to faster adopt new technologies which happened to be skill-biased. A large change in the skill supply results in a higher stock of skilled labour, including engineers and scientists. This might translate into a higher capacity to adopt technologies that were just developed - like the ICT in the 80s. If the new technologies are (by nature) skill-biased, a higher increase in skill premium in these countries will be observed.

To check for this possibility, we can include in the regression a control for the stock of highly-skilled labour (total hours worked) 10 years ago [9]. The column (5) in Table 2 shows that inclusion of this control does not change the results significantly - the effect of a change in the skill supply on the change in the skill premium is still siginificant at the 5% confidence level, and the coefficient has dropped marginally to 0.19. The regression shows also that the stock of skilled labour does not matter for the increase in the skill premium. The coefficient is significant only at a 10% level, and in fact it is negative. Almost exactly the same results are obtained if instead of controlling for the stock of skilled labour

[9] the lag seems to be necessary since we need to allow a time before the decision to adopt a new technology and a point at which the effect of adoption will be reflected in wage data

lagged 10 years, I control for the stock lagged 5 years and 15 years.

### 1.4.2 Spillover effect

The presence of spillover might be responsible for translating a higher skill supply into a higher skill premium. Suppose that how well a skilled worker uses a technology depends on how many other skilled workers are around. This might be because operating the technology requires a certain degree of experimentation, and sharing experience can facilitate the process and improve the outcome.

An illustrative example of such system of information exchange (although in the context of depeloped countries) is presented in the work of Bandiera and Rasul (2006). They studied the adoption of new crop varieties (a newly introduced technology) among farmers in Northern Mozambique and found that the output of the farmer depends crucially on the interaction with other farmers.

The effect does not have to be immediate - indeed it could be expected to take time before new workers establish connections with the old ones, before they trust each other, find a common language and learn how to utilize each other's experiences. Therefore, I would predict that the increase in the skilled workers supply first drives down their productivity due to diminishing returns to skills, but after a while the additional workers might contribute in knowledge sharing and increase the productivity of every skilled worker. Thus, the spillover effect can explain the pattern observed in the data.

To illustrate this line of thought, one might consider a formal model that includes the spillover effect. The productivity of skilled workers (how well they utilize the technology that is devoted to them) depends positively on the density of skilled workers in the economy, which is approximated with the ratio $\frac{L_s}{L_u}$ [10]. With the amended production function, the profit maximization for the firm $i$

[10] I do not include a similar effect for unskilled workers since spillover in their case is less likely. Nevertheless, inclusion of spillover for unskilled workers would not change the result.

in country $j$ is then:

$$\max_{L_{ijs},L_{iju}} P_{ij}\left[\left(\left(\frac{L_{js}}{L_{ju}}\right)^{\beta}A_{is}L_{ijs}\right)^{\sigma}+\left(A_{iju}L_{iju}\right)^{\sigma}\right]^{\frac{1}{\sigma}}-w_{js}L_{ijs}-w_{ju}L_{iju}$$

The combination of the two first-order conditions then gives:

$$\left(\frac{L_{ijs}}{L_{iju}}\right)^{\sigma-1}\left(\frac{L_{js}}{L_{ju}}\right)^{\sigma\beta}\left(\frac{A_{ijs}}{A_{iju}}\right)^{\sigma}=\frac{w_{js}}{w_{ju}}$$

And denoting $l_{ij}=\log\left(\frac{L_{ijs}}{L_{iju}}\right)$, $l_j=\log\left(\frac{L_{js}}{L_{ju}}\right)$, $a_{ij}=\log\left(\frac{A_{ijs}}{A_{iju}}\right)$ and $w_j=\log\left(\frac{w_{js}}{w_{ju}}\right)$:

$$\left(\sigma-1\right)l_{ij}+\sigma\beta l_j+\sigma a_{ij}=w_j$$

Adding the time indices (that take into account that the effect of spillover is not immediate):

$$\left(\sigma-1\right)l_{it}+\sigma\beta l_{jt-1}+\sigma a_{ijt}=w_t \tag{7}$$

If firms are symmetric in the sense that they face the same $a_{ijt}$ and the same $p_{ijt}$, then the firm indices can be dropped:

$$\left(\sigma-1\right)l_{jt}+\sigma\beta l_{jt-1}+\sigma a_{jt}=w_{jt}$$

Therefore, the spillover model predicts that, following the increase in the relative skill supply, the drop of the skill premium due to diminishing returns to the skill supply is first observed; later, its increase is observed, due to the spillover effect.

The problem with this hypothesis is that it cannot explain the long time lag (10 years) between changes in skill supply and changes in skill premium. Although, as argued before, establishing connections and learning how to share experiences might take some time, it is not plausible that some part of this effect would not be observed in five years. Yet, the data show no positive dependence of the skill premium on relatvie skills lagged five years.

The spillover effect might however play an important role if it is augmented with the endogenous technology choice effect: suppose that the firm knows about the spillover effects and it knows that a higher number of skilled workers implies that new technology directed to skilled workers will be used more efficiently. This creates additional incentive for the firm to shift towards such technology. More formally, this logic can be placed in the model that follows.

The line of the logic is best protrayed in the version of the model in which the firm has to pay (in the units of its final output) for the adoption of technology. Moreover, the adoption of the technology that assists skilled workers and the adoption of the technology that assists unskilled workers have different costs: $\frac{c}{\gamma}$ and $c$ respectively. Then, the profit maximization is given by:

$$\max_{L_{ijs},L_{iju},A_{ijs},A_{iju}} \left[ \left( \left( \frac{L_{js}}{L_{ju}} \right)^{\beta} A_{ijs} L_{ijs} \right)^{\sigma} + \left( A_{iju} L_{iju} \right)^{\sigma} \right]^{\frac{1}{\sigma}}$$
$$-w_{js}L_{ijs} - w_{ju}L_{iju} - \frac{c}{\gamma} A_{ijs}^{\omega} - cA_{iju}^{\omega}$$

Combining two first-order conditions with respect to $A_{ijs}$ and $A_{iju}$, we obtain:

$$\left( \frac{A_{ijs}}{A_{iju}} \right)^{\sigma-1} \left( \frac{L_{ijs}}{L_{iju}} \right)^{\sigma} \left( \frac{L_{js}}{L_{ju}} \right)^{\sigma\beta} = \frac{1}{\gamma} \left( \frac{A_{ijs}}{A_{iju}} \right)^{\omega-1}$$

Again changing the notation, as above, and adding the time indices[11]:

[11] The choice of production method is based on the information from two periods back.

$$(\sigma - 1)\, a_{ijt} + \sigma l_{ijt-2} + \sigma\beta l_{jt-2} = -ln\,(\gamma_{t-2}) + (\omega - 1)\, a_{ijt}$$

Rearranging:

$$a_{ijt} = \frac{ln\,(\gamma_{t-2}) + \sigma l_{ijt-2} + \sigma\beta l_{jt-2}}{(\omega - \sigma)} \tag{8}$$

and referring back to the first condition (7):

$$(\sigma - 1)\, l_{ijt} + \sigma\beta l_{jt-1} + \sigma \frac{ln\,(\gamma_{t-2}) + \sigma l_{ijt-2} + \sigma\beta l_{jt-2}}{(\omega - \sigma)} = w_{jt}$$

$$(\sigma - 1)\, l_{ijt} + \sigma\beta l_{jt-1} + \frac{\sigma}{\omega - \sigma} ln\,(\gamma_{t-2}) + \frac{\sigma^2}{\omega - \sigma} l_{ijt-2} + \frac{\beta\sigma^2}{\omega - \sigma} l_{jt-2} = w_{jt}$$

Dropping the firm indices in the equilibrium (however, not merging the terms in order to ease interpretation):

$$(\sigma - 1)\, l_{jt} + \sigma\beta l_{jt-1} + \frac{\sigma}{\omega - \sigma} ln\,(\gamma_{t-2}) + \frac{\sigma^2}{\omega - \sigma} l_{jt-2} + \frac{\beta\sigma^2}{\omega - \sigma} l_{jt-2} = w_{jt}$$

As in the previous model (without the endogenous choice of technology), there is a direct effect of spillover on skill premium (represented in the second term) taking place after first period (that may be 5 years). However, if $\beta$ is small, this effect may be very limited and even improperly reflected in the data. In addition to the direct effect, the spillover also has an indirect effect: since the returns to investing in the skill-biased technology depends positively on how well this technology is utilized and this depends in turn on the density of skilled workers, the increase in the relative skill supply will incentivise firms to invest more in skill-biased technology, thus providing an additional factor that shifts the skill premium. However, this effect will be introduced only in the second period since two periods are required before the firm spots the increase in the relative skill supply and then implements the new technology. Moreover, the indirect effect of spillover (through incentivising a skill-biased technology choice) might be stronger than the direct effect. If $\omega$ is not large and $\sigma$ is not too far from unity, the effect of spillover might appear only after second period.

### 1.4.3 Endogenous adoption of skill-biased technology.

Suppose that the technology choice mechanism does not work, and each technological paradigm offers only one new production method and firms cannot choose between various production methods). Now imagine that new technology paradigm (e.g. information and communication technology) and the (single) production method it offers guarantees higher productivity for both skilled and unskilled workers; however, compared to the previous technology it is clearly skill-biased in the sense that it benefits skilled workers much more than the unskilled. Suppose that the adoption of this technology is very costly (e.g. involves temporary loss of overall productivity) and not all countries want to immediately jump to the new production methods. One would expect that the first countries to adopt the new technology (and consequently move to a higher skill-premium) were the countries that can benefit most, i.e. those with a high stock of skilled labour. This could potentially explain the empirical results: countries with a high growth of skilled labour supply could have accumulated a high stock of skills. Those countries would adopt the new, skill-biased production method more rapidly and thus increase their skill-premium more than other countries.

Furthermore, the story may be continued beyond the first two periods: next period, another new technology (even more skill-biased) may appear and again countries that will adopt it first will be those with the biggest stock of skilled workers.

This hypothesis is in fact testable: it would imply that the skill premium increase should depend positively not only on the growth of skill supply but also on the level of skill supply. Yet, this is disproved by the regression in column

(5) of Table 2: there is no evidence that such positive association exist. The hypothesis would not be also able to explain the experience of Korea, which has reported a decline of relative skill supply and a decline of skill premium.

Nevertheless, elements of these mechanisms can be incorporated into the endogenous technology choice model. It may be that the new technology platform offers a range of production methods that boost skilled workers' productivity and only a few methods that improve unskilled workers' productivity (in the mode,l this is simply captured by the increase in $\gamma$ and modest increase in $B$) - indeed, it is very likely in case of ICT. All countries in such case will shift towards more skill-biased technologies. However, the countries that have large numbers of skilled workers and which before and after the change were operating relatively skill-biased technologies will witness a large jump away from the old technology (since there are so many new opportunities in the corner for skill-biased technologies) and, analogously, will experience a high degree of adoption of new technology. Conversely, countries that had a high number of unskilled workers, and which always positioned themselves at the unskilled corner of technolgy choices, will only move slightly away from their previous positions (although they might move towards more skill-biased technologies, they will still remain closer to the corner of unskill-biased technologies where not many new opportunities had been offered by the new technology platform).

Yet, the model does not predict the effect of the stock of skills on the change in skill premium - this is because countries with high stocks of skilled workers were already using more skill-biased technology; therefore, if the technology platform nature becomes more skill-biased itself, this will not have a higher effect on skill premium in these countries than in any other countries.

### 1.4.4 Stability of the effect

Interestingly, the positive effect of the supply of skills on future skill premiums seems to be diminishing in recent decades. Column (8) in Table 3 reports the results of the regression if the observations for years 2000 and 2005 are excluded. Comparing these results to the original regression (column (2) in Table 3) indicates that the coefficient on lagged growth of skills supply has almost doubled. If in addition observations for year 1995 are excluded, the coefficient increases further. In column (7), I also report the regression (for the full sample) which includes an interaction between lagged growth of skills supply and the time trend. The interaction term is negative and highly significant.

There are three possible explanations for these observations: first, as explained in the theoretical section, a delayed response of the skill premium to the skill supply will take place only if the shock to the supply of skills was unanticipated. This could have been true in the 70s, but it becomes less likely in 90s. The second possible reason is that over time, the labour market between countries became more integrated. Under mobility of skills, skill premium would be expected to become more exogenous, less responsive to local demographics and more dependent on global factors. The last reason could be that the adjustment of technological choices is easiest to be performed during a time of a major technological change - like a diffusion of information and communication technology in the 80s. Putting this last hypothesis in the context of the previous discussion, possibly global technological change is necessary to grease endogenous technological choices.

## 1.5 Concluding Remarks and Policy Implications

The purpose of this paper was to decompose the increase in the college premium growth across OECD countries into the growth caused by global forces (such as

| | full sample (2) | full sample (7) | before 2000 (8) | before 1995 (9) |
|---|---|---|---|---|
| skills supply growth {t} | -0.825*** | -0.679*** | -0.874*** | -0.729* |
| | (0.133) | (.108 | (0.139) | (0.387) |
| skills supply growth {t-5} | -0.224 | -0.313*** | -0.132 | -0.103 |
| | (0.154) | (0.122 | (0.236) | (0.547) |
| skills supply growth {t-10} | 0.217** | 83.598*** | 0.415*** | 0.461** |
| | (0.097) | (15.617) | (0.106) | (0.236) |
| d85 | -0.006 | -0.214*** | -0.05 | -0.016 |
| | (0.05) | (0.056) | (0.035) | (0.071) |
| d90 | -0.03 | -0.134*** | -0.053 | |
| | (0.048) | (0.043) | (0.035) | |
| d95 | 0.013 | -0.054* | | |
| | (0.038) | (0.032) | | |
| d00 | 0.013 | -0.015 | | |
| | (0.037) | (0.029) | | |
| skills growth {t-10} X year | | -0.042*** | | |
| | | (0.008) | | |
| constant | 0.163*** | 0.241*** | 0.126*** | 0.035 |
| | (0.033) | (0.03) | (0.030) | (0.073) |
| Rsquare | 0.5887 | 0.7504 | 0.8458 | 0.7747 |

Table 3: The dependent variable is a five-year change of the college wage premium (in logs). The independent variables are the 5-year change in the ratio of college graduates to remained of the labour force (in logs), its 5- and 10- year lags, dummy variables for each year, and the interaction term for the interaction between the time trend and the 5-year change in the ratio of college graduates to remained of the labour force lagged 10 years (in logs). All estimations come from random effect regressions.

skill-biased changes in the global technological paradigm, the new ICT) and the growth driven by local (national-level) forces related to the supply of the college workforce. I propose an empirical model which uses both the cross-section and the cross-time variation in the data to separate global from local factors. I find that countries with a higher growth of the college workforce experience a substantially higher college wage premium growth a decade later. Given that the independent variable of interest is lagged ten years, it is unlikely that the result is driven by reverse causality. The results of the regression which includes various control variables suggest that the dependence is not caused by a decreasing importance of trade unions, globalization forces, or the endogenous adoption of ICT technology. In this light, the most plausible candidate to explain this result is the endogeneity technology choice: if the number of college workers increases, firms have an incentive to choose the production methods (technologies) that are better suited for skilled workers. This, after some time, drives up the demand for educated workers and thus leads to an increase in the college wage premium. Using the results of the regression, I might conclude that a national-level mechanism driven by the supply of skilled workers (most likely endogenous technology choice at the local level) can explain 30% of the increase of the college premium in OECD countries.

To draw a policy implication, I need to consider all effects predicted by the model. The framework presented in section 1.2 (and supported by empirical evidence from section 1.3) predicts that the increase in the college workforce will first lead to the fall of the college wage premium (due to diminishing returns to skilled labour) and later (approximately in a decade) to its growth (due to the fact that technology choices of firms will be adjusted to a higher supply of skilled workers). The latter growth will be smaller than the initial drop, thus the net effect of the college workers' supply on their relative wages will be negative.

One important lesson policy makers can learn from the model is that a negative effect of the increase of the skill supply on wage inequality is smaller than predicted by the previous studies based solely on the short-term analysis. The second implication is that any unexpected increase in the college workforce will produce a substantial fluctuation of the college wage premium (it will first drop, then grow later). Such fluctuations might lead to suboptimal educational choices of workers. A suggestion for the policymakers might be therefore to implement policies increasing the number of college graduates gradually rather than rapidly.

# 2 Chapter II: Taste Heterogeneity, Elasticity of Demand and Endogenous Growth

## 2.1 Introduction

In a 1950 paper, Edward Chamberlin opposed "the supremacy of pure competition" and defended monopoly power of firms. He observed that allowing for some monopoly power is a necessary condition for sustaining product diversity. Further, he argued that variety might benefit the society because it gives a higher chance that the individual needs of consumers are satisfied.

Nearly 30 years later, Dixit and Stiglitz attempted to formalize Chamberlin's argument in an economic model. Their model possessed two features that turned out to be useful for other macroeconomic models: first, it endogenously generates product variety, and second, it provided a simple way to derive monopoly power and avoid an infinite elasticity of demand for firms. In fact, the elasticity of demand became a simple function of the elasticity of substitution between goods. These two features made the framework a convinient setup for numerous macroeconomic models.

Although the original motivation of Dixit and Stiglitz was to formalize Chamberlin's logic, the model is not an exact illustriation of his argument. Specifically, although the model shows how monopoly power could be derived from the consumers' love for variety, the source of the love for variety in the Dixit-Stiglitz model and in Chamberlin's argument differ substantially. In the Dixit-Stiglitz model, it originates from the convexity of indifference curves of a representative consumer. Chamberlin in turn argued that variety is desired because consumers' tastes differ, and a higher number of products implies a higher chance that the taste of an individual consumer will be matched.

To give an illustrative example, according to Chamberlin, the reason why Mercedes and Lexus could both enjoy monopoly power is that there are consumers who value the former much more than the latter brand, and there are other consumers who value latter more than the former. Hence, an increase in price of one of the brands will have little impact on consumer choices. According to the Dixit and Stiglitz model, the reason why Mercedes and Lexus could exercise monopoly power is that a representative consumer does not consider the two brands to be perfect substitutes.

Intuitively, the Dixit-Stiglitz model and Chamberlin's arguments are related: a higher heterogeneity of taste in population of consumers would correspond to a smaller elasticity of substitution of the representative consumer. However, no formal link has been ever established.

The model presented in this paper is designed to fill this gap. By taking the individual consumer optimization as a starting point, it sets the microfoundation for the Dixit-Stiglitz model. It derives how the convexity of indifference curves of the representative consumer (i.e. at the aggregate level) depends on the elasticity of substitution of *individual* consumers and the variance of valuation of goods by different consumers (heterogeneity of taste). It turns out that the elasticity of substitution of the representative agent (and the elasticity of demand for each good) can be written as a simple function of the two.

Relating the elasticity of substitution to taste heterogeneity has consequences for any macroeconomic model that uses the Dixit-Stiglitz setup. One example concerns the calibration of the model by Acemoglu, Aghion, Bursztyn and Hemous (2012). One of the key predictions of their model is that a higher elasticity of substitution between environmentally friendly and "dirty" goods implies higher social costs of delay in development of clean technologies. In the calibration, they consider only high values for the elasticity of substitution, arguing that "we would expect successful clean technologies to substitute for the functions of dirty technologies". Although, from an individual consumer perspective, clean and dirty technologies, such as electric and traditional automobiles, appear to be close substitutes, the relative valuation of the two types of technologies may differ substantially between consumers. The same electric car will be highly valued by a consumer in the city and provide little utility for the consumer in the country. If elasticity of substitution for an individual consumer is high, but the heterogeneity of taste between consumers is also high, the elasticity of substitution for the representative consumer remains ambiguous.

Another example, which I discuss in detail in sections 2.4-2.6, is the evaluation of Young's (1998) endogenous growth model. The original Young's model predicts that the equilibirum long-run growth of the economy does not depend on research subsidies, the total spending on R&D or the size of the economy. Its only determinants are the parameters of the research process and the elasticity of demand for goods. Since Young's (1998) model is built on the Dixit and Stiglitz framework, the elasticity of demand is solely determined by the elasticity of substitution for the representative consumer.

In section 2.4, I combine the Young endogenous growth model with the extention of Dixit-Stiglitz framework presented in this paper. The joint model predicts that higher heterogeneity leads to a lower rate of quality improvement. If the heterogeneity of taste is substantial, different consumers (or a group of consumers) have very different valuations of one product and little market can be gained by quality increase - those who have high valuation will buy the product anyway, while those with low valuation (and likely high valuation for other products) cannot be easily attracted by an increase in quality.

Interestingly, if technological progress is not driven by growth of quality, but growth of the number of products, the result is opposite. Increase in heterogeneity increases growth.

Furthermore, a minor ammendment of the model presented in section 2.5

allows for an insight on the effects of income inequality on the GDP growth. Given the logic presented above, it is expected that if a consumer's income is related to his or her tastes, higher income inequality should produce greater taste heterogeneity and lower growth. The simulation of the model (due to the loss of symmetry, the model can no longer be solved analytically) indeed confirms this intuition.

Finally, in the last section, I notice that the heterogeneity in consumers' valuation of a good can be reinterpreted as the uncertainty about a good's quality (e.g. prior to realization of the R&D process). The prediction of that reinterpretted model is that a higher uncertainty should lead to a larger product diversity - even if firms (or a planner) are risk neutral.

The remainder of this paper is structured as follows: section 2.2 describes the related literature, section 2.3 shows how elaticity of substitution can be presented as a function of taste heterogeneity meassure. Section 2.4 incorporates this result in the framework of Young's endogenous growth model to show how taste heterogeneity could affect R&D spending, technological change and long-run growth. Section 2.5 builds on the section 2.4 model to find how income inequality, if correlated with taste heterogeneity, could impact technological change and growth.

## 2.2 Related Literature

Chamberlin's argument deriving monopoly power from the differentiation in consumers' tastes has been formally described with economic models by a number of authors. One may distinguish two branches in this literature. The first consists of the models of spatial comptetion with the most prominent examples by Hotelling (1929), Salop (1979) and d'Aspremont, Gabszewicz and Thisse (1979). The second branch has been inspired by the econometric discrete choice theory (Manski and McFadden (1981) and Berry, Pakes and Levinsohn (1991)). This approach postulates using stochastic utility function to model taste heterogeneity. It was first proposed by Perloff and Salop (1985) and then adopted by Caplin and Nalebuff (1991) and Anderson, de Palma and Nesterov (1995). The paper by de Palma, Ginsburgh, Papageorgiou and Thisse (1985) combines the approaches from the two branches.

Contrary to all studies listed above, this paper is not an alternative to the Dixit and Stiglitz model. Instead, it aims to extend the Dixit and Stiglitz framework by allowing for heterogeneity of taste. As demonstrated in section 2.4, the compatability with the Dixit and Stiglitz framework allows merging the heterogeneity model with other macroeconomic models using Dixit and Stiglitz setup.

This paper is also closely related to the work by Anderson, de Palma and Thisse (1988, 1989), who show the equivalence between demand generated by the CES utility function and the logit model of the discrete choice theory. However, their model assumes that an individual consumer treats different goods as perfect substitutes. In contrast, the model presented in this paper allows the individual consumer to have a finite elasticity of substitution between goods.

The section of the paper that evaluates the microfoundation of Young's (1998) model aims to contribute to the literature of endogenous growth theory without the scale effect. The novelty of this paper is the addition of a richer demand structure analogous in some respects to the discrete choice theory - in particular, papers by Perloff and Salop (1985) and de Palma, Ginsburgh, Papageorgiou and Thisse (1985), and relating growth to consumers' heterogeneity of taste. Endogenous growth models without scale effects are presented in papers by Paretto (1998), Aghion and Howitt (1998), Dinopoulos and Thompson (1999) and Segerstrom (1998). I chose to base my model on Young's framework since, relative to other papers, his model produces the clearest link between growth and elasticity of demand with respect to quality.

## 2.3 Consumer's Heterogeneity and Elasticity of Substitution

The number of goods in the economy is given by $N_t$. Consumers' utility function takes the CES form:

$$U_{it} = \left( \sum_{j=1}^{N_t} (\theta_{ij} \lambda_{jt} x_{ijt})^{\rho} \right)^{\frac{1}{\rho}} \tag{9}$$

where $x_{ijt}$ is the quantity of product $j$ consumed by individual $i$ at time $t$, $\lambda_{jt}$ is the quality of product $j$ at time t and $\theta_{ij}$ is the idiosyncratic taste parameter [12].

In order to explore different types of competition for customers, I will consider two demand systems. In the first case, goods are imperfect substitutes for the individual consumer ($\rho < 1$). As a result, everyone consumes all products but possibly in different proportions. This partly addresses concerns raised by Pettengill (1979) that the Dixit-Stiglitz framework predicts that all consumers consume exactly equal (and very small) amounts of all products. One example of an application is the model with different types of labour (e.g. high skill, medium skill and low skill) as goods and heterogeneity of taste representing heterogeneity of production functions across sectors.

In the second scenario, goods are perfect substitutes ($\rho = 1$) and consumers choose only one product from the set of products available on the market. Again, each consumer might have his/her own valuation of each brand. This specification goes in line with discrete choice theory and corresponds to sectors like automobiles or personal computers.

[12] time indices are added to ease the incorporation of this framework into Young's endogenous growth model in section 2.4.

### 2.3.1 Imperfect Substitutes

In case of imperfect substitutes, corner solutions are ruled out. The demand of consumer $i$ for product $j$ is given by:

$$x_{ijt} = \frac{\left(\theta_{ij}\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k/p_k\right)^{\frac{\rho}{1-\rho}}} p_j^{-1} y_i \tag{10}$$

Notice that the value of $\phi = \frac{\left(\theta_{ij}\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k/p_k\right)^{\frac{\rho}{1-\rho}}}$ can be interpreted as a fraction of total real expenditure which consumer $i$ is willing to spend on the purchase of product $j$ if prices and qualities of all goods are the same.

Integrating over consumers with different tastes provides the total demand for good $j$:

$$Q_j = \int \dots \int \frac{\left(\theta_{ij}\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k/p_k\right)^{\frac{\rho}{1-\rho}}} p_j^{-1} y_i g\left(\underline{\theta}\right) d\underline{\theta}$$

When choosing the optimal level of quality and price, firm takes qualities and prices of others as given. The marginal change of quantity due to change of prices will be given by:

$$\frac{dQ_j}{dp_j} = \frac{-1}{1-\rho} \int\int \dots \int \left\{ \frac{\left(\theta_{ij}\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k/p_k\right)^{\frac{\rho}{1-\rho}}} p_j^{-2} y_i \right.$$
$$\left. -\rho \left( \frac{\left(\theta_{ij}\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k/p_k\right)^{\frac{\rho}{1-\rho}}} \right)^2 p_j^{-2} y_i \right\} g\left(\underline{\theta}\right) d\underline{\theta}$$

The elasticity of demand with respect to price will be given by

$$\frac{dQ_{jt}}{dp_{jt}} \frac{p_{jt}}{Q_{jt}} = -\frac{1}{1-\rho} \left( 1 - \rho \frac{\int\int \dots \int \phi_{ij}^2 y_i g\left(\underline{\theta}\right) d\underline{\theta}}{\int\int \dots \int \phi_{ij} y_i g\left(\underline{\theta}\right) d\underline{\theta}} \right) =$$

$$= -\frac{1}{1-\rho} \left( 1 - \rho \frac{E\left(\phi_j^2 y\right)}{E\left(\phi_j y\right)} \right)$$

If consumers' income is uncorrelated with their tastes, the expression further simplifies to:

$$\frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{Q_{jt}} = -\frac{1}{1-\rho}\left(1-\rho\frac{E\left(\phi_j^2\right)}{E\left(\phi_j\right)}\right) \tag{11}$$

Thus elasticity of demand is fully characterized by the substitutability parameter $\rho$ and the first two moments of the distribution of taste. In fact, in statistics, $D\left(\phi_j\right) = \frac{E\left(\phi_j^2\right)}{E\left(\phi_j\right)} - E\left(\phi_j\right)$ is the coefficient of dispersion of $\phi$'s distribution. The formula indicates that if these first two moments are the same for all goods, the demand curve will be the same for all goods. If all goods also have the same (upward sloping or flat) supply curves, the symmetric equilibrium exists.

For future reference, I also derive the cross-price elasticity of demand. For $k \neq j$

$$\frac{dQ_{jt}}{dp_{kt}}\frac{p_{kt}}{Q_{jt}} = \frac{\rho}{1-\rho}\frac{E\left(\phi_j\phi_k\right)}{E\left(\phi_j\right)}$$

Now define $\psi_{ij} = \frac{\theta_{ij}^{\frac{\rho}{1-\rho}}}{\sum_k \theta_{ik}^{\frac{\rho}{1-\rho}}}$. If the distribution of tastes is symmetric in the sense that $E\left(\psi_j\right) = E\left(\psi_k\right)$, $E\left(\psi_j^2\right) = E\left(\psi_k^2\right)$ and $Cov\left(\psi_j, \psi_k\right) = Cov\left(\psi_j, \psi_h\right)$ for any tripling of goods $j$, $k$ and $h$ and if all goods have the same supply curve, then symetric equilibrium exists, $p_j = p_k$, $\phi_{ij} = \psi_{ij}$, $E\left(\psi_j\right) = {}^1\!/\!_N$ and $Cov\left(\psi_j, \psi_k\right) = \frac{D(\psi_j)}{N(N-1)}$. Thus, under symmetry, we find that

$$\frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{Q_{jt}} = -\frac{1}{1-\rho}\left(1-\rho D\left(\psi_j\right) - \frac{\rho}{N}\right) \equiv -\epsilon \tag{12}$$

and

$$\frac{dQ_j}{dp_k}\frac{p_k}{Q_j} = \frac{\rho}{1-\rho}\left(-\frac{D(\psi_j)}{N-1} + \frac{1}{N}\right) \tag{13}$$

The elasticity, $|\epsilon|$, is a decreasing function of taste dispersion if $\rho > 0$ (goods would be gross substitutes in the absence of heterogeneity) and increasing function of taste dispersion if $\rho > 0$ (goods would be gross compliments).

By analogous derivations, the elasticity of demand with respect to quality is can be found:[13]

$$\frac{dQ_j}{d\lambda_j}\frac{\lambda_j}{Q_j} = \frac{\rho}{1-\rho}\left(1 - \mathrm{D}\left(\psi\right) - \frac{1}{N}\right) \equiv \epsilon - 1 \tag{14}$$

**Corollary: Representative Consumer and Heterogeneity of Taste**

Consider an economy in symmetric equilibrium (as defined above). Equations (12) and (13) predict that the Walrasian demand for good $j$ is

$$log\left(Q_j\right) = -\frac{1}{1-\rho}\left(1 - \rho D\left(\psi\right) - \frac{\rho}{N}\right) * log\left(p_j\right)$$
$$+\frac{\rho}{1-\rho}\left(\frac{-D\left(\psi\right)}{N-1} + \frac{1}{N}\right)\sum_{k\neq j}\log\left(p_k\right) + \log\left(\frac{y}{N}\right)$$

Now consider a consumer with utility function

$$U = \left(\left(0.5 * x_j\right)^{\eta} + \left(0.5 * x_k\right)^{\eta}\right)^{\frac{1}{\eta}} \tag{15}$$

where $\eta = \frac{\epsilon - 1}{\epsilon - \frac{1}{1}} = \rho\frac{1-\frac{1}{N}-D}{1-\frac{1}{N}-\rho D}$. The Walrasian demand for good $j$ in the economy populated only by this consumer is exactly the same as the one stated in equation (15). The consumer is therefore a representative consumer for an economy with heterogenous agents with utility function from equation (9).

There are two points that follows this observation. First, for the economy in symmetric equilibrium, the elasticity of substitution in the CES utility function (or production function) is a function of taste heterogeneity. If $\rho > 0$, lower

[13]The two elasticities can be also expressed in terms of variance of $\psi$: $\frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{Q_{jt}} = -\frac{1}{1-\rho}\left(1 - \rho N \mathrm{Var}\left(\psi_{ij}\right) - \frac{\rho}{N}\right) \equiv -\epsilon$ and $\frac{dQ_j}{d\lambda_j}\frac{\lambda_j}{Q_j} = \frac{\rho}{1-\rho}\left(1 - N\mathrm{Var}\left(\psi\right) + \frac{1}{N}\right) \equiv \epsilon - 1$

values of $\eta$ correspond to greater taste heterogeneity. If a model assumes very high values of $\eta$ implicitly, it assumes low heterogeneity of tastes.

Second, the corollary suggests that heterogeneity does not have to be explicitly modelled. Any model with CES utility function (or production function) implicitly allows for heterogeneity of taste between consumers. Whenever an analysis should account for taste heterogeneity of consumers, it is enough to perform comparative statics for parameter $\eta$. There is no need to build a sophisticated model with heterogenous tastes of agents.

### The Ideal Meassure of Heterogeneity

The choice of measure for taste heterogeneity is not trivial. First, two candidates are the variance and the dispersion of $\phi_j$, an income share devoted for good $j$. If income share devoted to good $j$ differs between consumers, it must be due to taste heterogeneity. The spread of income shares devoted to good $j$ across the population may therefore serve as a measure of heterogeneity in preferences. The major adventage of the two measures is that they are easily empirically observable. Their major problem is that the distribution of $\phi$ is defined on the simplex[14] and therefore depends on the number of goods: if a new good becomes available, the distribution of $\phi$ will change and thus its variance and dispersion.

In models where $N$ is fixed, $Var(\psi_j)$ and $D(\psi_j)$ can be treated as determined purely by factors outside the model. However, in models in which $N$ is endogenous, $Var(\psi_j)$ and $D(\psi_j)$ can no longer be taken as exogenous. How could we find the alternative? What remains exogenous is the distribution of taste in the utility function, $\theta$. Recall that $\theta_{ij}$ is the idiosyncratic taste parameter, a weight each customer $i$ puts on consumption of good $j$; since $\theta_{ij}$ can be any positive number chosen by the consumer and the sum $\sum \theta_{ij}$ does not have

[14] Since every consumer chooses an arrow $(\phi_1, \phi_2 ... \phi_N)$ such that $\sum_{j=1}^{N} \phi_j = 1$, the distribution of $\phi$ is a multivariate distribution with simplex $\Delta^{N-1}$ as a support.

to be unity, the distribution of $\theta_j$ can be completely independent of $N$.

A handy alternative meassure of heterogeneity of taste between consumers is a coefficient of dispersion of $\theta_j^{\frac{\rho}{1-\rho}}$, $D\left(\theta_j^{\frac{\rho}{1-\rho}}\right)$[15]. Consider the case of symmetric tastes, i.e. the same distribution of $\psi$ for all goods. If the goal is to express the right hand side of equation (12) in terms of $D\left(\theta^{\frac{\rho}{1-\rho}}\right)$, the relation between $\frac{E(\psi^2)}{E(\psi)}$ and $D\left(\theta^{\frac{\rho}{1-\rho}}\right)$ shall be found. The relation will depend on the particular distribution of $\theta^{\frac{\rho}{1-\rho}}$. An example of a distribution which allows for an elegant closed form solution is the gamma distribution.

If $\theta^{\frac{\rho}{1-\rho}} \sim Gamma\,(\alpha, \beta)$ then the dispersion of $\theta^{\frac{\rho}{1-\rho}}$ is $D = \frac{1}{\beta}$ and its expected value is $\mu = \alpha D$. Under the mean preserving spread - i.e. if upon increase in dispersion, parameter $\alpha$ adjusts to keep the mean unchanged - the distribution of income shares is Dirichlet, $\psi = \frac{\theta^{\frac{\rho}{1-\rho}}}{\sum_k \theta^{\frac{\rho}{1-\rho}}} \sim Dirichlet\left(\frac{\mu}{D}, \frac{\mu}{D}, ..., \frac{\mu}{D}\right)$ and

$$\frac{E\left(\psi^2\right)}{E\left(\psi\right)} = \frac{\mu + D}{N\mu + D} \tag{16}$$

which is an increasing function of $D$ for $n \geq 2$.

If the assumption on the symmetry of distributions is dropped and the taste for each product is allowed to follow its own distribution, $\theta_j^{\frac{\rho}{1-\rho}} \sim Gamma\,(\alpha_j, \beta_j)$, then $D_j = \frac{1}{\beta_j}$. The income share for product $j$ is distributed according to $\phi_j = \left(\theta_j\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}} \sim Gamma\left(\alpha_j, \beta_j\left(\lambda_j/p_j\right)^{-\frac{\rho}{1-\rho}}\right)$ and its expected value is $\mu_j = \frac{\alpha_j}{\beta_j}\left(\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}} = \alpha_j D_j \left(\lambda_j/p_j\right)^{\frac{\rho}{1-\rho}}$. Again, under the mean preserving spread, the distribution of income shares is Dirichlet, $\psi = \frac{\theta^{\frac{\rho}{1-\rho}}}{\sum_k \theta^{\frac{\rho}{1-\rho}}} \sim Dirichlet\left(\frac{\mu_1}{D_1}\left(\lambda_1/p_{1j}\right)^{-\frac{\rho}{1-\rho}}, ..., \frac{\mu_n}{D_n}\left(\lambda_n/p_n\right)^{-\frac{\rho}{1-\rho}}\right)$. It follows that

$$\frac{E\left(\phi_j^2\right)}{E\left(\phi_j\right)} = \frac{\frac{\mu_j}{D_j}\left(\lambda_j/p_j\right)^{-\frac{\rho}{1-\rho}} + 1}{\sum_k \frac{\mu_k}{D_k}\left(\lambda_k/p_k\right)^{-\frac{\rho}{1-\rho}} + 1}$$

The expression is decreasing in $D_j$ and increasing in $D_k$, implying that elasticity

[15] Perhaps a more natural meassure of heterogeneity would be $Var(\theta_j)$ or $D(\theta_j)$. It turns out, however, that such meassure involves much more problematic and less tractable derivations. In the appendix B, I show that, if $\theta_j^{\frac{\rho}{1-\rho}}$ is distributed with the gamma distribution, the ralation between $D(\theta_j^{\frac{\rho}{1-\rho}})$ and $D(\theta_j)$ is positive.

of demand for good $j$ increases when the dispersion of $\theta_j^{\frac{\rho}{1-\rho}}$ across population increases and decreases when the dispersion of $\theta_k^{\frac{\rho}{1-\rho}}$ with $k \neq j$ increases. If the dispersion of taste increases for all products by the same factor, elasticity of demand for each product decreases.

### 2.3.2 Perfect Substitutes

It turns out that mathematical analysis of this case is complex, and some simplifying assumptions on the distribution of idiosyncratic taste are necessary to proceed. In particular, in this subsection we assume that $\theta_{ij}$ is independently and identically distributed across products (the assumption which is typical in the discrete choice theory). Later, I will also assume that the log of $\theta_{ij}$ follows a logistic distribution. These assumptions reduce the generality of the problem but help with picturing the basic mechanism which this paper intends to describe.

Under the perfect substitution case, consumer maximization can be written as

$$U_i = \max \sum_{j=1}^{N} (\theta_{ij} \lambda_j x_{ij})$$

subject to the budget constraint

$$y = \sum_j x_{ij} p_j$$

In fact, this problem reduces to the simple choice of the product which gives the highest value to the consumer:

$$U_i = \max_j \left\{ \theta_{ij} \lambda_j \frac{y_i}{p_j} \right\} = \max_j \left\{ \ln \theta_{ij} + \ln \lambda_j + \ln y_i - \ln p_j \right\}$$

To find aggregate demand for each product, I follow the same strategy as in

Perloff and Salop (1985). The probability that consumer $i$ prefers product $j$ to product $k$:

$$\Pr(\ln\theta_{ij} + \ln\lambda_j + \ln y - \ln p_j > \ln\theta_{ik} + \ln\lambda_k + \ln y - \ln p_k) =$$

$$= \Pr(\ln\theta_{ik} < \ln\theta_{ij} + \ln\lambda_j - \ln\lambda_k - \ln p_j + \ln p_k) =$$

$$= G(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda_k - \ln p_j + \ln p_k)$$

where $G$ denotes the cumulative distribution function of $\theta_{ij}$.

Since $\theta_{ik}$ is independent and identically distributed across products, I find that the probability of choosing product $j$ given $\theta_{ij}$ is given by

$$\Pr(j \succ 1 \cap ... \cap j \succ j-1 \cap j \succ j+1 \cap ... \cap j \succ N) =$$

$$= \prod_{k \neq j} G(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda_k - \ln p_j + \ln p_k)$$

To find the aggregate demand for product $j$, I integrate over population:

$$Q_j = \frac{y}{p_j} L \int \prod_{k \neq j} G(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda_k - \ln p_j + \ln p_k)\, g(\ln\theta_{ij})\, d\ln\theta_{ij} \quad (17)$$

I consider only the symmetric equilibria. This allows further simplification of (17):

$$\frac{y}{p_j} L \int G(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda - \ln p_j + \ln p)^{N-1}\, g(\ln\theta_{ij})\, d\ln\theta_{ij} \quad (18)$$

Now I am able to derive the elasticities of demand with respect to price and quality. I start with the latter.

Using (18) I obtain:

$$\frac{\partial Q_j}{\partial p_j}\frac{p_j}{Q_j} = -1 - \frac{p_{jt}}{Q_j}\frac{y}{p_j^2}L\int (N-1)\,G\left(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda - \ln p_j + \ln p\right)^{N-2} *$$

$$*g\left(\ln\theta_{ij} + \ln\lambda_j - \ln\lambda - \ln p_j + \ln p\right) g\left(\ln\theta_{ij}\right) d\ln\theta_{ij}$$

and since in symmetric equilibrium $\lambda_j = \lambda$ and $p_j = p$, this simplifies to

$$\frac{\partial Q_j}{\partial p_j}\frac{p_j}{Q_j} = -\left(1 + \frac{p_{jt}}{Q_j}\frac{y}{p_j^2}L\int (N-1)\,G\left(\ln\theta_{ij}\right)^{N-2} g\left(\ln\theta_{ij}\right)^2 d\ln\theta_{ij}\right)$$

Analogous derivations gives the expression of elasticity of demand with respect to quality

$$\frac{\partial Q_j}{\partial \lambda_j}\frac{\lambda_j}{Q_j} = \frac{\lambda_j}{Q_j}\frac{y}{p_j\lambda_j}L\int (N-1)\,G\left(\ln\theta_{ij}\right)^{N-2} g\left(\ln\theta_{ij}\right)^2 d\ln\theta_{ij}$$

The assumption that $\ln\theta_{ij}$ follows the exponential distribution with $E\left[\ln\theta\right] = 0$ and $Var\left[\ln\theta\right] = \sigma^2$ allows to find a simple closed form solution.

$$\frac{\partial Q_{jt}}{\partial p_{jt}}\frac{p_{jt}}{Q_{jt}} = -\epsilon = -\left(1 + \frac{1}{\sigma}\right) \tag{19}$$

$$\frac{\partial Q_{jt}}{\partial \lambda_{jt}}\frac{\lambda_{jt}}{Q_{jt}} = \epsilon - 1 = \frac{1}{\sigma}$$

Table 1 presents the form of the elasticity under different types of distribution.

## 2.4 Young's Model of Endogenous Growth

Finding a functional relation between elasticity of demand and taste heterogeneity opens room for a reinterpretation of a wide class of macroeconomic models. Here, I present one example: an evaluation of Young's (1995) endognous growth

| Distribution | Distributed | $\epsilon$ | Comments on Distribution | Comments on Growth (refer to section 2.4) |
|---|---|---|---|---|
| Exponential | $\log(\theta)$ | $\frac{1}{\sigma}$ | Depicted on Figure 3. The support in terms of $\theta$: $\theta \in (1, \infty)$. Most consumers are neutral about the product ($\theta$ close to one). The group of product lovers is in the right tail. Number of product lovers (thickness of the tail) increases with variance, $\sigma^2$ | Under vertical spillover, growth is constant and negatively related to heterogeneity. Under horizontal spillover, growth is constant and positively related to heterogeneity. |
| Logistic | $\log(\theta)$ | $\frac{1}{\sigma}\frac{N_t-1}{N_t+1}$ | Depicted on Figure 4. The support in terms of $\theta$: $\theta \in (0, \infty)$. For large $\sigma$: most consumers have zero valuation of the product ($\theta$ close to zero). The group of product lovers is in the right tail. Number of product lovers (thickness of the tail) increases with variance, $\sigma^2$. For small $\sigma$ the distribution resembles log normal distribution ( with product lovers in a right tail) | Under vertical spillover, growth is constant and negatively related to heterogeneity. Under horizontal spillover, if $N_t$ is very large, growth is constant and positively related to heterogeneity. |
| Normal | $\log(\theta)$ | $\frac{1}{\sigma} * 0.5^{N_t} N_t *$ $* constant$ | This is a log normal distribution of $\theta$ | Under vertical spillover, growth is zero if $N_t$ goes to infinity. For any number of goods greater than one, elasticity is decreasing in number of goods - hence there is no clear relation between heterogeneity and growth |
| Uniform on $(-\sigma/2, \sigma/2)$ | $\log(\theta)$ | $\frac{N_t}{\sigma}$ | Depicted on Figure 5. The support in terms of $\theta$: $\theta \in \left(e^{-\sigma/2}, e^{\sigma}\right)$ | Under vertical spillover, growth is explosive if $N_t$ goes to infinity. There is no constant growth under horizontal spillovers |

Table 4: Elasticities of Demand under Different Distributions of Taste.

model. The original Young's model predicts that growth depends solely on elasticity of substitution between goods. This result, combined with the result in the previous section, implies a negative relation between taste heterogeneity and growth.

In the last subsection, I will elaborate on the possibility that taste heterogeneity originates from income inequality. In this case, the symmetry between goods is lost and growth of quality of each good is different. An increase in heterogeneity will reduce equilibrium growth of quality for at least some of the goods. There might be, however, a group of goods (particularly those goods which receive high valuation by high income consumers) whose quality growth is going to increase with an increase in inequality.

### 2.4.1 Competition for Consumers as an Engine of Growth.

The life-time utility is given by the discounted sum of logs of instantenous utilities :

$$V_i = \sum \beta^t \log \left(U_{it}\right)$$

As before, consumers' instantenous utility function takes the CES form:

$$U_{it} = \left( \sum_{j=1}^{N_t} \left(\theta_{ij} \lambda_{jt} x_{ijt}\right)^{\rho} \right)^{\frac{1}{\rho}}$$

Thus the demand side of the economy is the same as described in the previous section.

We assume that each firm produces only one product. All products are produced by the constant returns to scale technology using labor as the only input. The unit cost of production is the same for all goods and is denoted by $c$. Since I will use wage of workers as a numeraire, $c$ is also the inverse of the productivity of the worker. Every period, $t$ firms decide whether in period $t+1$

they wish to enter the market or stay aside (with the outside option giving zero profit). If a firm enters, it plays the Bertrand-Nash game in prices and qualities with other firms on the market in period $t+1$: it chooses the price and quality level to maximize profit given the choices of other firms. The fixed cost that firms have to cover before entering is the R&D spending in period $t$. This cost, however, may be tuned by the choice of quality level. Thus the maximization problem of firms' objectives can be formulated as follows:

$$\max_{\lambda_{jt},p_{jt}} \frac{(p_{jt}-c)\,Q_{jt}}{1+r_{t-1}} - F\left(\lambda_{jt},\overline{\lambda}_{j,t-1}\right)$$

where $F\left(\lambda_{jt},\overline{\lambda}_{j,t-1}\right)$ denotes the cost of developing quality $\lambda_{jt}$. I will use the same formulation as in Young's (1998) paper:

$$F\left(\lambda_{jt},\overline{\lambda}_{j,t-1}\right) = \begin{cases} fe^{a\lambda_{jt}/\overline{\lambda}_{j,t-1}} & \text{if } \lambda_{jt} \geq \overline{\lambda}_{j,t-1} \\ fe^{a} & \text{otherwise} \end{cases} \tag{20}$$

If good $j$ was previously produced, $\overline{\lambda}_{j,t-1} = \lambda_j^{max}$ is the highest quality of product $j$ produced up to time $t-1$. If good $j$ was not previously produced, $\overline{\lambda}_{j,t-1} = \max_k \left\{\lambda_{k,t-1}^{max}\right\}$ is the most advanced quality developed among all products available in period $t-1$. This means that all new products can use the last period frontier technology. This form of the cost function captures the idea of vertical spillover - the greater the advancements in past are, the easier it is to progress in development of product quality. Such form guarantees that constant spending on R&D across periods (or a constant number of researchers and engineers working on the development of one product) gives constant growth of quality. This form of the cost function implies also that there are some minimal fixed costs of entering the market - even if the firm chooses close to zero quality, it still has to pay $fe^a$ to cover the R&D on the initial development of the product.

I am interested in the symmetric subgame perfect nash equilibrium of the model. For firms which enter the market, first order conditions for optimal choices of prices and quality are

$$Q_{jt} + (p_{jt} - c)\frac{\partial Q_{jt}}{\partial p_{jt}} = 0$$

$$\frac{(p_{jt} - c)}{1 + r_{t-1}}\frac{\partial Q_{jt}}{\partial \lambda_{jt}} - \frac{\partial F\left(\lambda_{jt}, \overline{\lambda}_{j,t-1}\right)}{\partial \lambda_{jt}} = 0$$

The equilibrium will be also characterized by $N_t$ - the number of firms in the market which assures that, given choices in the second stage of the game, firms are indifferent between entering or not entering the market. This result in the zero profit condition:

$$\frac{(p_{jt} - c)\,Q_{jt}}{1 + r_{t-1}} = F\left(\lambda_{jt}, \overline{\lambda}_{t-1}\right)$$

This condition allows expression of FOCs in terms of elasticities:

$$\frac{p_{jt} - c}{p_{jt}} = -\left(\frac{\partial Q_{jt}}{\partial p_{jt}}\frac{p_{jt}}{Q_{jt}}\right)^{-1} = \frac{1}{E_{p_j}^{Qj}} \tag{21}$$

$$E_{\lambda_j}^{Qj} = \frac{\partial Q_{jt}}{\partial \lambda_{jt}}\frac{\lambda_{jt}}{Q_{jt}} = \frac{\partial F_{jt}}{\partial \lambda_{jt}}\frac{\lambda_{jt}}{F_{jt}} = E_{\lambda_j}^{Fj}$$

This leads to the important implication of the Young's paper: firms' optimization and zero profit conditions imply that firms set quality levels to equalize elasticity of demand with respect to quality and elasticity of R&D costs with respect to quality.

Under the production function of ideas specified in (20), the elasticity of costs with respect to quality is given by:

$$E^{Qj}_{\lambda_j} = a\lambda_{jt}/\overline{\lambda}_{jt-1} \tag{22}$$

Since I consider a symmetric equilibrium, I find that

$$\epsilon - 1 = a\lambda_{jt}/\overline{\lambda}_{jt-1} = a\lambda_t/\lambda_{t-1} \tag{23}$$

and

$$\frac{p_{jt} - c}{p_{jt}} = \frac{1}{\epsilon}$$

Following Young's paper, I can close the model with the labor resource constraint - labor in manufacturing and labor in the research sectors has to equal the total labor supply which - given that labor is supplied inelastically - is equal to the size of population, $L$. The amount of workers in the manufacturing sector can be found from the total amount of goods produced and workers' productivity:

$$L_m = N_t Q_t c = \frac{c}{p} E_t = \frac{\epsilon - 1}{\epsilon} E_t$$

where $E_t$ is a total expenditure. On the other hand, since wage was normalized to unity, labor employed in the research sector has to be equal to the total spending on R&D:

$$L_r = N_{t+1} f e^{\epsilon - 1}$$

This can also be expressed in terms of total expenditure by making use of the zero profit condition: total spending on R&D has to equate the total benefit from sell

$$L_r = N_{t+1} f e^{\epsilon - 1} = N_{t+1} \frac{(p - c)\, Q_{t+1}}{1 + r_t} =$$
$$= \frac{E_{t+1}}{\epsilon\,(1 + r_t)} = \frac{\beta E_t}{\epsilon} \tag{24}$$

where the last equality follows from consumer intertemporal optimization. Therefore, the labor resource constraint predicts

$$L = \left(\frac{\epsilon - 1 + \beta}{\epsilon}\right) E_t \tag{25}$$

Substituting it back to (24) gives the optimal number of firms:

$$N_{t+1} f e^{\epsilon - 1} = \frac{\beta}{\epsilon - 1 + \beta} L \tag{26}$$

### 2.4.2 Effect of Heterogeneity on Growth under Vertical Spillover and Imperfect Substitutability between Goods

Now I am able to combine the measure of consumer heterogeneity with endogenous growth theory model. I will first consider the case of vertical spillovers and imperfect substitability between goods.

To find a relation between growth and dispersion of income shares devoted for good $j$, $\psi_j$ equations (14) and (23) can be combined:

$$\frac{\lambda_t}{\lambda_{t-1}} = \frac{1}{a}\frac{\rho}{1-\rho}\left(1 - \mathrm{D}(\psi) - \frac{1}{N}\right) \equiv \epsilon - 1 \tag{27}$$

As discussed in section 2.3.1., in the context of models with endogenous number of goods, $D(\psi)$ cannot be treated as an exogenous parameter. Instead, I can represent the heterogeneity of taste with a dispersion of $\theta^{\frac{\rho}{1-\rho}}$. Assuming that $\theta^{\frac{\rho}{1-\rho}}$ follows a gamma distribution, I can obtain the relation between growth and $D\left(\theta^{\frac{\rho}{1-\rho}}\right)$ by combining equations (22), (11) and (16):

$$\begin{aligned}\frac{\lambda_t}{\lambda_{t-1}} &= \frac{1}{a}\frac{\rho}{1-\rho}\left(1 - \frac{\mu + D}{N\mu + D}\right) \\ &= \frac{1}{a}\frac{\rho}{1-\rho}\frac{(N-1)\mu}{N\mu + D}\end{aligned} \tag{28}$$

For constant $N$, an increase in the dispersion decreases elasticity and growth. If I allow $N$ to adjust, the decrease of elasticity will increase profits and thus create a room for new varieties (the mechanism is captured by equation (26)). A higher number of goods will have a positive effect on elasticity (as shown in equations (11) and (16)), thus partly offseting its initial drop.

The total effect of heterogeneity on the elasticity, however, remains negative. This can be easily demonstrated by contradiction: suppose that $\epsilon - 1$ increases after an increase in $D$. From equation (28), this would imply that number of products has to increase. As a result, the LHS of (26) (i.e. the total spending on research) would have to increase. However, as $\epsilon - 1$ increases it is known that in equilibrium, the RHS of (26) (i.e. equilibrium total benefit from sell) would need to fall - which is a contradiction.

As described with equation (22), a lower elasticity of demand leads to less intense competion and lower growth. Therefore, an increase in heterogeneity will have a negative effect on growth in the equilibrium.

**Welfare growth**

The final step in this subsection is to examine how technological progress translates into welfare growth under vertical spillovers.

Under symmetric equilibrium, the utility of each individual can be expressed as

$$U_{it} = \frac{y_{it}}{p}\lambda\left(\sum_{j=1}^{N_t}\left(\theta_{ij}\frac{\theta_{ij}^{\frac{\rho}{1-\rho}}}{\sum_k \theta_{ik}^{\frac{\rho}{1-\rho}}}\right)^{\rho}\right)^{\frac{1}{\rho}}$$

$$= \frac{y_{it}}{p}\lambda\left(\sum_k \theta_{ik}^{\frac{\rho}{1-\rho}}\right)^{\frac{1-\rho}{\rho}}$$

and hence total welfare is simply given by

$$W_t = \frac{\lambda_t}{p} y_t E\left(\left(\sum_k \theta_{ik}^{\frac{\rho}{1-\rho}}\right)^{\frac{1-\rho}{\rho}}\right) \tag{29}$$

As argued above, under vertical spillover, the number of products in equilibrium has to be constant, which implies that $E\left(\left(\sum_k \theta_{ik}^{\frac{\rho}{1-\rho}}\right)^{\frac{1-\rho}{\rho}}\right)$ is also constant. Additionally, elasticities of demand with respect to price and quality need to be constant, and this involves constant prices as well as - by the relation between total expenditure and labor (equation (25)) - constant per capita expenditure, $y_t$. Therefore, the growth of wealth in equilibrium is given exactly by the growth of the equilibrium quality level:

$$\frac{d\log(W_t)}{dt} = \frac{d\log(\lambda_t)}{dt}$$

### 2.4.3 Effect of Heterogeneity on Growth under Horizontal Spillover and Imperfect Substitutability between Goods

So far, I have analyzed a form of R&D production function, which assumes the presence of vertical spillovers: the higher the stock of knowledge is, the easier it is to develop a higher quality for each good in next period. On the other hand, the stock of knowledge embedded in the variety of goods has absolutely no impact on the R&D process.

In this subsection, I consider the opposite case - the case of horizontal spillovers: the current quality level does not ease development of the next period quality level; instead a higher number of goods already in place facilitates the development of new goods.

Following Young, horizontal spillovers are introduced here by modifying the

R&D cost function to

$$F\left(\lambda_{jt}, \overline{\lambda}_{t-1}\right) = \frac{f e^{a\lambda_{jt}}}{N_{t-1}} \tag{30}$$

Under this new specification, if heterogeneity is constant over time, and if the number of products is large enough (so that $1/N$ is negligible), the quality level is constant:

$$\lambda_t = \frac{1}{a}\frac{\partial F_{jt}}{\partial \lambda_{jt}}\frac{\lambda_{jt}}{F_{jt}} = \frac{\epsilon - 1}{a} = \tag{31}$$

$$= \frac{\rho}{1-\rho}\frac{1}{a}\left(1 - \frac{\mu + D}{N_t \mu + D}\right)$$

On the other hand, equation (26) becomes

$$\frac{N_{t+1}}{N_t} f e^{\epsilon - 1} = \frac{\beta}{\epsilon - 1 + \beta} L \tag{32}$$

predicting a constant growth of the number of goods in the economy:

$$g_N = \frac{N_{t+1}}{N_t} = \frac{\beta}{\epsilon - 1 + \beta} f e^{\epsilon - 1} L \tag{33}$$

The growth rate is a decreasing function of the elasticity of demand. Since the elasticities of demand, $(\epsilon)$, decreases with heterogeneity (see equation (31)), this growth rate increases with heterogeneity - the result which is opposite to the one under vertical spillover.

However, the size of the effect depends on the number of goods. As the number grows exponentially, the effect of heterogeneity on growth at some point becomes negligable.

**Welfare growth.**

If $\theta_j^{\frac{\rho}{1-\rho}} \sim Gamma\left(\frac{\mu}{D}, \frac{1}{D}\right)$, then $\sum_j theta_j^{\frac{\rho}{1-\rho}} \sim Gamma\left(\frac{N(t)\mu}{D}, \frac{1}{D}\right)$. It can be shown then that $E\left(\left(\sum_j \theta_j^{\frac{\rho}{1-\rho}}\right)^{\frac{1-\rho}{\rho}}\right) = D^{\frac{1-\rho}{\rho}} \frac{\Gamma\left(\frac{1-\rho}{\rho} + \frac{N(t)\mu}{D}\right)}{\Gamma\left(\frac{N(t)\mu}{D}\right)}$ where $\Gamma(.)$ is the gamma function. Using this result in equation (29)

$$W_t = \frac{\lambda}{p} y D^{\frac{1-\rho}{\rho}} \frac{\Gamma\left(\frac{1-\rho}{\rho} + \frac{N(t)\mu}{D}\right)}{\Gamma\left(\frac{N(t)\mu}{D}\right)} \tag{34}$$

The growth rate is then given by:

$$\frac{d\log(W_t)}{dt} = \frac{\mu}{D} e^{g_N t} \left( \psi^{(0)} \left( \frac{1-\rho}{\rho} + \frac{\mu}{D} e^{g_N t} \right) - \psi^{(0)} \left( \frac{\mu}{D} e^{g_N t} \right) \right)$$

where $\psi^{(0)}$ is the digamma function. The simulation reveals that in the limit (as t grows) this expression approaches $\frac{1-\rho}{\rho}$.

#### 2.4.4 Effect of Heterogeneity on Growth under Perfect Substitutability between Goods.

As argued in section 2.3.2, the elasticity of demand is negatively related to consumer heterogeneity for all distributions that were considered. On the other hand, the discussion in the section 2.4.1 implies a positive relation between elasticity of demand and growth in the setup with vertical spillovers. Heterogeneity must therefore reduce growth. To understand this result, consider the extreme case of heterogeneity in which each producer has a group of faithful consumers who strongly dislike all other goods. In this case, no quality improvement attracts new customers, as consumers outside the 'fan club' have their own favourites. This reduces the incentive for a firm to innovate.

To complete the analysis for general equilibrium, I shall take into account the endogenous response of a number of variaties, $N$. Under vertical spillovers, a decrease in elasticity of demand upon growth of heterogeneity will increase profit and attract new entrants. This response of $N$ will lead to an increase in elasticity of demand which offsets its initial fall. However, the total effect of heteronegeity on growth remains negative.

In case of horizontal spillovers, a negative relation between elasticity of demand and growth predicts a positive effect of heterogeneity on growth. Upon arrival of a new product, consumers will draw their valuations from the distribution. If the dispersion parameter of the distrubution is high, the potential entrant expects that a group of consumers will have valuations high enough to be faithful, irrespective of quality and price. This raises the monopoly power and increases profit opportunities promoting the entry. Under horizontal spillovers, the higher number of entrants leads to larger growth in long run.

Under horizontal spillovers, growth of the number of goods will eventually lead to constant and finite elasticity of demand under exponential and logistic distribution, infinite elasticity under uniform distribution and zero elasticity under normal distribution.

**Welfare Growth**

Again, to finalize this subsection, a link needs to be founds between technological progress and welfare growth. Under the setup drawn in this section, the total welfare will be the same as ex-ante utility of an agent (i.e. expected utility before his or her taste parameters are realized):

$$W_t = E_{\underline{\theta}}\left(U_{it}\right) = \lambda_t \frac{y_t}{p_t} E\left(\max_j \left\{\theta_{ij}\right\}\right)$$

As before, prices and per capita expenditure are constant, so the only source of growth is the improvement in quality (arising under vertical spillover) and the higher number of products (arising under horizontal spillover case). Since the number of goods in the vertical spillover case is constant, the link between wealth and quality growth is straightforward

$$\frac{d\log\left(W_t\right)}{dt}=\frac{d\log\left(\lambda_t\right)}{dt}$$

The growth in the horizontal spillover case comes through the increasing number of products - ex ante an agent might expect that with a bigger choice, he will be able to find a product that gives him a higher taste parameter, $\theta$. The mathematical evaluation of this effect is not obvious; however, following the assumption that $\log(\theta)$ is distributed with the exponential distribution, it can be shown that for large $N_t$

$$\max_j\left\{\log\left(\theta_{ij}\right)\right\}=\log\left(N_t\right)$$

(see table 1 in Gabaix, Laibson and Li (2010)). Since $\log E\left(\max_j\left\{\theta_{ij}\right\}\right)$ can be approximated (using first order Taylor approximation) with $E\left(\log\left(\max_j\left\{\theta_{ij}\right\}\right)\right)$ log welfare can be approximated with

$$\log\left(W_t\right)=\log\left(\lambda\right)+\log\left(y\right)-\log\left(p\right)+\log\left(N_t\right)$$

With constant elasticity, $\epsilon$, quality, prices and income are constant. As a result

$$\frac{d\log\left(W_t\right)}{dt}=\frac{d\log\left(N_t\right)}{dt}=\frac{1}{\sigma}$$

under exponential distribution of $log(\theta)$. The growth of welfare in this economy is given by the growth of the number of products, which - as shown - depends positively on heterogneity in the horizontal spillover case.

## 2.5 Income Inequality and Endogenous Growth.

This section shows how the model in section 2.4 can be extended to cover heterogeneity in income. It demonstrates that as long as tastes are correlated with the income of the individual, an increase in income inequality (leading to a higher taste heterogeneity) will have an implication for technological change and economic growth. Through this channel, a redistributive policy (or its lack) might have an indirect effect on long run economic growth.

Allowing for heterogeneity in productivity (and hence incomes) leads to the following form of elasticity of demand with respect to price

$$\epsilon = \frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{x_{jt}} = -\frac{1}{1-\rho}\left(1-\rho\frac{E\left(\phi_j^2 y\right)}{E\left(\phi_j y\right)}\right) \tag{35}$$

If there is no heterogeneity in taste across consumers, $\phi_j$ is a constant and under assumption of symmetry $\phi_j = \psi_j = \frac{1}{N}$. Then the expression for elasticity becomes

$$\epsilon = \frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{x_{jt}} = -\frac{1}{1-\rho}\left(1-\rho\frac{1}{N}\right)$$

thus the elasticity becomes independent of the distribution of income.

Similarly, if an individual's tastes are independent of his/her income ($Corr\left(\psi_j, y\right) = 0$ and $Corr\left(\psi_j^2, y\right) = 0$) equation (35) becomes

$$\epsilon = \frac{dQ_{jt}}{dp_{jt}}\frac{p_{jt}}{x_{jt}} = -\frac{1}{1-\rho}\left(1-\rho\frac{E\left(\phi_j^2\right)E\left(y\right)}{E\left(\phi_j\right)E\left(y\right)}\right)$$

$$= -\frac{1}{1-\rho}\left(1-\rho\frac{E\left(\phi_j^2\right)}{E\left(\phi_j\right)}\right)$$

and therefore reduces to the form in (12) - again elasticity of demand with respect to price is independent of the parameters of income distribution. This is in line with the conclusions of Foellmi and Zweimuller (2003), who find that when consumer preferences are homothetic, income inequality does not have an impact on elasticity of demand.

The situation changes, however, if a consumer's tastes are not independent of his/her income level. Below, I discuss some instances in which the dependance between the two might arise.

The first reason why the parameter might depend on income is that the CES form of utility does not capture consumer preferences correctly. If income can be treated as a proxy for the overall level of consumption, in many cases a dependence of theta on income might lead to a better description of preferences. One example could be the choice between public transportation vs. a car. The CES preference implies that, no matter what the consumer's income is, he or she will always choose the option that has the better value for the money. In practice, this is not the case - wealthier consumers tend to choose cars, while poorer consumers choose public transportation. Dependance of theta on income, although imperfectly, can capture this preference. Foelmi and Zweimuller (2003) provide a formal and more general discussion on how nonhomothetic preferences may lead to dependence of elasticity of demand on income distribution.

The Second reason for the correlation between tastes and income is that income might capture other characteristics of the consumer (normally not included as consumer goods), such as education or age. Before computer interfaces were simplified, their value to uneducated consumers was probably negligable, while educated consumers could enjoy exploring new technology and use it, for instance, for better time management. If income in the model serves simply as a proxy for other exogenous consumer characteristics, obviously the model cannot be used to evaluate the impacts of redistributive policies. However, it can serve as a tool to predict the implications of equal access to education policy or demography dynamics on long-run technological change.

The third reason that income might have an impact on tastes is because it often puts individuals in various social strata. An affilation with a particular social stratum has implications for consumer preferences. Thus, for middle-class consumers, possession of a car might carry an additional value because it signals affilation with the middle class. For consumers from the lower social strata, a mobile phone may play a more important role than for others because of security reasons - lower stratum consumers could be more exposed to danger and a mobile phone is essential in emergencies.

Finally, an example of a very direct effect of income on tastes can also be given: how much a consumer values the quality of financial services depends on how much he/she interacts with financial institutions. This, in turn, could depend on the income of a consumer.

In general, if income is correlated with taste parameters, the symmetry between goods is lost and thus equation (35) is not valid. There are two ways to proceede: one is to look for the functional relation between income and taste parameters that preserve the symmetry. The second is to simulate the problem numerically and find how an increase in income inequality affects elasticity of demand for each good.

The first approach - the imposition of a specific functional relation between income and taste parameters - greately reduces generality of the problem. However, due to the fact that symmetry between goods is preserved, the problem can still be solved analytically, and this allows me to precisely track the way that income inequality may affect elasticity of demand and long-run economic growth. The numerical simultation that I perform afterwards helps to ensure that the conclusions are not driven only by the assumption of the specific relationship between income and taste parameters.

In both analytical solutions and numerical simulations, I consider a simplified model with only two goods (A and B) and two groups of consumers: rich and poor. Let $\psi_{ij}$ denote the group's $i \in \{p, r\}$ taste parameter for product $j \in \{A, B\}$. The income of rich consumers is normalized to 1. The income of poor consumers is $k$, a fraction of rich consumers' income. Then equation (35) evaluated for product A becomes

$$\epsilon_A = \frac{dQ_{jt}}{dp_{jt}} \frac{p_{jt}}{x_{jt}} = \frac{1}{1-\rho} \left( 1 - \rho \frac{k * \psi_{p,A}^2 + \psi_{r,A}^2}{k * \psi_{p,A} + \psi_{r,A}} \right)$$

It can be shown that if $\psi_{p,A} = \frac{1}{2} - \frac{2\psi_{r,A} - 1}{2k}$ then elasticity of demand for good A and for good B are equal, $\epsilon_A = \epsilon_B$

Suppose that $\psi_{r,A} = 0.75$. Then $\psi_{p,A} = \frac{1}{2}\left(1 - \frac{1}{2k}\right)$, which is a decreasing function in $k$. Under this functional relation between one of the taste parameters and income, elasticity of demand for good A (and for good B) reduces to:

$$\epsilon = \frac{1}{1-\rho} \left( 1 - \frac{\rho}{2} - \frac{1 - 4\psi_{r,A}\left(1 - \psi_{r,A}\right)}{2k} \right)$$

which is an increasing function in $k$. Therefore, an increase in $k$ (corresponding to reduction in income inequality) leads to a reduction of taste heterogeneity and an increase in elasticity of substitution (or elasticity of demand). This in turn generates greater competition between producers of two goods and produces a higher growth of quality, greater technological change and higher growth in and the long-run.

As promised earlier, I also present the results of the numerical simulation of the model. As before, I assume that there are two types of consumers: poor and rich. I also assume that there are two categories of goods, labeled A and B.

In the simulations, the sizes of the populations of rich and poor consumers are equal to unity. I will focus on the case with $\rho = 0.33$, which implies a moderate substitution between goods in the absence of taste heterogeneity. The parameter $f$ has been calibrated to produce 10 goods in group A and 10 goods in group B in the equilibrium of the "baseline" scenario with no income inequality and no heterogeneity of taste. Parameter $a$ has been chosen to achieve a 10% equilibrium growth in the baseline scenario. The value of $\beta$ is 0.98.

In the first baseline scenario, the income of all consumers is equal to unity and each consumer values each good equally. As mentioned, the model has been calibrated to predict the entry of 20 goods - 10 in category A and 10 in category B. Because all goods are equally valued, the spending on each good will constitute 5% of each consumer's income. For the baseline scenario, the model predicts the elasticity of demand with respect to price, $\epsilon = 1.4679$ and 10% equilibrium growth of quality for each good.

Next, I introduce a moderate income inequality setting the income of the rich consumers to 1.2 and the income of the poor consumers to 0.8. Since we wish to explore what happens if the increase in income inequality is accompanied by an increase in heterogeneity of taste, I assume that the richer consumers value goods in category A two times more than goods in category B, while the poorer consumers value goods in category B two times more than goods in category A. The model predicts that in equilibrium, the total number of available goods in category A is going to rise to 16, while the number of goods in category B will drop to 4. Rich consumers will spend 5.4% of their income for each good in A and 3.5% for each good in B. The poor consumer will spend 4.5% for each good in A and 6.9% for each good in B. The elasticity of demand for goods in A is almost exactly the same as in the baseline scenario: $\epsilon = 1.4676$. For a good in B, the drop of elasticity relative to the baseline is more visible: $\epsilon = 1.4658$. As a result, the growth of quality for goods in A is 9.9% and for B good is, 9.5%.

The results are substantially different if income inequality and taste heterogeneity are further increased: suppose now that the income of the rich consumer is 1.4 and the income of poor consumer is 0.6. The rich consumer values good A 9 times more than good B, while the poor consumer values good B nine times more than good A. In equilibrium, the model predicts that this implies a spending of 5.4% of income on A goods and 1.7% on B goods by rich consumers, and 4.2% of income on A goods and 12% of income on B goods by poor consumers. The number of goods in categories A and B are 18 and 2, respectively. The elasticity of demand for goods in category A is $\epsilon = 1.4676$, while the elasticity of demand for goods in category B is $\epsilon = 1.4461$. As a consequence, the quality growth of good A is 9.9% - only marginally lower than in the baseline, while the qulity growth of good B is 4.9% - i.e. half of that in the baseline.

The dependence of growth on inequality increases for a larger value of parameter $\rho$. For example, if $\rho = 0.5$, an increase of inequality from $y_r = y_p = 1$ to $y_r = 1.2$ and $y_p = 0.8$ which is accompanied by an increase in dispersion in relative valuation of goods to 2:1 for rich consumers and 1:2 for poor consumers reduces A good's quality growth to 8.4% and B good's growth to 7.8%. If we increase inequality further, to $y_r = 1.4$ and $y_p = 0.6$ and simultaneously allow for an increase in dispersion in relative valuation of goods to 9:1 for rich consumers and 1:9 for poor consumers, the quality growth for good A is 8.6%, while good B is not imroved at all.

## 2.6 R&D Uncertainy and Endgenous Growth

### 2.6.1 R&D Uncertainty and Optimal Product Diversity

One of the key contributions of Young's endgeonous growth model is an observation that a high number of variety limits the rate of technological progress becuase research effort needs to be partitioned to increase the quality of numerous goods. The more carriages there are, the more horses per carriage are needed to keep a cavalcade moving. This leads to the question of weather it would be optimal to ban variety and concentrate all research effort on development of one good, thereby achieving growth. In the origrinal paper by Young, sustaining variety is necessary because all goods have some degree of complimentarity and each consumer has a positive demand for each good, even if its price is very high. Section 2.4, and in particular subsection 2.4.4, suggests that even if goods are perfect substitutes, variety is valuable because one good is highly valued only by a narrow part of the population and other consumers may have high valuations of other goods. These goods must be developed also. The third reason to sustain variety - presented in this subsection - is the uncertainty of the R&D process.

Consider a simplified version of Young's model: there is only one period and R&D process require investment in terms of final output. Since I want to focus on optimal product diversity, I present a solution to the social planner problem. The planner allocates an inelastically supplied resource (e.g. labour) across production sides of various goods and chooses a number of varieties and R&D investments in order to maximize consumption. The outcome of the R&D process is uncertain: R&D spending of $F(\lambda) = fe^{a\lambda}$ produces quality level $\lambda\theta$, where $\theta$ is a random variable. The problem may be therefore stated as:

$$\max_{N,\lambda,L_i} \left\{ E\left[ \left( \sum_{i=1}^{N} (\lambda\theta_i L_i)^{\rho} \right)^{\frac{1}{\rho}} \right] - N * fe^{a\lambda} \right\}$$

subject to $\sum L_i = L$. The planner chooses the number of varieties and R&D investment before realization of $\theta$'s. The decision on the allocation of labour may be taken after the realization.

Consider the case of the perfect substitute: $\rho = 1$. The problem then becomes:

$$\max_{N,\lambda} \left\{ \lambda L E \left[ \max \left\{ \theta_i \right\} \right] - N * f e^{a\lambda} \right\}$$

If the uncertainty parameter is distributed with the Frechet Distribution: $\theta_i \backsim Frechet\,(\alpha, s, 0)$ with $\alpha > 1$ [16], then $\max\{\theta_i\} \backsim Frechet\left(\alpha, N^{\frac{1}{\alpha}} s, 0\right)$ and $E\left[\max\left\{\theta_i\right\}\right] = N^{\frac{1}{\alpha}} s \Gamma\left(1 - \frac{1}{\alpha}\right)$. Hence the planner's problem simplifies further:

$$\max_{N,\lambda} \left\{ \lambda L N^{\frac{1}{\alpha}} s \Gamma \left(1 - \frac{1}{\alpha}\right) - N * f e^{a\lambda} \right\}$$

Clearly, although additional varieties involve more R&D effort, they also bring a benefit: every additional brand may turn out to be more valuable than the previous set of varieties. To relate this to a real life situation, we may consider renewable sources of energy. Although investing in photovoltaic panels, wind turbines and hydroelectric power require three times more resources than focusing solely on the wind turbines, it may beneficial to sustain the portfolio, as there is a positive probability that investment in solar power turns out to be unexpectedly fruitful. Furthermore, the benefit from additional varieties depends on the dispersion of the $\theta$ distribution, which (for constant $\alpha$) is proportional to the parameter $s$.

The first order conditions to the planner's problem are:

$$\frac{1}{\alpha} \lambda L N^{\frac{1-\alpha}{\alpha}} s \Gamma \left(1 - \frac{1}{\alpha}\right) = f e^{a\lambda}$$

$$L N^{\frac{1}{\alpha}} s \Gamma \left(1 - \frac{1}{\alpha}\right) = N f a e^{a\lambda}$$

Combining the two conditions:

[16] the lower limit is set to ensure finite moments of the distrbution and to ensure that the second order conditions are satisfied.

$$\lambda = \frac{\alpha}{a}$$

$$N = \left( \frac{\lambda L \Gamma\left(1 - \frac{1}{\alpha}\right)}{\alpha f e^{a\lambda}} s \right)^{\frac{\alpha - 1}{\alpha}}$$

Note that for this simple model and for the Frechet distribution of $\theta$, optimal R&D investment does not depend on dispersion of $\theta$ (again assuming, $\alpha$ is constant). However, the optimal product differentation is a positive function of the dispersion.

### 2.6.2 R&D Uncertainty and Decentralized Equilibrium

In this subsection we consider the model from section 2.4 except that now consumers tastes are homogenous and the outcome of an R&D process is uncertain. In particular R&D spending of $F(\lambda)) = fe^{a}\lambda$ brings quality improvement by a factor $\theta\lambda$ where $\theta$ is a random variable with $E(\theta) = 1$. If firm needs to choose price before realization of R&D outcome then its problem is

$$\max_{p_i, \lambda_i} (p_i - c_i) E[Q_i] - F(\lambda_i)$$

subject to

$$E(Q_i) = \int \ldots \int \frac{\left(\theta_{ij}\lambda_j / p_j\right)^{\frac{\rho}{1-\rho}}}{\sum_k \left(\theta_{ik}\lambda_k / p_k\right)^{\frac{\rho}{1-\rho}}} p_j^{-1} y_i g\left(\underline{\theta}\right) d\underline{\theta}$$

This problem is identical to the problem of the monopolist in the section 2.4. The results must be identical too:

$$\frac{\lambda_t}{\lambda_{t-1}} = \frac{1}{a} \frac{\rho}{1-\rho} \left(1 - \mathrm{D}(\psi) - \frac{1}{N}\right) \equiv \epsilon - 1$$

An increase in uncertainty - captured by an increase in dispersion of possible realizations of $\theta$ - results in the reduction of R&D spending and lower growth. Notice that this result follows for risk-neutral firms. The intuition is similar to the intuition of the taste heterogneity model: under uncertainty, there is a positive probability that a firm's competitors may be (relatively) very succesful or very unfortunate. In either of these two cases, R&D effort performed by the firm will be irrelevant: in the first case, even a large quality improvement will not be enough to steal a market from the successful competitor; in the second case, the firm will capture the market, even with low R&D effort.

## 2.7 Conclusion

Although Chamberlin's proposition that differences in taste between consumers can be a source of desire for products variety has been extensively discussed in numerous microeconomic models (e.g. Salop (1979), Hoteling (1929), Perloff and Salop (1985), de Palma et al. (1985)), this paper is the first study that formalizes Chamberlin's proposition in direct reference to the setup of the Dixit and Stiglitz model. In particular, it demonstrates that in symmetric equilibrium - whenever it exists - elasticity of substituion, which governs optimal product differentiation in the Dixit and Stiglitz model, can be shown to be a decreasing function of consumer taste dispersion. More generally, for any kind of equilibrium, elasticity of demand for a good can be expressed as a dereasing function of dispersion (across consumers) of income shares devoted for this good.

The result amends interpretation of numerous papers that are based on the Dixit and Stiglitz framework. For example, in Young's model of endogenous growth, elasticity of demand determines equilibrium innovation effort and the rate of technological progress. Together with result derived in section 2.3, this produces a negative relation between economic growth and taste heterogeneity. In section 2.5, the model is extended to include income inequality. Simulation results reveal that if income is correlated with taste, higher income inequality is associated with lower quality improvement for every good. However, the drop of quality growth that follows mean preserving inequality increase is significantly larger for those goods which are favoured by poorer consumers. This result is in line with the intuition of the directed technological change hypothesis (Acemoglu 2002, 2007) that R&D investment in productivity of a good depends on the value of its market. Application of the directed technological change hypothesis to evaluate quality growth of goods devoted for poor vs. quality growth of goods devoted for rich consumers has not been studied so far and appears to be a promising path for future research.

The set up in section 2.3 permits reinterpretation of heterogeneity in consumers' tastes as uncertainty about product quality, e.g. prior to the realization of R&D. In this context, the model may give predictions about determinants of the optimal R&D portfolio: higher uncertainty about R&D process gives incentive to diversify the portfolio, i.e. invest in a higher number of varieties. The result occurs, despite the risk neutrality of firms.

# 3 Chapter III: Catching up in Central and Eastern Europe

## 3.1 Introduction

At the beginning of the 90s, when it became clear that a group of economically backward countries from Central and Eastern Europe will strive to join the European Union, allow trade and investment from abroad and adopt stable institutions, many economists anticipated a textbook example of economic convergence, similar to the one between Southwestern and Northwestern European countries that Europe witnessed in the 60s. A number of economists rushed to form predictions about the speed of convergence and warn of potential problems. (Sachs (1991), Fisher, Sahay, and Vegh (1998a, 1998b), Boldrin and Canova (2003) and Caselli and Tenreyro (2005)). 15 years after the economies embarked on the convergence path seems to be a good starting point to examine where the countries stand, how did they go so far and what they can expect in the future. For economists, this is an interesting opportunity to confront their expectations with true outcomes, as well as to evaluate the predictions of various models such as the Solowian convergence model (Solow (1956)) or technology imitation models (Nelson and Phelps (1966), Barro and Sala-i-Martin (1997), Howitt (2000)). For the policy makers, it might reveal what sources of conveg-ence are already exploited and what sources still remain potential (possibly still not activated) engines of growth.

At the same time, today might be the last chance to perform such an analysis: the economic crisis that began in 2008 put many of the Central and Eastern European economies out of the equilibrium convergence path. It might be that economists will need to wait a long time before the effect of (at the end exogenous) the shock diminishes and the convergence again could be analysed along the standard growth models.

The growth accounting for Central and Eastern European regions was performed earlier in the World Bank Report on Productivity Growth in Eastern Europe and Former Soviet Block (2008). However, that analysis is at the regional level and does not evaluate the growth decomposition for particular countries. Growth accounting for Poland, Czech Republic, Hungary, Slovakia and Slovenia is explored in Doyle, Kuijs and Jiang's (2001) analysis that only covers growth until 1999. Iradian (2007) presents growth accounting for transition economies, but only covers former Soviet Union republics. The similar growth accounting exercise, which also includes the effect of labour composition on growth, has been presented in Gradzewicz, Growiec, Kolasa, Postek and Strzelecki (2014). However, the central focus of their paper is to study the drivers of Polish relative success during the European crisis. The aim of my study is the illustration of forces driving the 12-year growth in Central and Easter European Economies relative to the growth in German economy. Two articles, Baran (2013) and Growiec (2008) shows the results on the convergence of central and eastern european growth using the data envelopment analysis, which is an interesting alternative to the approach followed in this paper.

The strategy for this paper is to first identify various theoretical sources of convergence (e.g. capital differences between Poland and Germany), then, to examine to what extent these sources were exploited (e.g. by looking at the growth rates of capital between 1995 and 2005) and finally to assess the potency of the source for the future (e.g. by comparing levels of capital in 2005).

The theoretical sources of convergence are identified along two decompositions. The first one decomposes growth into growth of labour input (total hours worked), physical capital and a term that combines human capital and TFP growth. To further decompose the growth of this last term, I propose a new method that takes advantage of the availability of the data on relative wages of workers with different educational levels. This method allows me to separate the effect of increased shares of well educated workers and the effect of higher productivity of more abundant educational groups. Furthermore, if workers with different education levels are not perfect substitutes, I might also isolate the negative effect of one type of workers becoming very scarce.

The second decomposition uses sector level data to distinguish whether growth comes from growth of productivities of industries or from moving labour towards more productive industries.

## 3.2 Data

The data that is necessary for the first decomposition comes from two sources. Real GDP (2005 prices in international dollars - after PPP conversion) data is taken from Penn World Tables, and the number of hours worked and the total number of persons engaged comes from the EU KLEMS dataset. Persons engaged includes both employees and the self-employed.

The choice of the target period is debatable. The key data is available in national statistics agencies for all countries until 2012. However, I decided to leave 2007 as the ending date. The reason for this is that the 2007 economic crisis put many of the Central and Eastern European economies out of the equilibrium convergence path.

The most problematic task was the computation of capital levels for the three countries. The standard method to obtain time series for capital is the perpetual inventory method: given the initial level of capital, each capital level is constructed by taking the previous capital level, adding investment and subtracting depreciation - normally assumed to be 6% of the previous capital level (Barro, Mankiw and Sala-I-Martin (1992), Nehru and Dhareshwar (1993), Caselli and

Tenreyro (2005), Iradian (2007)). The investment series in Penn World Tables is given at the PPP coverted 2005 constant prices. The initial capital level can be easily computed if it is assumed that the economy is in the long run equilibrium and growth of capital is equal to the growth of output. Then initial capital is given by[17]

$$K_0 = \frac{I_0}{g + \delta}$$

where $g$ is the growth rate of output and $\delta$ is the depreciation rate.

Regarding Hungary and Poland, I compute $g$ by taking the average across the annual growth rates in years 1970-1974, and following Iradiam (2007), I center the initial level of capital in the middle of this period, which is 1972. The depreciation rate in this period is assumed to be 0.06.

Is the assumption about equal growth rates of capital and ouput reasonable? Was the economy in 1970 already on the balanced growth path after the war? The communist central planner did push for the rapid physical capital accumulation from the early post-war period, which might suggest that by the early 70s, the steady state of capital was restored. On the other hand, anecdotal evidence says that the 70s included construction of impressive factories (the example is the Katowice steelworks project), whose purpose was capital formation far beyond replacing depreciated capital and keeping pace with TFP growth. As a result, the growth rate of capital might be biased downward (and so the initial level of capital would be biased upward). However, this problem should not alter the results since the effect of misscalculation of initial capital level diminishes over time, and the first capital observation I used is in 1995 - a quarter of a century after the estimation of the 1970 capital level.

A more serious problem is associated with the assumption of 6% depreciation. It is very likely that, especially at the beginning of the 90s, depreciation

[17] This follows from: $K_1 = K_0 (1 - \delta) + I_0$. Rearranging, $(-\delta) + \frac{I_0}{K_0} = \frac{K_1 - K_0}{K_0} = g$

levels were way above the standard level observed in other countries. In the Polish case, since 1991 the OECD has provided data on gross consumption of capital (which indeed turns out to be extraordinarily high until 1994-1995). Thus, since 1991, instead of deducting the previous capital level multiplied by the 6% depreciation rate, I deduct the gross consumption of capital from the OECD dataset. In the Hungarian case, the data on capital consumption is available only from 1995. Since, bearing in mind the Polish case, 6% depreciation between 1991-1994 appears to be over-optimistic, I assume that in this period the consumption of capital is at the level of the average consumpton in the period of 1995-1999 (9%). [18]

In the case of the Czech Republic, the procedure to compute capital levels was similar. However, due to limited availability of data pre-1990, it required a series of additional assumptions. I construct a proxy for GDP in the period of 1983-1990 by assuming that the yearly growth rates in Czech Republic were the same as in Poland in the period of 1982-1989 (one year lag is taken because the recession in Czech Republic started one year later). I also assume that the investment to GDP ratio in this period was the same as in Poland. Using these approximated series, I compute the initial level of capital for 1985 using the same procedure as before for Hungary and Poland. Until 1990, I assume the depreciation at the level of 6% per year. For the period of 1991-1992, I assume that the consumption of capital (as % of GDP) is at the level of average consumption in the period of 1993-1997 (19%).

As mentioned earlier, to compute the effect of skill upgrading, I use the data on relative wages of highly skilled, medium-skilled and low-skilled workers. The data on hourly wages for these three groups is again taken from EU KLEMS. The dataset labeled as highly skilled workers as those who have received university

[18]Since in Poland depreciation between 1995-1999 was lower then between 1991-1994, such approximation might be still over-optimistic

degrees, medium-skilled workers as those with high school degrees and low-skilled workers as the remaining group. In this paper, I combine medium-skilled and highly skilled workers to form the group that from now on I will label as high-skilled workers. This is based on the presumption that a high school or vocational degree indicates skills that go beyond basic literacy and simple arithmetic skills. Thus those with at least a high school degree might be used for very different tasks than those without a high school degree.

The two key variables for the second decomposition (along sectors) are the value added and the number of hours by people engaged per each sector. These data is available in the EU KLEMS dataset. Since value added is given in nominal terms, I deflate each sector value added by this sector price index (taking 1995 as a base; value added price indices for each sector are also supplied by the EU KLEMS dataset). Then I weight each observation bythe PPP international dollars conversion factor in 1995 (the conversion factor is taken from stats.OECD.org dataset).

## 3.3 Factor Accumulation and TFP Growth Decomposition

I start with the most standard decomposition of growth into growth of physical capital, labour input and the residual (capturing, among others, human capital and technological progress). This decomposition is based on the Cobb-Douglas production function. It follows the idea that growth of output might be broken into growth of factors weighted by their income shares:

$$\frac{y_{t+1} - y_t}{y_t} = (1-\alpha)\left(\frac{A_{t+1} - A_t}{A_t}\right) + (1-\alpha)\left(\frac{L_{t+1} - L_t}{L_t}\right) + \alpha\left(\frac{k_{t+1} - k_t}{k_t}\right) \tag{36}$$

The labour share of income is assumed to be 0.33 - as normally assumed in

| | Poland | Hungary | Czech |
|---|---|---|---|
| growth of GDP | 72% | 56% | 52% |
| growth of hours per person engaged | -2% | -3% | -3% |
| growth of number of persons engaged | -2% | 8% | 1% |
| growth of capital | 52% | 37% | 75% |
| residual growth | 90% | 58% | 44% |

Table 5: Engines of Growth

the literature.

This first decomposition indicates the tremendous increase in the TFP in Poland. The productivity has increased by 90% over the twelve-year period (on average 5.5% increase per annum). The capital during this period did not keep pace with the productivity improvement and has risen only by 52% (on average 3.5% per annum). The labour input in turn has fallen: the total number hours worked by persons engaged in 2005 was 96% of the level from 1995. It might be also interesting to decompose the fall in number of hours worked into changes in total number of persons engaged and in number of hours per person engaged. It turns out that both figures have fallen: number of persons engaged has decreased by 1.8%, and the number of hours per person engaged decreased by 2.1%.

In Hungary, the pattern of growth, resembling the Polish case, features very high growth of TFP (58% increase over the twelve-year period or 3.9% on average per annum) and capital growth that did not keep pace with productivity (over twelve years, capital grew by 37%, i.e. 2.7% on average per annum [19]). The amount of labour has increased marginally (in total by 5% over the period). In fact, the number of hours per person engaged has decreased by 2.6%; however, this was compensated by the 7.8% increase in the number of persons engaged.

Czech Republic witnessed the most unbalanced growth: while the TFP has

[19] Between 1995 and 1997, due to the very high depreciation, the capital actually fell by 8%. Afterwards, depreciation remained high (until 2002); however, high investment rates led to a growth of capital at the average rate of 3.4% per annum.

increased by 72% over 12 years, capital has increased only by 22%. It must be noted, however, that the estimates of capital accumulation and TFP are subject to substantial measurement error since the data on GDP and investment are available only from the beginning of the 90s. As in other countries, the number of hours per person engaged has decreased marginally by 3.3%. The total number of persons engaged has increased by 1%. The amount of labour has decreased by 2.2%

To decompose convergence in GDP to Germany, I use the same strategy, except now each variable is replaced with its value relative to Germany (thus e.g. $y$ becomes Polish GDP divided by German GDP). Since labour input did not change much in any country, I focus on the values per hour of person engaged (p.h.p.e.). In Poland, the relative GDP p.h.p.e. has increased by 47% (from 0.33 to 0.49). The relative capital p.h.p.e. has increased by 27% (from 0.18 to 0.23), and the relative residual has increased by 58% (from 0.45 to 0.72). Translating it into contributions in the convergence of output (by weighting capital by its share in income and TFP by labour share in income), the 47% increase in relative output may be decomposed into a 39% increase due to residual growth and a 9% increase due to capital accumulation. This is in line with the conclusions from the previous paragraph that the Polish catching up to German economy cannot be explained by capital accumulation.

Over the 12-year period, the ratio of Hungarian GDP p.h.p.e to German has increased by 23% (from 0.38 in 1995 to 0.46 in 2007), which - as in the Polish case - was primarily fueled by the increase in relative TFP (increase from 0.47 in 1995 to 0.61 in 2007) and, to a significantly lower extent, by growth in relative capital (from 0.25 to 0.26)[20].

Czech Republic had exactly the same starting point as Hungary - in 1995 its

[20] When interpreting a small increase of relative capital, I need to take into account the remarkable fall of the Hungarian capital in the first part of the period, as described in footnote 3.

GDP was 38% of German GDP. Compared to Hungary, it grew slightly faster to reach 49% of German GDP in 2007. Quite surprisingly - and contrary to the Solow model predictions - the capital relative to Germany has not changed: both in 1995 and in 2007, the ratio of capital in the Czech Republic to the capital in Germany was 0.31. Despite no growth, in 2007 the level of capital per hour of person engaged in the Czech Republic was highest among the three economies. Relative productivity has grown substantially - in line with the technology transfer hypothesis - from 0.43 to 061.

To explore the potency for future growth, I shall now focus on the levels of capital and residual in 2005. Given the results presented so far it is immediate that a large, and so far not properly exploited, reservoir of growth is the capital accumulation. Scholars who study the theory of economic growth will find it astonishing that while the TFP of one country is about three quarters than that of the other country, its capital per labour is three or five times smaller. The optimistic view would be that TFP growth was so rapid during these years that capital simply couldn't keep pace. This would imply that the coming years are yet to witness capital accumulation and the continuation of convergence is almost guaranteed. The pessimistic view is that there are some substantial barriers for capital accumulation that slowed down convergence and, unless resolved, will continue to slow down the economy in the future.

A factor that could have played a role was the international investors' uncertainty about political and economic stability in Poland, Hungary and Czech Republic in this period. This could have changed after the three countries joined EU. For this reason, I explore the growth of capital in the period of 2004-2007. In Poland, capital relative to Germany was growing on average at 2% per annum compared to 2% on average in the entire 12-year period. In Hungary, between 2004 and 2007, relative capital actually grew slower than in the previous period - 1% average growth compared to the average of 2% growth between 1995 and 2007. In Czech Republic, as in the entire period relative capital did not move. Thus either EU membership was not perceived by investors as warranty of stability or risk is not the key obstacle to capital accumulation.

The sluggish accumulation of capital compared to the rapid growth for productivities also has a consequence on evaluating how well the Solow model might explain the convergence between the three economies to the old Europe. In fact, this is devastating for the Solow model prediction: in 1995 in Poland, capital per efficiency unit (capital devided by total hours worked and productivity residual) was 2.14, compared to 5.47 in Germany. With the free flow of capital between countries, capital per efficiency units should be equalized in order to equalize the return from capital in both countries. If I abandon the assumption on free flow of capital, I might compute the steady state level of capital per efficiency unit from savings rate, $s$, depreciation, $\delta$, TFP growth, $g$, and labour input growth, $n$, using the steady state condition: $\frac{k}{AL} = \left(\frac{s}{\delta+g+n}\right)^{\frac{1}{1-\alpha}}$. The savings rate (investment over GDP) in Poland is given in the Penn World Tables, and its average value between 1995 and 2007 is 0.193, the average annual TFP growth in this period was 5.5%, the annual growth of hours of persons engaged was negative and equal to -0.3% and the depreciation rate is assumed to be 6%. Given these values, I can compute the steady state level of capital per efficiency unit in Poland to be 2.48, above the observed value of 2.14. Thus, over ten years it would be expected that the flow of capital into Poland (under its free flow assumption) or its internal accumulation (in case of close economy) moves capital per efficiency unit closer to the German level or its steady state value. Instead of moving this direction, the capital per efficiency unit in Poland actually decreased to 1.78 in 2007 (comparing to 5.68 in Germany).

The same trend is confirmed in case of Hungary and Czech Republic. In

Hungary in 1995, the observed capital per efficiency unit was 2.97, which is higher than in Poland but way below the Hungarian theoretical steady state level of 3.27 or the German level of 5.47. The theory would therefore predict an increase of capital per efficiency unit. Instead, its value has dropped over the 12-year period to 2.46 in 2007.

Czech Republic had the highest value of capital per efficiency unit, which was 3.95. Compared to the other countries in the region it was closest to the steady state - the theoretical steady state value of capital per efficiency unit was 4.38. However, over time, the economy has distanced from the steady state. In 2007, the capital per efficiency unit was 3.3

Obviously the failure of the Solowian prediction does not have to be driven by capital accumulation that is too slow but rather by the rapid growth of productivity. Nevertheless, this short exercise shows that it is difficult to evaluate the convergence of output using the Solowian model alone.

### 3.3.1 Human Capital and Technology Decomposition

Given the spectacular increase in the residual, it might be worth attempting to throw some additional explanatory power into the model. The data on the number of workers in each educational group and their wages might allow exploration of the extent in which growth comes from an increased share of more educated workers and if the technological progress was concentrated on the growth of productivity of high-skilled or low-skilled workers[21].

#### The Form and the Calibration of the Production Function

To accomodate the differences in the productivities of different types of workers, I will use a production function similar to that of Caselli and Coleman (2005)

[21] The idea to use data on relative factor prices in accounting for cross-country productivity differences was first suggested by Caselli and Coleman (2005).

and Ottaviano and Peri (2008):

$$y = \left( \left( A^h L^h \right)^\sigma + \left( A^l L^l \right)^\sigma \right)^{1-\alpha/\sigma} k^\alpha \tag{37}$$

where $L_l$ denotes the amount of high-skilled labour, $L_h$ is the amount of low-skilled labour, $A_l$ and $A_h$ are the productivity parameters and $\sigma$ is the parameter ruling the substitability between two types of workers. Isolating the total labour input:

$$y = \left( \left( A^h s^h \right)^\sigma + \left( A^l s^l \right)^\sigma \right)^{1-\alpha/\sigma} L^{1-\alpha} k^\alpha$$

where $s_h$ is the share of high-skilled workers in the total labour supply, $s_l$ is the share of low-skilled workers.

If I equate factor prices with their marginal products (as implied by representative firm first order conditions for optimal choice of inputs), I obtain the following condition:

$$\frac{w_h}{w_l} = \left( \frac{L_h}{L_l} \right)^{-(1-\sigma)} \left( \frac{A_h}{A_l} \right)^\sigma \tag{38}$$

Combining this with the production function, I can easily derive the expression for $A_h$ and $A_l$:

$$A_l = \frac{y^{\frac{1}{1-\alpha}} k^{\frac{-\alpha}{1-\alpha}}}{L_l} \left( \frac{w_l L_l}{w_h L_h + w_l L_l} \right)^{\frac{1}{\sigma}}$$

$$A_h = \frac{y^{\frac{1}{1-\alpha}} k^{\frac{-\alpha}{1-\alpha}}}{L_h} \left( \frac{w_h L_h}{w_h L_h + w_l L_l} \right)^{\frac{1}{\sigma}}$$

Since the calibration of $\sigma$ has a crucial impact on the results, I shall give it some attention. The choice is problematic. One of the first - and so far the most popular - estimates of $\sigma$ was delivered by Katz and Murphy (1992), who found that the value of elasticity of substitution (i.e. $\frac{1}{1-\sigma}$) is equal to 1.41. They obtain it from a simple regression derived from the first order conditions (38)

$$log\left(\frac{w_h}{w_l}\right) = -\left(1-\sigma\right) log\left(\frac{L_h}{L_l}\right) + \sigma log\left(\frac{A_h}{A_l}\right) \tag{39}$$

Katz and Murphy estimate this regression using US data between 1963 and 1997. To proxy for the last term they control for a linear trend.

However, a simple intuition questions the validity of the estimates. If relative labour supply responses to relative efficiency deviations from the linear trend, then the explanatory variable is correlated with the error term and the regression estimates are biased. Since it is expected that there will be a positive response of relative supply to positive relative efficiency shock, the coefficient on relative supply is likely to be biased upwards (towards zero) - intuitively, since higher relative efficiency and higher relative labour supply tend to go together, I cannot observe a negative response of wages to the higher relative labour supply. Thus, the bias is towards higher substitability across two types of workers.

Equation (39) has received a lot of attention in the labour literature, and various authors tried to find instruments that could help to avoid the endogeneity bias described above. One of the convincing IV estimates was presented by Ciccone and Peri (2005), who instrument relative labour supply with compulsory school attendance laws (at state level) and arrive to the value of 1.5. The time series they use covers the period of 1950-1990 in the US - similar to the one from Katz and Murphy. Thus, it seems that if there is an endogeneity bias in the US, it is not very significant.

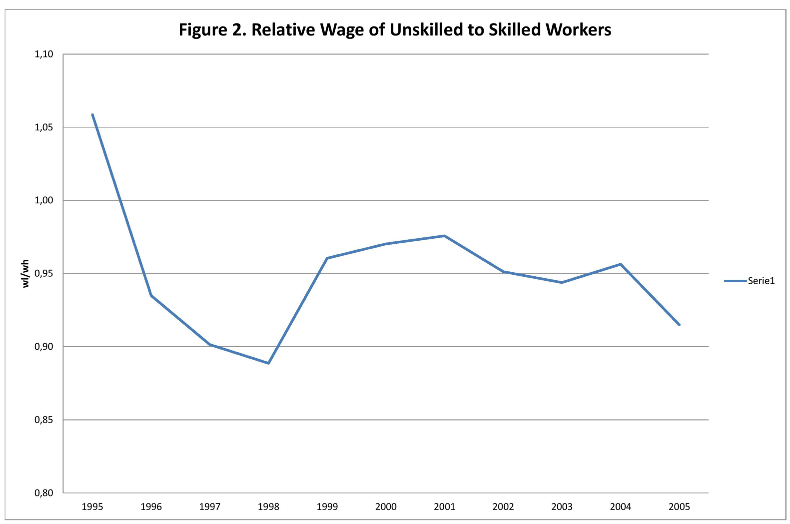
The reason why I describe the various estimates of $\sigma$ in so much detail is that my own estimates of sigma in Europe (both in Germany and Poland) are significantly different, even though my estimates for the US match the evidence from the Ciccone and Peri and Katz and Murphy papers. When I regress equation (39) using the EU KLEMS data for Poland between 1995 and 2005 (and as Katz and Murphy, control for linear trend), the coefficient on relative labour supply is very close to zero: 0.15 (and not statistically significicantly different from zero). Similarly, in Germany, the coefficient is very close to zero (with the value of -0.15). This implies a very high elasticity of substitution. Of course this might be driven by the endogeneity problem outlined above. But why has no bias in US data been seen (as there is no significant difference between my, Peri and Ciccone's and Katz and Murphy's estimates)?

Thus there are two options to choose from: either I assume that elasticity of substitution in European countries and in the US is similar, although this would imply that the responsiveness of relative labour supply to relative efficiency (that drives the bias) in Europe is much stronger than in US; or I assume that in Europe and the US the bias is similar (i.e. negligable) and thus there is a significant difference in elasticity of substitution.

One way to solve this problem is to look for an exogenous shock to the relative labour supply and observe the reaction of the relative wage. Luckily, in the middle of the period in Poland, a major educational reform was introduced that extended obligatory schooling by one year and the obligatory education was finished at the age of 16 rather than at the age of 15. As a direct result of this change, while the cohort born in 1987 could enter the market as low-skilled after summer in 2000, the cohort born in 1986 could enter the market as low-skilled only after summer 2002. This one year gap in the inflow of new low-skilled workers into the labour market should be reflected in the fall of the supply of the unskilled relative to the supply of skilled workers around this time. This time series is shown in Figure 1.

Indeed, the figure shows that after years of no change, the relative supply of unskilled to skilled workers drops dramatically in 2001. If I take into account that those graduating from primary school need around half of a year to find a job, I should find that while in 2001 there was a normal low-skilled labour inflow, in 2002 there was no inflow (maybe except for those who found a job immediately after graduating in the summer of 2002). In fact, it is evident that the decline has continued. This might be a result of the fact that the reform (that has also reduced the years of high school by one year) has encouraged students to continue education and thus acquire more skills.

What impact did this shock have on the relative wage of unskilled and skilled workers? If there is some degree of complimentarity between skilled and unskilled workers ($\sigma$ is different than zero), there should be an increase in the wage of unskilled workers, due to the relative scarcity of unskilled labour. Instead, Figure 2 shows that the effect was none. If anything, the relative wage experienced a small fall around 2002. The graph suggests that since 1999, relative wages have been experiencing a slow and rather stable decline (probably following skill-biased technological change) uninterrupted by any changes in relative supply of skills. This might serve as additional evidence that, at least in Poland, as suggested by the simple Katz-Murphy regression, the skilled and unskilled workers are characterized by high degrees of substitutability.


Below I present results for both cases: if I assume $\sigma = 1$ and if I assume $\sigma$ is at the US level (estimated by Katz and Murphy).

**Growth Decomposition Using Wages for Skilled and Unskilled Workers**

Given all the information on the production function form and its key parameters ($A_h$, $A_l$ and $\sigma$), I can recover additional information about the nature of convervence. Using equation (37), I immediately arrive to the simplest decomposition of growth rates:

$$\frac{y_{t+1}}{y_t} = \left( \frac{\left(\left(A^h_{t+1}s^h_{t+1}\right)^\sigma + \left(A^l_{t+1}s^l_{t+1}\right)^\sigma\right)^{1/\sigma}}{\left(\left(A^h_t s^h_t\right)^\sigma + \left(A^l_t s^l_t\right)^\sigma\right)^{1/\sigma}} \right)^{1-\alpha} \left(\frac{L_{t+1}}{L_t}\right)^{1-\alpha} \left(\frac{k_{t+1}}{k_t}\right)^{\alpha}$$

The comparison with equation (36) reveals that in the previous analysis I have captured the first term in the above expression as a residual. Now the additional information on $A$'s might help to decompose this residual into further elements. The first step is to decompose it in the manner similar to the way nominal GDP growth is decomposed into real GDP growth and inflation (by Paasche or Laspeyres decomposition):

$$\frac{y_{t+1}}{y_t} = \left( \frac{\left(\left(A^h_{t+1}s^h_{t+1}\right)^\sigma + \left(A^l_{t+1}s^l_{t+1}\right)^\sigma\right)^{1/\sigma}}{\left(\left(A^h_t s^h_{t+1}\right)^\sigma + \left(A^l_t s^l_{t+1}\right)^\sigma\right)^{1/\sigma}} \frac{\left(\left(A^h_t s^h_{t+1}\right)^\sigma + \left(A^l_t s^l_{t+1}\right)^\sigma\right)^{1/\sigma}}{\left(\left(A^h_t s^h_t\right)^\sigma + \left(A^l_t s^l_t\right)^\sigma\right)^{1/\sigma}} \right)^{1-\alpha} \left(\frac{L_{t+1}}{L_t}\right)^{1-\alpha} \left(\frac{k_{t+1}}{k_t}\right)^{\alpha} \tag{40}$$

This decomposition resembles a counterfactual exercise: the first term states what would be the growth of output if the efficiency parameters, $A_l$ and $A_h$ are allowed to jump to the new levels while keeping the shares of skilled and unskilled labour fixed across two periods. The second term instead shows what would be the change in ouput if efficiency parameters are fixed at their initial level but I let shares of skilled and unskilled workers change to their new levels.

The interpretation of these terms, however, is not straightforward (at least as long as $\sigma$ is not equal to one). One may be tempted to interpret the second term as the growth coming from the increased share of a more productive factor.

However, this is not exactly right. Consider what would happen if productivity parameters ($A$'s) of both factors (here, both types of labour) would be exactly equal. Then we would expect that the term should be one (no growth) irrespective of how the shares change. However, this will not be the case if $\sigma$ is not one.

The only way to obtain a term that would be robust to this problem is to split the second term in (40) along the following lines:

$$\frac{\left(\left(A_t^h \frac{s_{t+1}^h}{\left({s_{t+1}^h}^\sigma + {s_{t+1}^l}^\sigma\right)^{\frac{1}{\sigma}}}\right)^\sigma + \left(A_t^l \frac{s_{t+1}^l}{\left({s_{t+1}^h}^\sigma + {s_{t+1}^l}^\sigma\right)^{\frac{1}{\sigma}}}\right)^\sigma\right)^{\frac{1}{\sigma}}}{\left(\left(A_t^h \frac{s_t^h}{\left({s_t^h}^\sigma + {s_t^l}^\sigma\right)^{\frac{1}{\sigma}}}\right)^\sigma + \left(A_t^l \frac{s_t^l}{\left({s_t^h}^\sigma + {s_t^l}^\sigma\right)^{\frac{1}{\sigma}}}\right)^\sigma\right)^{\frac{1}{\sigma}}} * \frac{\left(\left(s_{t+1}^h\right)^\sigma + \left(s_{t+1}^l\right)^\sigma\right)^{\frac{1}{\sigma}}}{\left(\left(s_t^h\right)^\sigma + \left(s_t^l\right)^\sigma\right)^{\frac{1}{\sigma}}} \tag{41}$$

The first term in this expression is an indicator of how much output change comes from the increase in share of a more productive factor. The term will always take the value of one if factors have the same productivity - regardless of what will be the change of shares of these factors.

The second term is a residual, but in fact, it does have a clear interpretation. To understand it, note that if production function has the form given in (37), there will be some optimal shares of skilled and unskilled labour that will maximize output given other parameters. A change of shares will move the ouput towards or away from this optimum. If factors are perfect substitutes ($\sigma = 1$), moving resources towards more productive factor will always increase output. However, if inputs have some degree of complimentarity, setting one input to zero will always involve a loss for the economy, even if this factor has a lower productivity parameter, $A_l < A_h$. Thus the optimal output will involve shares that will be on one hand skewed towards the more efficient input, but on the other hand will not be too far from equal balance. The second term in the expression in (41) will exactly capture the cost (or benefit) from moving away (or towards) equally balanced shares. Note finally that if inputs are perfect substitutes, the term drops. Thus this decomposition is only important if it is believed that Polish, Hungarian, Czech and German elasticities of substitution between inputs are not infinite.

By an exactly analogous argumen, I might also decompose the first term in expression (40) into two terms:

$$\frac{\left(\left(\frac{A_{t+1}^h}{\left({A_{t+1}^h}^\sigma+{A_{t+1}^l}^\sigma\right)^{1/\sigma}}s_{t+1}^h\right)^\sigma+\left(\frac{A_{t+1}^l}{\left({A_{t+1}^h}^\sigma+{A_{t+1}^l}^\sigma\right)^{1/\sigma}}s_{t+1}^l\right)^\sigma\right)^{1/\sigma}}{\left(\left(\frac{A_t^h}{\left({A_t^h}^\sigma+{A_t^l}^\sigma\right)^{1/\sigma}}s_{t+1}^h\right)^\sigma+\left(\frac{A_t^l}{\left({A_t^h}^\sigma+{A_t^l}^\sigma\right)^{1/\sigma}}s_{t+1}^l\right)^\sigma\right)^{1/\sigma}}*\frac{\left(\left(A_{t+1}^h\right)^\sigma+\left(A_{t+1}^l\right)^\sigma\right)^{1/\sigma}}{\left(\left(A_t^h\right)^\sigma+\left(A_t^l\right)^\sigma\right)^{1/\sigma}} \tag{42}$$

Then I might intepret the first term in the above expression as a growth that originates from the fact that more abundant factor becomes relatively more productive. The second term will capture two effects: the first is the overall productivity growth, and the second captures that the change in relative productivity might help to balance inputs (or rather their efficiency equivalents).

#### Decomposition Results for Poland, Hungary and Czech Republic

Since EU KLEMS data on wages for various types of labour is available only until 2005, in this section, the analysis is focused on the convergence between 1995 and 2005. First, I proceed with the results for the value of $\sigma = 0.3$ (corresponding to $\frac{1}{1-\sigma} = 1.4$, as estimated by Katz and Murphy). Recall that the residual growth between 1995 and 2005 in Poland (earlier labeled as TFP) was equal to 67%. This might be decomposed as follows. The first term in expression (41) (growth due to higher share of a more productive factor, in this case - increasing shares of high skilled workers) implies a growth of 7.0%, a rather modest contribution.

Elasticity of Substituion = 1.42

| | Poland | Hungary | Czech |
|---|---|---|---|
| (1) Skills upgrading | 6.98% | 8.51% | 13.21% |
| (2) Skills balance | -6.84% | -6.60% | -10.77% |
| (3) Favourable technology bias | 8.20% | 7.33% | 6.12% |
| (4) Overall | 55.39% | 48.10% | 19.47% |
| TOTAL TFP growth | 67.56% | 61.11% | 28.06% |

Elasticity of Substitution = $\infty$

| | | | |
|---|---|---|---|
| (1) Skills upgrading | 0.07% | 1.72% | 1.37% |
| (2) Skills balance | - | - | - |
| (3) Favourable technology bias | 2.66% | 3.78% | 0.95% |
| (4) Skills neutral tech. change | 63.12% | 52.63% | 25.14% |
| TOTAL TFP growth | 67.56% | 61.11% | 28.06% |

Table 6: Growth and Skills Mix

The second term in (41) (growth due to better balance between factors) was in fact negative (unskilled labour in Poland is scarce, thus reducing its share leads to more imbalances and loss of output) and implies a decline of 6.8%. The first term in expression (42) (growth due to higher productivity of a more abundant factor) implies a growth of 8.2%. Finally, the last term in (42) that capture the overall improvement in productivities (weighting the improvement in technology for high-skilled and low-skilled equally) brings a growth of 55%. The key conclusion from this decomposition is that skill upgrading played some role in TFP growth, but this role was far from being crucial.

If I take the value of $\sigma$ to be a unity, the role of skill upgrading diminishes to zero: the first term in (41) is 1.0007. Also, the growth due to higher productivity of a more abundant factor is very modest - the first term in expression (42) implies a growth of 2.7%. The second component in (41) is obviously a unity. The entire remaining growth is attributed to the residual - overall increase in productivity.

Skill upgrading in Poland did not play a significant role in growth, regardless of whether I assume that workers with different skill levels are perfect substitutes. In the case of imperfect substitution, the reason for this is that the positive effect of skill upgrading is offset by the fact that I reduce the share of the already scarce factor. If in turn I assume that two types of labour are perfect substitutes, it turns out that productivity differences between two types of workers are not large (in 1995 $A_l = 16.1$ and $A_h = 17.6$)

Under the assumption of a 1.42 value of the elasticity of substitution, in Hungary, the skill upgrading could be associated with a growth of 8.51% - thus it played a more significant role than it did in Poland. However, as in the Polish case, the positive effect of an increased share of productive factor was almost entirely offset by a negative effect of increasing the disbalance between skilled and unskilled workers. The favourable technology bias - i.e. increase in relative productivity towards a more abundant factor - can be associated with a growth of 7.33%. The most important contributor was the change in the overall productivity change index: it can explain the 48.1% growth.

If skilled and unskilled workers are perfect substitutes, the change in skills composition lead only to a very modest growth. In constrast to the Polish economy, the wage difference between skilled and unskilled workers was substantial (almost two fold). The small contribution of skills recomposition is explained by the small size of skills accumulation (number of unskilled workers dropped from 18.9 to 14.5% over the ten years).

In Czech Republic, skills upgrading was rather low - unskilled workers in 1995 constituted 9.5% of the workforce; ten years later it was 6.2%. If I consider the finite elasticity of susbtitution and if I isolate the negative effect coming from the higher imbalance between skilled and unskilled workers in CES production function, skills upgrading ends up being one of the major forces in productivity growth: out of 28% growth of the residual computed in section 3.2, skills upgrading was associated with 13% growth. However, reducing the number of unskilled workers from 9.5% to 6.2% also had a large negative consequence on prodcution - at least as long as a finite elasticity of substitution is assumed. 6

Under the perfect substitution assumption, neither skills upgrading nor favourable skill bias could be associated with significant contribution. Again, the negligible role of skills accumulation can be explained with the small size of this process: the share of unskilled workers dropped by only 3%.

## 3.4 The Sectoral Decomposition

Following the strategy of Caselli and Tenreyro (2005), I can use sectoral data to find whether growth comes from growth of productivities of industries or from moving labour towards more productive industries.

Let $y^i_{jt}$ be the productivity of labour (output devided by total number of hours worked by persons engaged) in country $i$, in sector $j$ at time $t$, and let $a^i_{jt}$ be the share of employment (number of hours worked in the sector devided by total number of hours worked in entire economy) for sector $j$ in country $i$, at time $t$. Then productivity of labour might be expressed as the weighted sum of sectoral productivities:

$$y^i_t = \sum_j a^i_{jt} y^i_{jt} \tag{43}$$

I am interested in the change of Polish labour productivity relative to German labour productivity: $\Delta \frac{y^P_t}{y^G_t}$. Using equation (43), after some algebraic manipulations, I can decompose the change in relative productivities into three terms:

$$\Delta \frac{y^P_t}{y^G_t} = \sum_j \overline{a^P_{jt}} \Delta \left( \frac{y^P_{jt} - y^G_{jt}}{y^G_t} \right) +$$

$$+ \sum_j \overline{\left( \frac{y^P_{jt}}{y^G_t} \right)} \Delta a^P_{jt} - \sum_j \overline{\left( \frac{y^G_{jt}}{y^G_t} \right)} \Delta a^G_{jt}$$

$$+\sum_{j}\left(\overline{a_{jt}^{P}}-\overline{a_{jt}^{G}}\right)\Delta\left(\frac{y_{jt}^{G}}{y_{t}^{G}}\right) \tag{44}$$

where $\Delta x_{jt} = x_{jt} - x_{jt-1}$ and $x_{jt} = \frac{x_{jt}+x_{jt-1}}{2}$.

The term in first line captures the growth of relative productivity coming from growth within each sector. The sectoral growth rates are weighted by the sector's average employment share. The first term in the second line is the growth that comes from shifting labour towards more productive sectors. Because I am interested in the growth of relative productivity, the corresponding growth on the German side is deducted. Finally, the last term originates from the fact that the relative growth might also follow if the sector in which Poland has a disproportionally higher share of employment (compared to Germany) experiences extraordinary productivity growth.

The economy-wide increase in relative productivity of labour was 0.118 (from 0.341 in 1995 to 0.459 in 2006). This comes, to a large extent from the first component - the within-industry productivity growth lead to a 0.087 decline in relative productivity differences. The biggest contribution in this figure was manufacturing (0.036), wholesale (0.018) and financial intermediation (0.017) and real estate, renting and business activities (0.023) (figures in brackets are declines in sectoral relative productivity differences already weighted by average employment shares of these sectors).

The shift of labour towards more productive sectors in Poland had a very modest contribution of 0.008. Interestingly, at least until 2006, there is no evidence of a shift away from agriculture, where productivity - in comparison to German economy - is the lowest. In fact, regarding growth rooted in shifts of labour, Poland was outpaced by Germany that by moving labour towards better sectors has increased the productivity distance to Poland by 0.059. This result seems to be driven by the astonishing effect of moving labour towards real

| | Poland | Hungary | Czech |
|---|---|---|---|
| Productivity relative to Germany (1995) | 0.341 | 0.381 | 0.400 |
| Productivity relative to Germany (2006/2007*) | 0.459 | 0.477 | 0.489 |
| Change in relative productivities | 0.118 | 0.095 | 0.089 |
| Decomposition: | | | |
| Weighted growth of individual industries: | 0.089 | 0.074 | 0.107 |
| Shift of Labour to more productive industries | | | |
| in the country: | 0.008 | 0.020 | 0.005 |
| in Germany: | 0.059 | 0.063 | 0.063 |
| Favourable Technological Change: | 0.080 | 0.063 | 0.040 |

Table 7: The Sectoral Decomposition of Growth in Relative Productivity. *2006 for Poland, 2007 for Hungary and Czech Republic.

estate, renting and business activities sector in Germany (this alone increases the distance to Poland by 0.110 points).

Finally, Poland has benefited substantially from the growth of productivities in sectors in which Poland had disproportionally higher shares of employment (compared to Germany) - the last component has decreased Polish distance to Germany by 0.079 points. This is mainly due to the increase in the productivity of agriculture, where Poland has an abundant workforce and real estate, renting and business activities sectors. The latter sector had a 5% bigger share of labour in Germany than in Poland. Thus a drop of productivity in this sector (relative to the entire economy) has contributed to the shortening of the distance between Germany and Poland.

In Hungary, the productvity relative to Germany has increased by 0.095 - from 0.381 to 0.477. The decomposition attributes the largest role in the increase to growth of relative productivities in individual industries. The drop in the productivity gap between Germany and Hungary has been particularly visible in wholesale and retail trade (contributed 0.010 in closing the gap between average productivities), financial itermediation (contributed 0.019), real estate, renting and business activities (contributed 0.017) and education (contributed 0.02).

The shift of labour to a more productive sector led to more substantial contribution in growth of relative productivities than in the case of Poland. The second term in equation (44) takes the value of 0.020. The substantial role was played by the real estate, renting and business activities sector, which grew from 3.8% of the labour force in 1995 to 7.3% in 2007. However, as noted above, a similar shift towards that sector was present in Germany. As a result, shifts of labour between sectors in both countries resulted ultimately in increase of distance between Germany and Hungary.

Throughout the period, the portion of labour working in real estate, renting and business activities in Germany was higher than in Hungary. Since over the 12 years the productivity of this sector relative to other sectors was decreasing, the German benefit over Hungary was shrinking. In fact, this force has contributed 0.044 points to the increase of the Hungarian-German productivity ratio.

The experience of Czech Republic is similar to that of Poland and Hungary: Relative productivity of Czech Republic has increased by 8.9%: while in 1995, Czech productivity was exactly 40% of Germany, and in 2007 it was almost 49%. The weighted growth of individual industries alone could actually lead to growth by 10.7%; however, it was offset by the shifts of labour to a more productive sector in Germany - as described below. The fastest growing sectors were total manufacturing (contributed 1.5% in the 10.7% distance reduction), wholesale and retail trade (contributed 4.3%) and real estate and business activities (5.1%).

As in Poland shift of labour towards more productive sectors brought only a marginal benefit. The exception was a shift of labour towards the real estate and business activites sector. Since I could observe an even larger shift of labour towards this sector in Germany, altogether the shifts of labour between sectors has led to an increase in distance between Germany and Czech Republic.

Finally, Czech Republic could benefit from changes in the relative productivities of various sectors. In particular, since a larger share of the labour force works in the manufacturing in Czech Republic than in Germany, an increase of relative productivity of that sector helped to increase the distance to Germany.

## 3.5 Summary

In the first part of the paper, I decompose economic growth in Poland, Hungary and Czech republic into three elements: growth of capital, growth of labour supply and the total factor productivity term that captures the increase of human capital, transfer of technology and institutional change. Since in 1995, in all three countries, the levels of capital and the TFP were very low relative to German levels, the Solow model and the technology imitation models predicted capital accumulation and an increase in residuals that are larger than in Germany. Over the 13-year period between 1995 and 2007, TFP has converged to Germany's level in all three countries. Relative capital has increased in Poland and in Hungary, but it stayed constant in Czech Republic. Even in the latter two countries, the convergence of capital has been subtantially slower than the convergence of TFP. In fact, contrary to the prediction of the Solow model, the distance between observed capital per efficiency unit and its steady state level increased. The results therefore imply that the Solow model alone is too weak to explain the observed pattern of convergence between countries and it must be suplemented with the technology transfer model.

Given the extraordinary role of the TFP growth, I further decompose it into the increased share of high-skilled (better educated) workers, the increase in relative productivity of a more abundant type of labour and skill-neutral technological progress. The latter two factors played only a minor role in all three countries.

In the second part of the paper, I use sector level data to distinguish whether convergence to the German economy could be explained by the shift of labour towards more productive industries, by an average growth of within-industry productivity or by a growth of relative productivity in sectors in which a converging economy has a higher share of labour than Germany. The results suggests that only the last two factors are equally important in explaining the convergence. The shift of labour to more productive sectors did not play a significant role.

# Bibliography


- Acemoglu, Daron, 1998. “Why Do Technologies Complement Skills? Direct Technical Change and Wage Inequality”, Quarterly Journal of Economics, 113, 1055-1090.
- Acemoglu, Daron, 2000. ”Technical Change, Inequality, and the Labor Market,” NBER Working Papers 7800, National Bureau of Economic Research, Inc.
- Acemoglu, Daron, 2007. ”Equilibrium Bias of Technology,” Econometrica, Econometric Society, vol. 75(5), pages 1371-1409, 09.
- Acemoglu, Daron, 2014. ”Localized and Biased Technologies: Atkinson and Stiglitz’s New View, Induced Innovations, and Directed Technological Change,” NBER Working Papers 20060, National Bureau of Economic Research, Inc.
- Aghion, Philippe, 2002. ”Schumpeterian Growth Theory and the Dynamics of Income Inequality,” Scholarly Articles 3350067, Harvard University Department of Economics.
- Anderson, Simon P., de Palma, Andre and Jacques FranÃ§ois Thisse (1988). ”The CES and the logit : Two related models of heterogeneity,” Regional Science and Urban Economics, Elsevier, vol. 18(1), pages 155-164, February.
- Anderson, Simon P., de Palma, Andre and Jacques FranÃ§ois Thisse (1989). ”Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach”, The Review of Economic Studies, Vol. 56, No. 1 (Jan., 1989), pp. 21-35
- Anderson, Simon P., de Palma, Andre, Nesterov, Yurii (1995). ”Oligopolistic competition and the optimal provision of products”, Econometrica 63, 1281-1302.
- d’Aspremont, Claude, Jean Jaskol Gabszewicz, Jacques FranÃ§ois Thisse (1979). ”On Hotelling’s "Stability in Competition",” Econometrica, Econometric Society, vol. 47(5), pages 1145-50, September.
- Bandiera , Oriana and Imran Rasul (2006). ”Social Networks and Technology Adoption in Northern Mozambique,” Economic Journal, Royal Economic Society, vol. 116(514), pages 869-902, October.
- Baran, Katarzyna (2013), The Determinants of Economic Growth in Hungary, Poland, Slovakia and the Czech Republic in the years 1995-2010, Equilibrium, 8, Issue 3, issue , pages 7-26.
- Barro, Robert J., N. Gregory Mankiw and Xavier Sala-i-Martin (1995). ”Capital Mobility in Neoclassical Models of Growth,” American Economic Review, American Economic Association, vol. 85(1), pages 103-15, March.

- Barro, Robert J. and Xavier Sala-i-Martin (1997). ”Technological Diffusion, Convergence, and Growth,” Journal of Economic Growth, Springer, vol. 2(1), pages 1-26, March.

- Berman, Eli, John Bound and Zvi Griliches (1994): “Changes in Demand for Skilled Labor Within US Manufacturing: Evidence from the Annual Survey of Manufactures”, Quarterly Journal of Economics, 109:2, pp.367-97.

- Berry, Steven, James Levinsohn and Ariel Pakes (1995). ”Automobile Prices in Market Equilibrium”, Econometrica, Vol. 63, No. 4. (Jul., 1995), pp. 841-890.

- Boldrin, Michele and Fabio Canova (2001). ”Inequality and convergence in Europe’s regions: reconsidering European regional policies,” Economic Policy, CEPR;CES;MSH, vol. 16(32), pages 205-253, 04.

- Caplin, Andrew and Nalebuff, Barry (1991). ”Aggregation and Imperfect Competition: On the Existence of Equilibrium,” Econometrica, Econometric Society, vol. 59(1), pages 25-59, January.

- Caselli, Francesco and Wilbur John Coleman II (2006). ”The World Technology Frontier,” American Economic Review, American Economic Association, vol. 96(3), pages 499-522, June.

- Caselli, Francesco and Silvana Tenreyro (2005). ”Is Poland the Next Spain?, ”CEPR Discussion Papers 4877, C.E.P.R. Discussion Papers.

- Chamberlin, Edward H. (1950). ”Product Heterogeneity and Public Policy”, The American Economic Review, Vol. 40, No. 2, Papers and Proceedings of the Sixty-second Annual Meeting of the American Economic Asociation (May, 1950), pp. 85-92

- Ciccone, Antonio and Giovanni Peri (2005). ”Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950-1990,” The Review of Economics and Statistics, MIT Press, vol. 87(4), pages 652-663, November.

- DiNardo, John, Nicole Fortin and Thomas Lemieux (1995). “Labor market institutions, and the distribution of wages, 1973-1992: A semiparametric approach,” Econometrica 64, (1995).

- Dinopoulous, Elias and Peter Thompson (1998). ”Schumpeterian Growth Without Scale Effects”, Journal of Economic Growth, 3: 313–335 (December 1998)

- Dixit, Avinash K. and Joseph E. Stiglitz (1977). ”Monopolistic Competition and Optimum Product Diversity”, The American Economic Review, Vol. 67, No. 3 (Jun., 1977), pp. 297-308

- Fischer, Stanley, Ratna Sahay and Carlos Vegh (1998a). "From Transition to Market; Evidence and Growth Prospects," IMF Working Papers 98/52, International Monetary Fund.
- Fischer, Stanley, Ratna Sahay and Carlos Vegh (1998b). "How Far is Eastern Europe From Brussels?," IMF Working Papers 98/53, International Monetary Fund.
- Foellmi, Reto and Josef Zweimuller (2004). "Inequality, market power, and product diversity," Economics Letters, Elsevier, vol. 82(1), pages 139-145, January.
- Gabaix, Xavier; David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, Casper G. de Vries (2013). "The Impact of Competition on Prices with Numerous Firms" Working Papers 13-07, Chapman University, Economic Science Institute.
- Gradzewicz, Michal, Jakub Growiec, Marcin Kolasa, Lukasz Postek and Pawel Strzelecki (2014). "Poland's exceptional performance during the world economic crisis: New growth accounting evidence," National Bank of Poland Working Papers 186, National Bank of Poland, Economic Institute.
- Growiec, Jakub (2008). "Productivity Differences Across OECD Countries, 1970-2000: The World Technology Frontier Revisited," IBS Working Papers 1/2008, Instytut Badan Strukturalnych.
- Growiec, Jakub (2008). "A new class of production functions and an argument against purely labor-augmenting technical change," International Journal of Economic Theory, The International Society for Economic Theory, vol. 4(4), pages 483-502.
- Growiec, Jakub (2013). "A microfoundation for normalized CES production functions with factor-augmenting technical change," Journal of Economic Dynamics and Control, Elsevier, vol. 37(11), pages 2336-2350.
- Hicks, J. (1932): "The Theory of Wages". London: Macmillan. [1374]
- Hotelling, Harold (1929). "Stability in Competition", the economic journal Vol. 39 No. 153, pp. 41-57
- Howitt, Peter (1999). "Steady Endogenous Growth with Population and R&D Inputs Growing", Journal of Political Economy, Vol. 107, No. 4 (August 1999), pp. 715-730
- Howitt, Peter (2000). "Endogenous Growth and Cross-Country Income Differences," American Economic Review, American Economic Association, vol. 90(4), pages 829-846, September.
- Iradian, Garbis (2007). "Rapid Growth in Transition Economies; Growth-Accounting Approach," IMF Working Papers 07/164, International Monetary Fund.

- Jiang G., Peter Doyle and Louis Kuijs (2001). "Real Convergence to EU Income Levels; Central Europe From 1990 to the Long Term," IMF Working Papers 01/146, International Monetary Fund.

- Jones, Charles I (1995). "R&D-Based Models of Economic Growth," Journal of Political Economy, University of Chicago Press, vol. 103(4), pages 759-84, August.

- Jones, Charles I. (2005). "The Shape of Production Functions and the Direction of Technical Change," The Quarterly Journal of Economics, MIT Press, vol. 120(2), pages 517-549, May.

- Katz, Lawrence F and Kevin M Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," The Quarterly Journal of Economics, MIT Press, vol. 107(1), pages 35-78, February.

- Leamer, Edward E. 1995. "Trade, Wages and Revolving-Door Ideas," NBERWorking Paper No. 4716 (1995).

- Nehru, Vikram and Ashok M. Dhareshwar (1993). "A new database on physical capital stock : sources, methodology and results," Revista de analisis economico, Vol. 8.1993, 1, p. 37-59

- Nelson, Richard R. and Edmond S. Phelps (1965). "Investment in Humans, Technological Diffusion and Economic Growth," Cowles Foundation Discussion Papers 189, Cowles Foundation for Research in Economics, Yale University.

- Ottaviano, Gianmarco P. and Giovanni Peri (2008). "Immigration and National Wages: Clarifying the Theory and the Empirics, No 6916, CEPR Discussion Papers, C.E.P.R. Discussion Papers.

- Ottaviano, Gianmarco Ireo Paolo and Jacques-Francois Thisse (1999). "Monopolistic Competition, Multiproduct Firms and Optimum Product Diversity," CEPR Discussion Papers 2151, C.E.P.R. Discussion Papers.

- de Palma, A., Ginsburgh, V., Papageorgiou, Y., Thisse, J.-F. (1985). "The Principle of Minimum Differentiation Holds under Sufficient Heterogeneity", Econometrica, Vol. 53, No. 4 (Jul., 1985), pp. 767-781

- Peretto, Pietro F. (1995). "Variety, Spillovers and Market Structure in a Model of Endogenous Technical Change", Working paper. Durham, N.C.: Duke Univ., 1995.

- Peri, Giovanni (2009). "The Effect of Immigration on Productivity: Evidence from US States," NBER Working Papers 15507, National Bureau of Economic Research, Inc.

- Perloff, Jeffrey M. and Steven C. Salop (1985). "Equilibrium with Product Differentiation", The Review of Economic Studies, Vol. 52, No. 1 (Jan., 1985), pp. 107-120

- Pettengill, J.S. (1979). "Monopolistic competition and optimal product diversity: COmment", American Economic Review, no.2, 265-279.
- Sachs, Jeffrey (1994). "Poland's Jump to the Market Economy," MIT Press Books, The MIT Press, edition 1, volume 1, number 0262691744, August.
- Salop, Steven C. (1979). "Monopolistic Competition with Outside Goods", Bell Journal of Economics 10, No. 1 (Spring 1979): 141-156
- Samuelson, Paul, (1960). "An Extension of the LeChatelier Principle" Econometrica, 28, 368-379
- Samuelson, Paul A. (1965). "A Theory of Induced Innovation along Kennedy-Weisecker Lines", Review of Economics and Statistics, Vol. 47, No. 4 (Nov., 1965), pp. 343-356.
- Segerstrom, Paul S. (1998). "Endogenous Growth without Scale Effects," American Economic Review, American Economic Association, vol. 88(5), pages 1290-1310, December.
- Solow, Robert M. (1956), "A Contribution to the Theory of Economic growth," Quarterly Journal of Economics, 70, 1 (February), 65-94.
- Wood, Adrian (1994). "North-South Trade, Employment and Inequality: Changing Fortunes in a Skill Driven World", Oxford, Clarendon Press, 1994.
- Van Reenen, John (2011). "Wage inequality, technology and trade: 21st century evidence," Labour Economics, Elsevier, vol. 18(6), pages 730-741.
- Young, Alwyn (1995). "Growth without Scale Effect", working paper no. 5211, National Bureau of Economic Research, August 1995
- Young, Alwyn (1998). "Growth without scale effects", The Journal of Political Economy; Feb 1998; 106, 1. pg. 41

# 4 Appendix A

## A1 Tradeoff in Production Methods Menu.

In this section, I elaborate in more detail the ideas drafted in section 1.2.1 - I present and formally discuss two arguments on why the tradeoff between technologies that assign high productivity to skilled workers and technologies that assign high productivity to unskilled workers can be expected. The first argument is based on the presumption that the adoption of each technology might be costly - a firm might not find it optimal to adopt technologies that increase productivity of unskilled workers if there are very few unskilled workers. The second argument is that the random nature of production method discoveries might generate the trade-off itself. Both arguments are described with formal models.

### A1.1. Costly adoption

One implicit, though potentially problematic, assumption in the analysis presented so far is that technology is taken to be a single object associated with some productivities for skilled and unskilled workers. Instead, in the real world, technologies devoted to skilled workers and technologies devoted to unskilled workers might exist separately. Mathematically, this would imply that each technology is no longer characterized by a vector $(A_s, A_u)$, but by scalars: technologies for skilled workers are characterized by scalar $A_s$, technologies for unskilled workers are characterized by a scalar $A_u$. In such case, firms will simply take the best technology available for skilled workers (buy the fastest available PCs, apply best practices for HR management etc., i.e. maximize $A_s$) and take the best available technology for unskilled workers (purchase the most productive machines, apply best practices in production line organization etc. i.e., maximize $A_u$). Why would the purchase of better computers necessarily involve a necessity to use poorer production machinery?

Such trade-off might arise if firms face adoption costs. Suppose that the more advanced the technology is that the firm aims to adopt, the higher the cost of adoption. Moreover, progressing on the advancement of technologies devoted to skilled workers and technologies devoted to unskilled have different costs: $\frac{c}{\gamma}$ and $c$ respectively. The firm's optimization problem can then be stated as:

$$\max_{L_{is},L_{iu},A_{is},A_{iu}} [(A_{is}L_{is})^{\sigma} + (A_{iu}L_{iu})^{\sigma}]^{\frac{1}{\sigma}} - w_s L_{is} - w_u L_{iu} - \frac{c}{\gamma} A_{is}^{\omega} - c A_{iu}^{\omega}$$

subject to $A_s \leq \bar{A}_s$ and $A_u \leq \bar{A}_u$ where $\bar{A}_s$ and $\bar{A}_u$ are the frontier technologies.

Of course firms might hit the frontier for both technologies. However, if the costs are high enough, this will not happen and firms will not find it optimal to adopt technologies that increase productivity of unskilled workers if there are very few unskilled workers. In fact, if a firm is not choosing frontier technologies, the first order condition will be exactly the same as before, with $\gamma$ ruling the tradeoff between the optimal choices of $A_s$ and $A_u$.

Another possibility is that firms face the credit constraints for adoption of technologies. In such a scenario, a firm's optimization will be given by

$$\max_{L_{is},L_{iu},A_{is},A_{iu}} [(A_{is}L_{is})^{\sigma} + (A_{iu}L_{iu})^{\sigma}]^{\frac{1}{\sigma}} - w_s L_{is} - w_u L_{iu} - \frac{c}{\gamma}A_{is}^{\omega} - cA_{iu}^{\omega}$$

subject to $A_s \leq \bar{A}_s$ and $A_u \leq \bar{A}_u$ and $\frac{c}{\gamma}A_{is}^{\omega} + cA_{iu}^{\omega} \leq B$, where $B$ is the borrowing constraint for the firm. As long as $B$ is not high enough, the firm will not be able to adopt the best technologies for either skilled and unskilled workers, and it will face a tradeoff between investing in the two types of technologies. In such case, the first-order conditions will be exactly the same as before.

### A1.2. Random Discoveries

The purpose of this subsection is to demonstrate that if one is ready to assume that each technology is associated with a pair of productivities: for skilled and unskilled workers (rather than some technologies determining technologies for skilled workers and other technologies ruling productivity of unskilled workers), then the tradeoff can be derived easily by allowing these pairs to be random draws from a bivariate distribution. A similar framework in the context of capital-augmenting and labour-augmenting technology parameters has been proposed by Jones (2005) and developed in Growiec (2008, 2013).

Before proceeding, I should first consider if the assumption of a single technology for both unskilled and skilled workers is defendable. It is hard to justify the assumption if the roles of skilled and unskilled workers are clearly defined and separated and when technologies are only used to help workers perform their duties better. Two types of technologies would probably be observed - one devoted to skilled workers and one devoted to unskilled workers - and the two types are unlikely to be negatively correlated.

However, what happens if the roles of unskilled and skilled workers are not independent of technology? If technology (especially if it is defined broadly, including management strategies and organization of production) itself determines the roles and their division between two types of workers, it has to be treated as a unitary object - firms cannot choose technologies for skilled and unskilled workers separately and independently.

The logic above suggests that if the tasks of workers are predefined and technology determines the division of these tasks between skilled and unskilled workers, the trade-off between skill-biased and unskill-biased technologies appears immediately: firms have to decide to adopt technology where skilled workers assist unskilled workers (key roles in production go to the unskilled) or the technology with unskilled workers assisting the skilled (the key roles go to skilled). However, what if the set of roles is not predetermined but defined by technology? The technology might determine the duties for skilled workers independently of duties for unskilled. The model below captures this idea and shows how random generation of technology might explain the trade-off that is essential for the endogenous technology choice hypothesis.

Imagine a Central Science University that has just devised a new civilizational milestone (such as steam power, semi-conductors, or radioactive decay). The finding has been passed to Central Engineering University, which will try to determine how to combine the new scientific disovery and two types of labour inputs to generate a final good. In fact, they might have various ideas on how to do it, and each idea will involve some degree to which the newly discovered law of nature can compliment the work of skilled and unskilled humans. Thus each idea can be represented with the production function (1) with parameters $(A_{is}, A_{iu})$.

How the ideas look (what are the pairs $(A_{is}, A_{iu})$ that engineers could come up with) depends partly on chance and partly on the nature of the fundamental scientific discovery made in Central Science University. Therefore, one might think about each idea, or rather a pair $(A_{is}, A_{iu})$, that characterizes it as a draw from the bivariate distribution whose parameters depend on the nature of discovery (some fundamental scientific discoveries might be skill-biased by nature in the sense that the explored law of nature compliments ideally with the effort of educated workers - then engineers have much higher chances of discovering production methods with very high $A_s$). Engineers have $n$ ideas and thus $n$ production methods (with $n$ associated $(A_{is}, A_{iu})$ pairs) appear as possibilities to be picked up by firms.

As in the other parts of the paper, the production function is assumed to take a CES from for all available technologies:

$$F_i = [(A_{is}L_s)^\sigma + (A_{iu}L_u)^\sigma]^{\frac{1}{\sigma}}$$

Suppose that it is known that a representative firm uses the technology represented by a solid point on Figure 3. An isoquant going through this point indicates that in areas A and B there must not be any available technology (otherwise firm did not choose the optimal technology). Suppose now that the skilled and unskilled labour supply has changed - in particular relative to the number of unskilled workers the number of skilled workers has increased. This makes the isoquant flatter, as illustrated on the graph. I know for certain that a firm will not jump to technologies in areas A and B because there are no available technologies there. It also do not choose technology from area D because it is suboptimal - it is better to stay with the current technology (represented with a solid dot). The only possibility is then that a firm might find other available technologies in area C - then it will decide to shift. There will be no jump to the technologies that lay below the dotted line. The proposition follows:

**Proposition 1.** Upon a relative increase in the supply of one of the types of labour, the representative firm will never jump to technologies that disfavour this type of labour.

In fact we can tell much more about the changes of optimal choices of the company if we consider a particular bivariate distribution of technologies. We

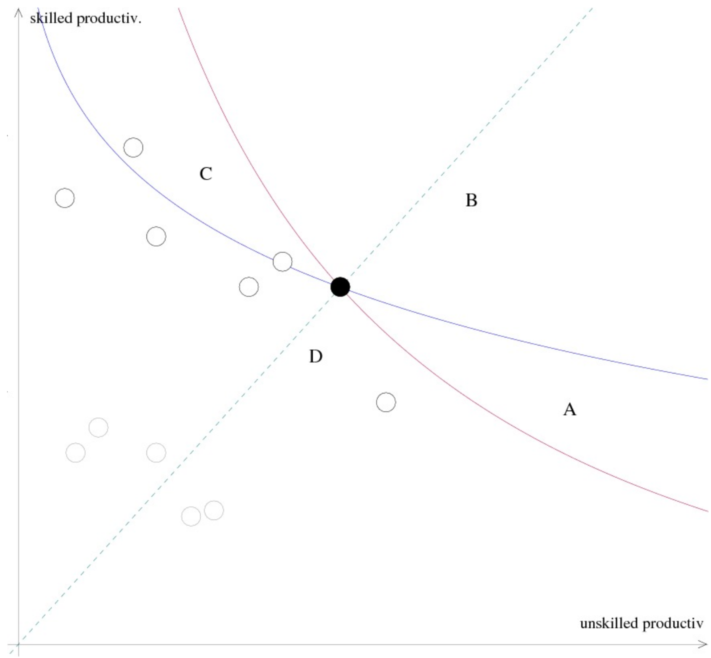



Figure 3: Optimal Choices of Technologies under changing relative supply of skills

might assume that this distribution is a bivariate normal distribution with no correlation between $A_s$ and $A_u$ (to form a production method engineer first draws $A_s$ from a normal distribution, then draw $A_u$ from another normal distribution (but independent of the value of the first draw) and then puts the two draws together). With this assumption, another proposition can be developed:

**Proposition 2.** If a pair $(A_s, A_u)$ is drawn from the bivariate normal distribution with no correlation between $A_s$ and $A_u$, then the most likely ex ante (i.e. before the realization of the draws) choice of technology (the mode of the optimal technology choice distribution) has to satisfy:

$$\log\left(\frac{A_s}{A_u}\right) = \frac{1}{2-\sigma}\log(\gamma) + \frac{\sigma}{2-\sigma}\log\left(\frac{L_s}{L_u}\right)$$

*Proof:* With the bivariate normal distribution with no correlation the equidensity contours in the $(A_s, A_u)$ space takes the form $A_s^2 + \gamma A_u^2 = B$. Each contour is associated with some density level $\pi$ (i.e. if point $(A_s, A_u)$ lays on the contour $k$, the probability (or probability density) that a random draw from the distribution will be $(A_h, A_l)$ is equal to $\pi^k$)

Suppose that there are $n$ independent draws of inventions (and so $n$ points $(A_s, A_u)$ a firm can select from). Each draw will be indexed by $i$. Further, let $F^j = \left(\left(A_s^i L_s^j\right)^\sigma + \left(A_u^i L_u^j\right)^\sigma\right)^{1/\sigma}$ be the output of firm $j$ if it picks the draw $i$.

Take any point in the $(A_s, A_u)$ space, call it point P. The probability (or probability density) that this point ex-ante (before the realization of the draws) is the optimal point for firm $j$ is given by the probability that the first draw happens to be at point P multiplied by the probability that the first draw is optimal for firm $j$ among all the other draws plus the probability that the second draw will happens to be at point P multiplied by the probability that the second draw is best etc.

$$Pr\left(\left(A_s^P, A_u^P\right) \text{ is selected by country } j\right) =$$

$$= \sum_i Pr\left(\left(A_s^i, A_u^i\right) = \left(A_s^P, A_u^P\right)\right) *$$

$$*Pr\left(\left(A_s^P, A_u^P\right) \text{ is optimal for } j\text{among all the draws}\right)$$

If the point $\left(A_s^P, A_u^P\right)$ lays on the contour $k$ then probability that a draw is exactly equal to $\left(A_s^P, A_u^P\right)$ is given by $\pi^k$ . The probability that draw $i$ is optimal for country $j$ among all the other draws is in turn equal to probability that the use of any other technologies that popped out will give smaller output:

$$Pr\left(\left(A_s^P, A_u^P\right) \text{ is optimal for } j\text{among all the draws}\right) =$$

$$= \prod_{s \neq i} Pr\left(F^j\left(A_s^s, A_u^s\right) < F^j\left(A_s^P, A_u^P\right)\right)$$

Now consider particular point E that lays on the contour $k$ depicted on figure 4. Probability that a draw happens to be exactly at point E is $\pi^k$. Probability that this draw is the best option among all the other draws is the probability that no other draw appears in the shaded area above the isoquant passing through point E - if it does, then selection of that alternative draw (and the associated production function) would result in a higher output and selection of point E would be suboptimal. Observe that point H has exactly the same probability that it will appear as an optimum as point E: it lays on the same contour $k$ and it has exactly the same probability that no other draw will give higher output (i.e. the probability that the shaded area remains empty). Now notice that if we consider any point on the contour $k$ between points E and H, say point F, again they have exactly the same probability they will be drawn in one of the draws (they lay on the same contour) but they have strictly higher probability that no other draw will give higher output. This is because the probability that there exist other draw that will give better output than point F is the probability that this draw will appear in the double shaded area. This area is strictly contained within the single shaded area. Thus the probability that poin E is outperformed has to be higher than the probability that point F is outperformed. The only point on the contour $k$ that for which we cannot find a point with higher probability of being and optimum (among other points on contour $k$) is point of tangency of the contour with the isoquant. Therefore point G has the maximum probability of being chosen among all the points on contour.

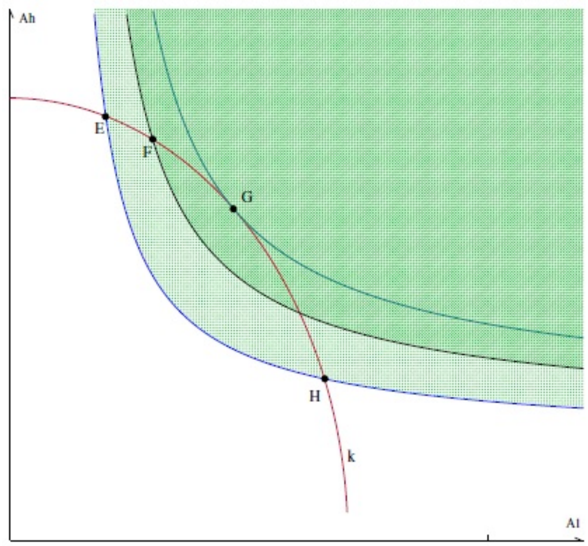



Figure 4: Derivation of density distribution of optimal technology choices

A simple algebra - analogous to the one from Caselli and Coleman model - shows that the point of tangency between equidensity contours and isoquants has to satisfy

$$\left(\frac{A_h}{A_l}\right)^{2-\sigma} = \gamma \left(\frac{L_h}{L_l}\right)^{\sigma}$$

Notice that this condition does not depend on which contour we consider. This means that the mode - the point in $(A_h, A_l)$ space that is most likely to be selected has to satisfy this condition. The condition might be rearranged to the form in the proposition. $\square$

Now consider a bivariate distribution of optimal choices of technologies (ex ante - i.e. before we know which technologies are available to firms). Since proposition 1 tells us that none of the firms will move towards technologies disfavouring skilled workers after they became more numerous (relative to unskilled) and proposition 2 tells us that at least some mass of the distribution of optimal technology choices has to move towards choices that favour more abundant factor we arrive to the final proposition of the paper:

**Proposition 3.** Consider a bivariate distribution of optimal choices of technologies (ex ante - i.e. before we know which technologies are available to firms). The expected optimal choice must shift towards more skill-biased technologies after an increase in the supply of skilled labour

# 5 Appendix B

In this section of the appendix we derive the first two moments and the dispersion of $\theta$ if $x = \theta^{\frac{\rho}{1-\rho}} \sim Gamma\left(\alpha, \beta\right)$

We start with the second moment:

$$D\left(\theta_j\right) = \frac{E\left(\theta^2\right)}{E\left(\theta\right)} - E\left(\theta\right)$$

$$E\left(\theta^2\right) = E\left(\left(x^{\frac{1-\rho}{\rho}}\right)^2\right) = E\left(x^{2\frac{1-\rho}{\rho}}\right)$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \int x^{2\frac{1-\rho}{\rho}} x^{\alpha-1} e^{-\beta x} dx =$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \int z^{2\frac{1-\rho}{\rho}+\alpha-1} \beta^{-2\frac{1-\rho}{\rho}-\alpha+1} e^{-z} \frac{dz}{\beta} =$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \beta^{-2\frac{1-\rho}{\rho}-\alpha} \int z^{\left(2\frac{1-\rho}{\rho}+\alpha\right)-1} e^{-z}$$

$$E\left(\theta^2\right) = \frac{\beta^{-2\frac{1-\rho}{\rho}}}{\Gamma\left(\alpha\right)} \Gamma\left(2\frac{1-\rho}{\rho} + \alpha\right)$$

The first moment can be derived analogously:

$$E\left(\theta\right) = E\left(x^{\frac{1-\rho}{\rho}}\right) = \int x^{\frac{1-\rho}{\rho}} \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} x^{\alpha-1} e^{-\beta x} dx =$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \int x^{\frac{1-\rho}{\rho}+\alpha-1} e^{-\beta x} dx =$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \int z^{\frac{1-\rho}{\rho}+\alpha-1} \beta^{-\frac{1-\rho}{\rho}-\alpha+1} e^{-z} \frac{dz}{\beta} =$$

$$= \frac{\beta^\alpha}{\Gamma\left(\alpha\right)} \beta^{-\frac{1-\rho}{\rho}-\alpha} \int z^{\left(\frac{1-\rho}{\rho}+\alpha\right)-1} e^{-z} dz$$

$$E\left(\theta\right) = \frac{\beta^{-\frac{1-\rho}{\rho}}}{\Gamma\left(\alpha\right)} \Gamma\left(\frac{1-\rho}{\rho} + \alpha\right)$$

We can combine the two to derive the dispersion of $\theta$ if $x = \theta^{\frac{\rho}{1-\rho}} \sim Gamma\left(\frac{\mu}{D}, \frac{1}{D}\right)$

$$D\left(\theta_j\right) = \frac{E\left(\theta^2\right)}{E\left(\theta\right)} - E\left(\theta\right) =$$

$$= D^{\frac{1-\rho}{\rho}} \frac{\Gamma\left(2\frac{1-\rho}{\rho} + \frac{\mu}{D}\right)}{\Gamma\left(\frac{1-\rho}{\rho} + \frac{\mu}{D}\right)} - \frac{D^{\frac{1-\rho}{\rho}}}{\Gamma\left(\frac{\mu}{D}\right)} \Gamma\left(\frac{1-\rho}{\rho} + \frac{\mu}{D}\right) =$$

$$= D^{\frac{1-\rho}{\rho}} \left[ \frac{\Gamma\left(2\frac{1-\rho}{\rho} + \frac{\mu}{D}\right)}{\Gamma\left(\frac{1-\rho}{\rho} + \frac{\mu}{D}\right)} - \frac{\Gamma\left(\frac{1-\rho}{\rho} + \frac{\mu}{D}\right)}{\Gamma\left(\frac{\mu}{D}\right)} \right]$$

Simulations for various value of $\mu$ show that this is function is increasing in $D$.