Estimating country heterogeneity in aggregate capital-labor substitution using panel data: A bayesian fixed effects approach

Lucciano Villacorta†

January 14, 2016

For the latest version, please click *here*

Abstract

The aggregate elasticity of substitution between labor and capital is central in understanding the global decline in the labor share and cross-country heterogeneity in technologies. However, very different explanations arise depending on whether the elasticity of substitution is below or above one. In this paper, I develop a flexible framework to estimate the aggregate elasticity of substitution between labor and capital for a panel of countries. In contrast to previous studies, my framework considers country heterogeneity in the elasticity of substitution. The growth rates of labor- and capital- augmenting technologies are modeled as a dynamic factor model. Estimation is based on posterior distributions in a Bayesian fixed effects framework. I propose a computationally convenient procedure to compute posterior distributions in two steps that combines the Gibbs and the Metropolis-Hasting algorithm. Using the EU KLEMS database, I find evidence of heterogeneity in the elasticity of substitution across countries, with a mean of 0.92, a median of 0.87 and a standard deviation of 0.23. The bias in the technical change is the dominant mechanism in explaining the labor share decline in the majority of countries. However, the increase in the capital-labor ratio (or the decline in the price of investment goods) is also an important mechanism for some countries. Finally, I find that the growth rate of the labor augmenting technology relative to the growth rate of the capital augmenting technology correlates negatively with the growth rate of the capital-labor ratio.

Keywords: CES, Elasticity of substitution, capital- and labor- augmenting technologies, nonlinear panel data, unobserved heterogeneity, factor model, Bayesian fixed effects , Gibbs, Metropolis-Hasting.

JEL codes: C110, C150, C330, C380, E230

*I am deeply thankful to Manuel Arellano and Stephane Bonhomme for their guidance and extensive discussions. Also I would like to thank Dante Amengual, Jesus Carro, Julio Crego, Rosario Crinò, Ivan Fernandez Val, Chris Hansen, Loukas Karabarbounis, Monica Martinez, Pedro Mira, Aureo de Paula, Josep Pijoan, Julio Galvez, Luca Repetto, Raul santaeulalia-Llopis, Enrique Sentana, Alonso Villacorta, Fabrizio Zilibotti and Jeffry Wooldridge for useful comments and discussions. Also, audiences at the 11th World Congress of the Econometric Society, the Bristol Econometric Study Group 2014, the Symposium of the Spanish Economic Association, the Peruvian Economic Association Congress 2014, CEMFI PhD workshop, ENSAI, the Central Bank of Peru and the Central Bank of Chile. Support from the European Research Council/ ERC grant "Estimation of nonlinear panel models with unobserved heterogeneity" (grant agreement n. 263107) is gratefully acknowledged. Errors and omissions are exclusively my own responsibility.

†Central Bank of Chile. Email: lucciano.villacorta@gmail.com

1 Introduction

The aggregate elasticity of substitution between labor and capital is central in answering several fundamental questions in macroeconomics. It plays a crucial role, for example, in understanding the global decline in the labor share. However, very different explanations arise depending on whether the elasticity of substitution is below or above one. If it is higher than one, the decline in the labor share is associated either with a decrease in the price of investment goods (Karabarbounis and Neiman, 2014) or with an increase in the capital-output ratio (Piketty and Zucman, 2014). Conversely, if the elasticity of substitution is lower than one, the decline in the labor share is associated with an increase in labor-augmenting technology (Acemoglu, 2002). The elasticity of substitution is also crucial in models of directed technical change. Such models are useful in explaining country heterogeneity in efficiency of the factors of production and in development accounting. These models also have different implications depending on the value of the elasticity of substitution. Indeed, Caselli (2005) concludes that the most important outstanding question in development accounting is to know the value of the aggregate elasticity of substitution.

The existing literature has estimated the aggregate elasticity of substitution using either time series variation for a particular country or cross-country variation under the assumption of country homogeneity in the elasticity of substitution. However, these two approaches deliver different results. The time series approach, which typically assumes a constant growth rate of the technology, finds an elasticity lower than one (Antras, 2004; Herrendorf, Herrington and Valentinyi, 2013). Conversely, the cross-country approach finds an elasticity of substitution higher than one under the assumption of a common elasticity of substitution and a common growth rate of technology across countries (Karabarbounis and Neiman, 2014).[1]

In this paper, I develop a flexible framework to estimate the elasticity of substitution from a panel of countries, allowing for unobserved heterogeneity. In contrast to previous studies, my framework considers country heterogeneity in the elasticity of substitution and the growth rates of both the labor- and capital- augmenting technologies. Additionally, I endeavor to keep as flexible as possible the assumptions on the technology process. In particular, the growth rates of the capital- and labor-augmenting technologies are allowed to vary over time and across countries while retaining some commonalities across the panel via underlying factors.

There are theoretical and empirical reasons to expect that there is important heterogeneity across countries in both the elasticity of substitution and the growth rates of technical change. For instance, a recent contribution by Oberfield and Raval (2014) shows that the aggregate elasticity of substitution depends on the heterogeneity in capital intensity across firms.[2] Hence, differences in the capital intensity distribution across countries may imply cross-country heterogeneity in the elasticity of substitution. Moreover, models of directed technical change suggest cross-country heterogeneity in the capital- and labor-augmenting technologies (Caselli and Coleman, 2006;

[1] Duffy and Papageorgiou (2000) also find an elasticity of substitution higher than one. They estimate the Constant Elasticity of Substitution production function for 82 countries over a 28-year period assuming homogeneity of all the parameters of the corresponding production function and Hicks-neutral technology.

[2] Oberfield and Raval (2014) estimate the elasticity of substitution using plant-level data and then aggregate across plants to derive the aggregate elasticity, allowing for reallocation of inputs across plants. They find an aggregate elasticity of 0.7 for the U.S, 0.84 for Chile and 1.11 for India.

Acemoglu, 2002).[3] Finally, the data show evidence of considerable cross-country heterogeneity in the evolution of the labor share, even in countries with similar patterns of capital-output ratio or factor prices (see Piketty and Zucman (2014) for an example). Such evidence suggests cross-country differences in the elasticity of substitution and/or the path of the technical change.

I identify the elasticity of substitution and the technical change from the supply system model, which consists of the Constant Elasticity of Substitution (CES, henceforth) production function and the price equations (the labor share equation and the capital share equation). The supply system is a suitable model in terms of estimation as it reflects both optimizing behavior and technology. It also incorporates additional moment conditions with respect to estimating a single equation, which might potentially lead to an improvement in the precision of the estimates. I use the EU KLEMS database, which is an unbalanced panel of 20 countries observed from 1970 to 2007, where the sample size of the time series in each country ranges from 15 to 38 years.[4] The model is an unbalanced nonlinear panel system of equations with random coefficients (country-specific elasticities of substitution) and unobserved time varying-specific factors that interact with country-specific loadings (growth rates of the capital- and labor-augmenting technologies). This setup is non-standard, and its estimation is challenging given the non-linearity and the amount of unobserved heterogeneity allowed, both in the cross-sectional and the time-series dimension.

One possibility to estimate the model is to follow a fixed effects approach, that is, treating each of the country- or time- specific effects as parameters and estimate them via GMM. This approach has the advantage of generality but has two main drawbacks. First, the (large-N and large-T) statistical properties of the fixed effects estimator in set-ups with country- and time- specific effects have only been studied in linear models(Bai, 2009; Moon and Weidner, 2014), and in some specific nonlinear models (Fernández-Val and Weidner, 2013). My model does not belong to any of these cases. Second, the fixed effects estimator of this model is computationally problematic due to the nonlinearity and the high dimensionality of the model. Another possibility is a random effects approach, that is, completing the model with distributional assumptions for the country- and time-specific latent variables and computing estimates directly from the marginal likelihood. This approach has the advantage of computational tractability but the disadvantage of sensitivity to misspecification of the distribution of the latent variables.

In this paper I adopt an alternative approach, known as the Bayesian fixed effects estimator (BFE), which also specifies distributions for the latent variables but they are used as priors in a Bayesian fashion. These priors are combined with data to form posterior distributions which are then used in obtaining estimates of average effects in my model. This is in contrast to the random effects approach where priors are not updated since they are taken as the truth. The BFE can be expected to have similar asymptotic properties to the standard fixed effects when N and T go to infinity, but the use of priors may lead the BFE estimator to have better properties in small samples.[5] In addition, the Bayesian fixed effects approach provides a computationally tractable way of estimating the model

[3]In these models, the direction of the technical investment depends on the relative factor endowment $\frac{K}{L}$ and the elasticity of substitution between factors.

[4]Only 20 of the 30 countries of the EU KLEMS database have information in the variables I use for the estimation.

[5]Arellano and Bonhomme (2009) prove the consistency of the Bayesian fixed estimator (BFE) of average effects when T goes to infinity in nonlinear models with just individual specific effects, even when the parametric model of the individual effect is misspecified. The BFE will be also consistent in my setup as N and T go jointly to infinity.

by using Markov Chain Monte Carlo (MCMC) methods. Given the non-linearity of the supply system model, I construct the Markov chain using the Metropolis-Hasting (M-H) within Gibbs algorithm. The computational challenge in this algorithm is to choose an adequate starting value for each latent variable, and more importantly, to set proposal distributions for each of them (more than 200 in my model). To this end, I propose a feasible and computationally efficient procedure to obtain posterior distributions, by dividing the estimation in a two-step procedure. The over-identification of the model allows me to compute the posterior distribution of all the latent variables of the model by working either with the price equations, the nonlinear CES or both. Given that the price equations have a linear structure, in the first step I derive closed form solutions for each conditional posterior distribution and compute the posterior distributions by using a Gibbs sampling algorithm.[6] In the second step, I use the distributions simulated in the first step as the proposal distributions in the M-H within Gibbs algorithm to estimate the complete supply system model.

This framework allows to quantify the extent of cross-country heterogeneity in the elasticity of substitution and the pattern of technological change. This allows me to shed light on questions related to the global decline in the labor share and the direction of technological change. First, it allows me to assess which of the two competing mechanisms that explain the evolution of the labor share is more relevant in a particular country.[7] Second, it allows me to test the predictions of models of directed technical change. Finally, my framework provides a flexible manner of controlling for unobserved heterogeneity both in the cross-section and the time series dimensions, which helps in identifying the elasticity of substitution.

The results show heterogeneity in the elasticity of substitution with a mean of the cross-country distribution equal to 0.92, a median equal to 0.87 and a standard deviation of 0.23. Most countries have an elasticity of substitution below one, but a few countries (e.g., Spain or Portugal) have an elasticity higher than one. I also find that the bias in the technical change is the dominant mechanism in explaining the evolution of the labor share for the majority of countries in the database. However, the increase in the capital-labor ratio (or the decline in the price of investment goods) is also an important mechanism for some countries. The results also show heterogeneity in the growth rates of the capital- and labor-augmenting technologies. Moreover, I find that the growth rate of the labor-augmenting technology relative to the growth rate of the capital-augmenting technology correlates negatively with the growth rate of the capital-labor ratio. These result is in line with the directed technical change models, in which countries invest in technologies that increase the efficiency of the scarce factor.

The rest of the paper is organized as follows. Section 2 introduces the CES framework and its relation with the literature of labor share and development accounting. Section 3 presents the panel model with unobserved heterogeneity. Section 4 discusses the methodological approach. Section 5 proposes a method of computation of the posterior distributions and section 6 shows the results. In section 7 I conclude.

[6] In the Gibbs sampler, I do not need to set any proposal distribution, given that I can draw directly from the known conditional distributions of the latent variables.

[7] The elasticity of substitution governs how aggregate factors react to changes in factor prices.

2 Preliminaries

I start by describing the CES model and its relation with the literature of labor share and directed technical change. I also describe the main identification problems in the estimation of the elasticity of substitution. In the next section I present the panel model with unobserved heterogeneity.

2.1 The aggregate CES production function with factor augmenting technology

Consider an aggregate constant elasticity of substitution (CES) production function that relates the total amount of income Y in a economy with the the total amount of some measure of aggregate capital K and the total amount of some measure of aggregate labor L.

$$Y = \left[\delta(A_K K)^{\rho} + (1-\delta)(A_L L)^{\rho}\right]^{\frac{1}{\rho}} \tag{1}$$

where A_K and A_L are two separate technology terms, $\delta \in (0,1)$ is a distribution parameter which determines how important the two factors are, and $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution between capital and labor, the factors of production.[8] When $\rho > 0$ ($\sigma > 1$) the two factors are gross substitutes and when $\rho < 0$ ($\sigma < 1$) they are gross complements. When $\rho = 0$ ($\sigma = 1$), the production function is Cobb-Douglas. The terms A_K and A_L capture capital and labor augmenting progress, respectively. The factor augmenting technology is a specific technical change that complements one of the factors. If $A_K > A_L$, we say that the economy uses capital more efficiently and if $A_L > A_K$ the economy uses labor more efficiently. δ is the distribution parameter and ensures that the labor share and the capital share are different from zero when $\rho = 0$.

2.1.1 Non constant factor shares

The CES production function allows for time varying evolution of the capital and the labor share. Under the assumption of competitive markets and denoting r as the user cost of capital and w as the market price of a unit of human capital, the following equations hold in equilibrium:

$$sh_K \equiv \frac{rK}{Y} = \delta A_K^{\rho}\left(\frac{K}{Y}\right)^{\rho} \tag{2}$$

$$sh_L \equiv \frac{wL}{Y} = (1-\delta)\, A_L^{\rho}\left(\frac{L}{Y}\right)^{\rho} \tag{3}$$

$$sh_{K/L} \equiv \frac{rK}{wL} = \frac{\delta}{1-\delta}\left(\frac{A_K}{A_L}\right)^{\rho}\left(\frac{K}{L}\right)^{\rho} \tag{4}$$

where sh_K is the capital share, sh_L is the labor share and $sh_{K/L}$ is the relative share. In the case of $\rho > 0$, the relative capital share $sh_{K/L}$ depends positively on the capital-labor ratio ($\frac{K}{L}$) and on the bias of technical change

[8] The distribution parameter δ is the capital share in case of a Cobb Douglas production function, i.e $\rho = 0$.

($\frac{A_K}{A_L}$). Instead when $\rho < 0$ the relation is the opposite. In contrast, the Cobb-Douglas function features constant shares.

Several papers have documented a decline in the labor share in many countries for the last thirty years (e.g; Jones (2005); Karabarbounis and Neiman (2014); Piketty and Zucman (2014)).[9] There are two main competing mechanisms for explaining the declining labor share. However, each of these mechanisms relies on different values of the aggregate elasticity of substitution. One mechanism works through changes in factor prices. Example of this are Piketty and Zucman (2014), who suggest that the increase in capital accumulation leads the decline of the labor share, and Karabarbounis and Neiman (2014), in which the decline in the relative price of investment generates the decline in the labor share. These explanations rely on $\rho > 0$. The other mechanism works through a change in technology. For example, Acemoglu (2002) developed a model of directed technical change in which the bias of the technical change explains the decline in the labor share. This explanation relies on a $\rho < 0$.

Since my framework allows for country-specific elasticity of substitution and country-specific factor bias in the technical change, I can study which of the competing mechanisms in explaining the evolution of the labor share is most relevant in a particular country.

2.1.2 Non-neutral technical change

The CES framework with factor augmenting technology allows for non-neutral technical change, which means that shifts in production technologies could favor the marginal productivity of a specific factor more. Instead, in the most conventional Cobb-Douglas case, technical change is factor-neutral. Importantly, how each factor's marginal productivities are affected with a specific change in technology depends on the value of the elasticity of substitution. To see this, consider the relative marginal product of the two factors:

$$\frac{MP_K}{MP_L} \equiv \frac{r}{w} = \frac{\delta}{1-\delta}\left(\frac{A_K}{A_L}\right)^{\rho}\left(\frac{K}{L}\right)^{\rho-1} \tag{5}$$

If $\rho > 0$ ($\rho < 0$), an increase in the capital augmenting technology relative to the labor augmenting technology $\left(\uparrow \frac{A_K}{A_L}\right)$ is capital (labor) biased since it favors more the marginal productivity of capital (labor). Note that in the Cobb Douglas case ($\rho = 0$) , the technical change is neutral since it does not affect the relative marginal product of the factors.

Structural models of technical change. Since the CES production function allows for the possibility of non-neutral technical change, it plays a central role in explaining the direction of technical change. Acemoglu (2002), Acemoglu and Zilbotti (2001) and Caselli and Coleman (2006) use the CES framework with factor- augmenting technology to develop models of endogenous technical change. In these models, the direction of the technical investment depends on the relative factor endowment $\frac{K}{L}$ and the elasticity of substitution between factors. If the elasticity of substitution is low enough (high enough), countries invest in technologies that increase the

[9] Karabarbounis and Neiman (2014) document a 5 percentage point decline in the share of global corporate gross value added paid to labor over the past 35 years. They also show that 38 out of the 56 countries analyzed, exhibited downward trends in their labor share. Piketty and Zucman (2014) show that capital shares did rise in rich countries during the 1970-2010 period, from about 15%-25% in the 1970s to 25%-35% in the 2000s-2010s, with large variations over time and across countries.

efficiency of the scarce factor (abundant factor). These models have been used to explain efficiency differences across countries and play an important role in development accounting.

Acemoglu and Zilbotti (2001) developed a model where new technologies used by the less developed economies are originated in developed economies and are designed to make optimal use of their factors. Given the difference in factor endowment between less developed economies and developed economies , the new technologies are not "appropriate for the less developed ones". Therefore, even when all countries have equal access to new technologies (ideas can flow rapidly across countries and machines that incorporate better technologies can be imported by less developed countries), the technology-skill mismatch leads to sizable differences in factor productivity. Conversely, Caselli and Coleman (2006) developed a model in which each country can choose their own "appropriate" technology from a distribution of technologies determined by the technology frontier in each country. As in Acemoglu and Zilbotti (2001), the direction of the technical change depends on the factor endowment and the elasticity of substitution between the factors. Therefore, rich countries might be more efficient in the use of one factor, whereas poor countries might be more efficient in the use of the other factor.[10]

Since my framework allows to recover the process of the capital and the labor augmenting technology, I am able to test the hypothesis of these structural models.

2.2 Identification and Estimation Issues

Identification. Identifying the elasticity of substitution is difficult. The elasticity of substitution can not be identified from time series data on output, inputs, and prices, unless we impose assumptions on the path of the technical change (Diamond and McFadden, 1965). In the production function setup, the first order conditions of a firm maximization problem suggest that the observed inputs are chosen as a function of the unobserved technologies, causing OLS estimates to be biased. In this sense, both aggregate capital and labor are functions of aggregate productivity. Additionally, the models of directed technical change also state a correlation between the labor- and capital- augmenting technology, the elasticity of substitution and the endowment of capital and labor.

The literature has generally circumvented this problem by imposing some type of structure on the functional form of technical change.[11] Nevertheless, as was pointed out by Antras (2004), results of the estimation of the elasticity of substitution might be driven by the assumptions imposed on the behavior of the technology.[12] These

[10]Working a CES framework with non-neutral technology as in 1 and using data on output, physical capital and human capital for many countries in 1996, Caselli (2005) finds that poorer countries use physical capital more efficiently whereas richer countries use human capital more efficiently. In particular, Caselli (2005) assumes that there is no measurement error in the labor share and capital share equation (3 and 2) and recover directly A_K and A_L for a given value of ρ that is assumed to be common for all the countries.

[11]The state of the art to deal with this endogeneity problem is to use observed input decisions to recover the unobserved productivity and control for it in the production function equation, (see Olley and Pakes (1996), (Levinsohn and Petrin, 2003; Doraszelski and Jaumandreu, 2013). This control function approach has been applied in a large number of recent empirical papers that estimate Cobb Douglas production at the firm level.

Aside from the collinearity problem and the timing assumptions stressed by (Ackerberg, Caves and Frazer, 2006), the presence of factor augmenting technology, the non-linearity of the CES and the measurement error in the labor and capital compensation at the aggregate level prevent the use of the control function approach, given that the dimensionality assumption of the first stage is violated. Therefore, is not possible to recover A_K and A_L from the input decisions as in Olley and Pakes (1996), Levinsohn and Petrin (2003) and Doraszelski and Jaumandreu (2013).

[12]To stress the last point, consider, for instance, the U.S. case for the period 1970 to 2005, a period in which the labor share has remained somehow stable, whereas the capital-labor ratio has increased dramatically. Under the assumption of Hicks neutrality i.e.

different assumptions may explain the wide range of estimates in the literature.

For instance, Karabarbounis and Neiman (2014) estimate an elasticity of substitution above one using a very carefully constructed database of corporate labor compensation and price of investment goods in 57 countries. They assume a common growth rate of capital-augmenting technology and a common elasticity of substitution across countries. Two sources of bias may affect the estimation. First, models of directed technical change suggest that less developed countries have a higher capital augmenting technology. This might create an upward bias in the estimation if less developed countries have experienced a lower decrease in the price of investment goods than developed countries.[13] Second, in the case of heterogeneous elasticities of substitution, estimating a common parameter might be interpreted as an estimator of the cross-country average of the elasticity of substitution. This estimator might present an upward bias if the countries with lower elasticity of substitution are the ones which experienced a higher decrease in the price of investment goods.

Estimation model. The estimation of the parameters of a CES production function typically has proceeded by estimating the price equations (equations 2 and 3). The advantage of using the price equations is that after taking logs, they become linear in the inputs. However, a drawback of the estimation based on the price equations is the lack of long time series data on labor share and capital share, which introduce small sample bias. In addition, observations on compensation of inputs, such as interest rate and wages, tend to present considerable measurement error, rendering the variance of the estimates a first-order concern.

The supply system which consists of the joint estimation of the production function and the price equations is a suitable model in terms of estimation as it helps with all these empirical issues. The incorporation of more moment conditions via the CES as well as the lower variance of the measurement error of output data in relation to prices, potentially lead to an improvement in the precision of the estimates. León-Ledesma, McAdam and Willman (2010) show the superiority of the supply system model over the prices equations by Monte Carlo simulations. Moreover, the estimation of the CES is particularly important in the case of misspecification of the price equations due to imperfect competition.

3 Empirical Panel Data Framework with Unobserved Heterogeneity

Consider the supply system model for a panel of countries $i = 1, \ldots, N$ in the period $t = 1, \ldots, T$.

$$log(Y_{it}) = \frac{1}{\rho_i} log\left[\delta_i \left(A_{K,it}K_{it}\right)^{\rho_i} + (1-\delta_i)\left(A_{L,it}L_{it}\right)^{\rho_i}\right] + \varepsilon_{Y,it} \tag{6}$$

$A_K = A_L \Longrightarrow sh^{K/L} = \left(\frac{K}{L}\right)^{\rho}$, the only value for the elasticity of substitution that is consistent with the previous facts is one ($\rho = 0$), which is delivered by a Cobb Douglas function. If we allow for bias in the technical change $A_K \neq A_L \Longrightarrow sh^{K/L} = \left(\frac{A_K}{A_L}\right)^{\rho}\left(\frac{K}{L}\right)^{\rho}$ we can replicate the same stable path of labor share and increasing capital-labor ratio with an elasticity of substitution lower than 1 ($\rho < 0$) and predominant labor augmenting technical change.

[13]The argument works for an elasticity of substitution less than one. When the elasticity of substitution is higher than one, the models of directed technical change imply a higher capital-augmenting technology in developed countries. In this case, an upward bias is possible if less developed countries have experienced a lower decrease in the price of investment goods than developed countries, otherwise the bias will be downward. Caselli (2005) shows that poor countries have higher A_K and rich countries have higher A_L.

$$log(sh_{L,it}) = \rho_i log\left(1-\delta_i\right)\rho_i + log\left(A_{L,it}\right) + \rho_i log\left(\frac{L_{it}}{Y_{it}}\right) + \varepsilon_{W,it} \tag{7}$$

$$log(sh_{K,it}) = \rho_i log\left(\delta_i\right) + \rho_i log\left(A_{K,it}\right) + \rho_i log\left(\frac{K_{it}}{Y_{it}}\right) + \varepsilon_{R,it} \tag{8}$$

where equation 6 is the log of the CES production function in 1 , and 7 and 8 are the log of the labor share equation in 3 and the log of the capital share equation in 2, respectively. ρ_i is a country-specific elasticity of substitution and δ_i is a country-specific distribution parameter.[14] The variables $A_{L,it}$ and $A_{K,it}$ are the level of labor and capital augmenting technology of country i in period t. Finally $\varepsilon_{Y,it}$, $\varepsilon_{W,it}$ and $\varepsilon_{R,it}$ represents the measurement errors of the log of the output, the log of the labor share and the log of the capital share respectively.

Heterogeneity in elasticity of substitution. The heterogeneity in the elasticity of substitution ρ_i is the first departure from previous studies and is a key feature of my model. In a novel contribution Oberfield and Raval (2014) show how to derive the aggregate elasticity of substitution from the firm-level elasticity of substitution, by allowing for reallocation of capital and labor across firms. They show that the aggregate elasticity of substitution depends on the firm-level elasticity of substitution and the heterogeneity in capital intensity among firms. They also find heterogeneity in the estimates of the aggregate elasticity of substitution among the U.S, Chile, Colombia and India, mainly due to differences in the capital intensity distribution in these countries. Oberfield and Raval (2014) is the first attempt that studies cross-country heterogeneity in the elasticity of substitution. My paper differ from them in various aspects. First, I estimate the elasticity of substitution for a larger number of countries, allowing for common dynamic fluctuations in the technology patterns across countries rather than estimating each of the countries separately. Second, while Oberfield and Raval (2014) estimate the price equations from a firm-level perspective, I estimate the complete supply system model using aggregate data. Third, I model the capital- and labor- augmenting technology using a factor structure while they recover the technology as a residual. Finally, the estimation approach is also different.

The country-specific elasticity is also motivated by the considerable heterogeneity in the path of the labor share across countries. Despite the well documented global decline in the labor share, the data shows considerable evidence of country heterogeneity in the pattern of the labor share, even in countries with similar patterns in the capital output ratio and in the factor prices. The former suggests cross-country differences in the elasticity of substitution and/or the path of the bias in the technical change.

The country-specific elasticity of substitution is assumed to be time-invariant, as changes in technologies that might change the way countries substitutes capital and labor will be captured by the variables $A_{K,it}$ and $A_{L,it}$ which are allowed to change over time. Moreover, one reason that might explain heterogeneity in the capital intensity across firms is misallocation of resources which is associated to quality of institutions, labor market policies, financial constraints, and other sources of social infrastructure that might change slowly over time.[15]

[14] A country-specific δ_i allows for a country-specific labor share in the case of $\rho_i = 0$.

[15] Oberfield and Raval (2014) find that the elasticity of substitution has been stable over the past forty years. They show that the aggregate elasticity of substitution has risen slightly from 0.67 in 1972 to 0.75 in 2007

The Factor model as a reduced form model of the labor and capital augmenting technology. Given that the assumptions imposed over the technology process are crucial to identify the elasticity of substitution, I keep the technology process as flexible as possible. The second feature of my framework is to consider a growth rate of the capital- and labor- augmenting technology that is allowed to be different in every period of time and in every country, while retaining some commonalities across the panel via underlying factors.[16]

$$A_{K,it} = A_{K,i0} exp\left(\sum_{s=0}^{t} \gamma_{K,is}\right) \tag{9}$$

$$A_{L,it} = A_{L,i0} exp\left(\sum_{s=0}^{t} \gamma_{L,is}\right) \tag{10}$$

The initial level of capital and labor augmenting technology $A_{K,i0}$ and $A_{L,i0}$ are allowed to be heterogeneous across countries. The annual growth rate of technology is modeled as a country fixed effect plus one underlying factor.

$$\gamma_{K,it} = \alpha_{K,i} + \lambda_{K,i} f_{K,t} \tag{11}$$

$$\gamma_{L,it} = \alpha_{L,i} + \lambda_{L,i} f_{L,t} \tag{12}$$

The terms $\alpha_{K,i}$ and $\alpha_{L,i}$ capture a country-specific growth rate of technology. The terms $f_{K,t}$ and $f_{L,t}$ capture a growth rate of technology that might change over time and affect all countries with different intensities $\lambda_{K,i}$ and $\lambda_{L,i}$.

The factor model may capture, in a reduced-form fashion, some important features of the structural models of directed technical change. For example, a common factor is consistent with the idea of Acemoglu and Zilbotti (2001) where technologies are created in developed economies and acquired by the rest of countries. The loadings in the factor, $\lambda_{L,i}$ and $\lambda_{K,i}$, imply a heterogeneous acquisition of technology across countries. This could be interpreted as heterogeneous barriers in countries that prevent the immediate adoption of technologies;[17] or as different incentives in the adoption of new technologies. The terms $\alpha_{K,i}$ and $\alpha_{L,i}$ capture country specific technologies, consistent with the story of Caselli and Coleman (2006) where countries develop their "own appropriate technology".[18]

[16]The factor model is a way to reduce the dimensionality of a two dimensional variable γ_{it} into a few number of underlying factors f_t with a country specific loading λ_i. A sufficiently large number of factors can approximate any two dimensional variable. Obviously, there is a tradeoff between number of factors and accurate estimation. Nevertheless, as emphasized by Acemoglu and Zilbotti (2001) there are few countries that develop new technology, so a small number of factors can do a good job in capturing the behavior of the technological process.

[17]Hall and Jones (1999)and Lagos (2006) emphasized that the quality of institutions, labor market policies, financial constraints, and others sources of social infrastructure might work as barrier that prevent the immediate adoption of the technology process.

[18]To see the latter, let us consider the technology of country i : $A_{L,it} = G_{L,it}^{\eta_{L,i}} F_{L,t}^{\lambda_{L,i}}$ and $A_{K,it} = G_{K,it}^{\eta_{K,i}} F_{K,t}^{\lambda_{K,i}}$ as a combination of two different complementary technologies. Two country-specific technologies $G_{L,it}$ and $G_{K,it}$ that grow at a constant rate $g_{L,i}$, $g_{K,i}$ and two common

The technology structure in equations 11-12 provides more flexible patterns of the evolution of technology with a wide range of possible time paths and country heterogeneity. This has two main advantages. On the one hand, it allows me to test models of directed technical change.[19] On the other hand, the reduced form model allows me to control for unobserved heterogeneity in the cross-section and the time-series, in the same fashion as in Bai (2009).[20] The latter helps with the identification concerns in the estimation of the elasticity of substitution in comparison to previous studies.

Objects of interest. The elasticity of substitution in each country ρ_i is the primary objective of interest in this study. Nevertheless, in my model, some other objects of interest are averages over individual effects, such as marginal average effects. As a first example, consider an object that summarizes the effect of the capital accumulation across the world on the global decline in the labor share. This average marginal effect is just the mean of the elasticity of substitution across countries. (see equation 8)

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial\triangle log(sh_{it}^{K})}{\partial\triangle log\left(\frac{K_{it}}{Y_{it}}\right)}\right]=\frac{1}{N}\sum_{i=1}^{N}[\rho_i] \tag{13}$$

Another object of interest is the world average effect of a change in the growth rate of the world labor- augmenting technology relative to the world capital- augmenting technology on the global decline of the labor share. Combining equations 7, 8, 9,10, 11 and 12 I can expressed this effect by:

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial\triangle\, log(sh_{it}^{L})}{\partial\left(f_{L,t}-f_{K,t}\right)}\right]=\frac{1}{N}\sum_{i=1}^{N}\left[\rho_i\left(\lambda_{L_i}-\lambda_{K_i}\right)\right]$$

As in Oberfield and Raval (2014), I can assess the cumulative contribution of the bias in the technical change A_K/A_L and the cumulative contribution of the increase in the capital-labor ratio K/L to the evolution of the labor share for every country in the sample. Combining equations 7, 8, 9,10, 11 and 12 , the cumulative contribution of the bias in the technical change between the years t_{0i} and T_i, holding the capital-labor ratio as constant, can be express as:

$$\widetilde{sh}_{L,iT_i}=sh_{L,i0_i}\left(1-sh_{K,i0_i}\left[\rho_i\left(\alpha_{K,i}-\alpha_{L,i}\right)+\rho_i\sum_{s=1}^{T_i}\lambda_{K,i}f_{K,s}-\lambda_{L,i}f_{L,s}\right]\right) \tag{14}$$

where $sh_{L,i0_i}$ is the value of the labor share in country i in the initial year available in the database and $\widetilde{sh}_{L,iT_i}$ is the counterfactual value of the labor share generated by the increase in the bias in the technical change. Note that

technologies $F_{L,t}$ and $F_{K,t}$ which are allowed to grow at a non constant rate $f_{L,t}$ and $f_{K,t}$.The growth rate of labor and capital augmenting technology can be written as in (11) and (12), where $\alpha_{L,i}=\eta_{L,i}g_{L,i}$ and $\alpha_{K,i}=\eta_{K,i}g_{K,i}$.

[19]Note that I do not consider the structural models of directed technical change to impose a parametric structure over the technology, but the reduced form model is flexible enough to capture features of these models that can be tested after the estimation.

[20]There are two differences from the panel model in Bai (2009). The first one is that in Bai (2009), the model is linear whereas in my model the factor enters in a nonlinear way inside the CES. The second difference is that Bai (2009) focuses on a common coefficient, whereas in my model I have random coefficients (the elasticity of substitution is heterogeneous across countries).

the object in 14 is an average effect in the time series dimension. In the same way, I can calculate the cumulative contribution of the increase in the capital-labor ratio to the labor share, holding the technology as constant.

$$\tilde{\tilde{sh}}_{L,iT_i} = sh_{L,it_{0i}} \left(1 - sh_{K,it_{0i}} \left[\rho_i \left(\frac{K_{iT_i} - K_{i0_i}}{K_{i0_i}} - \frac{L_{iT_i} - L_{i0_i}}{L_{i0_i}}\right)\right]\right)$$

where $\tilde{\tilde{sh}}_{L,iT_i}$ is the counterfactual value of the labor share generated by the increase in the capital-labor ratio.

Other objects of interest are the ones related to the models of directed technical change. These models suggest that the direction of the technical change A_L/A_K depends on the relative endowment of human capital and physical capital L/K and the elasticity of substitution. Using the ingredients of my model, I can carry out two different type of regressions in order to test the implications of these models. The first regression exploits the time series variation in the growth rate of the capital- and labor- augmenting technologies.

$$f_{L,t} - f_{K,t} = \beta \left[\frac{1}{N}\sum_{i=1}^{N} \rho_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right] + \varepsilon_t \tag{15}$$

where $f_{L,t} - f_{K,t}$ is the difference between the common factor of the labor- augmenting technology and the common factor of the capital- augmenting technology in period t. $\left[\frac{1}{N}\sum_{i=1}^{N} \rho_i^j \left(\gamma_{L,it} - \gamma_{K,it}\right)\right]$ is the average across countries of the cross product of ρ_i and the difference between the observed growth rate of the human capital and the observed growth rate of the physical capital in country i for period t. The object of interest is the OLS estimator of equation 15:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} \left(f_{L,t} - f_{K,t}\right) \left(\frac{1}{N}\sum_{i=1}^{N} \rho_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right)}{\sum_{t=1}^{T} \left[\frac{1}{N}\sum_{i=1}^{N} \rho_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right]^2}$$

The second regression exploits the cross-country variation of the growth rate of the capital- and labor- augmenting technology:

$$\bar{\gamma}_{A_{L,i}} - \bar{\gamma}_{A_{K,i}} = \beta\rho_i \left(\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}\right) + \varepsilon_i \tag{16}$$

where $\bar{\gamma}_{A_{L,i}} - \bar{\gamma}_{A_{K,i}} = \left(\alpha_{L,i} - \alpha_{K,i}\right) + {}^{1}\!/_{T} \left(\sum_{t=1}^{T} \lambda_{L,i} f_{L,t} - \lambda_{K,i} f_{K,t}\right)$ is the difference between the average annual growth rate of labor augmenting technology and the average annual growth rate capital augmenting technology in country i. $\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}$ is the difference between the average annual growth rate of the human capital and the average annual growth rate of the physical capital growth rate in country i. The object of interest is the OLS estimator of equation 16:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \left(\bar{\gamma}_{A_{L,i}} - \bar{\gamma}_{A_{K,i}}\right) \left(\rho_i \left[\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}\right]\right)}{\sum_{i=1}^{N} \left(\rho_i \left[\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}\right]\right)^2}$$

4 Identification and Estimation

The empirical model I take to the data is the supply system with country-specific elasticity of substitution with a factor model structure for the growth rate of the capital- and the labor- augmenting technical change, as described in section 3.

$$
\begin{aligned}
log(Y_{it}) \quad = \quad & \frac{1}{\rho_i} log\Big[\delta_i \left(A_{K,i0}\, exp\left(\sum_{s=0}^{t} \alpha_{K,i} + \lambda_{K,i} f_{K,t}\right) K_{it}\right)^{\rho_i} \\
& + (1-\delta_i)\left(A_{L,i0}\, exp\left(\sum_{s=0}^{t} \alpha_{L,i} + \lambda_{L,i} f_{L,t}\right) L_{it}\right)^{\rho_i}\Big] + \varepsilon_{Y,it}
\end{aligned} \tag{17}
$$

$$
log(sh_{L,it}) = \rho_i log\,(1-\delta_i) + \rho_i log\left(A_{L,i0}\right) + \rho_i \left(\sum_{s=0}^{t} \alpha_{L,i} + \lambda_{L,i} f_{L,t}\right) + \rho_i log\left(\frac{L_{it}}{Y_{it}}\right) + \varepsilon_{W,it} \tag{18}
$$

$$
log(sh_{K,it}) = \rho_i log\,(\delta_i) + \rho_i log\left(A_{K,i0}\right) + \rho_i \left(\sum_{s=0}^{t} \alpha_{K,i} + \lambda_{K,i} f_{K,t}\right) + \rho_i log\left(\frac{K_{it}}{Y_{it}}\right) + \varepsilon_{R,it} \tag{19}
$$

Equations 17-19 is an unbalanced nonlinear panel system with unobserved heterogeneity in both the time series and the cross sectional dimensions. Equation 17, the log of the CES production function, is a non-standard panel equation with two unobserved factors and country fixed effects, both entering non-linearly. Equation 18 -the log of the labor share- and equation 19 -the log of the capital share- are linear regression models with country fixed effects, a one factor model and a random coefficient. Note that equations 18 and 19 in first differences are similar to the linear panel model with interactive fixed effects studied in Bai (2009) and Moon and Weidner (2010*a*). The difference is the presence of random coefficients in my model.

Normalization. As was emphasized in Klump and de La Grandville (2000) and de La Grandville and Solow (2005) it is necessary to fix benchmark values for the level of production, factor inputs and factor income shares in order to have a meaningful interpretation of the elasticity of substitution and the growth rate of the capital- and labor- augmenting technology.[21] Moreover, León-Ledesma, McAdam and Willman (2010) show the superiority of estimates a normalized supply system as opposed to estimate a non-normalized one. Following León-Ledesma, McAdam and Willman (2010), I set values for the distribution parameter in each country to be equal to the capital share in a reference point in time t_{REF} , i.e $\delta_i = sh_{K,it_{REF}}$. The labor- augmenting technology and the capital-augmenting technologies for the reference year are set to be equal to the labor-output ratio and the capital-

[21]The importance of explicitly normalizing CES functions was discovered by La Grandville (1989), further explored by Klump and La Grandville (2000), La Grandville and Solow (2006), and first implemented empirically by Klump, McAdam and Willman (2007). Normalization starts from the observation that a family of CES functions whose members are distinguished only by different elasticities of substitution need a common benchmark point. Since the elasticity of substitution is originally defined as point elasticity, one needs to fix benchmark values for the level of production, factor inputs and for the marginal rate of substitution, or equivalently for per capita production, capital deepening and factor income shares.

output ratio in the reference year, i.e $A_{L,it_{REF}} = \frac{L_{it_{REF}}}{Y_{it_{REF}}}$ ($A_{K,it_{REF}} = \frac{K_{it_{REF}}}{Y_{it_{REF}}}$). Hence, the elasticity of substitution and the growth rate of the capital- and labor- augmenting technology are identified from the growth rate of the labor share, output, physical capital and human capital. I use $t_{REF} = 1995$ for two reasons . First, given that the factor $f_{K,t}$ and $f_{L,t}$ are identified from the cross sectional variation, the reference point has to be common for all the countries in the panel (my panel is unbalanced and for many countries the year 1995 is the first year of the sample). Second, the data on output, physical capital and human capital in the EU KLEMS database is constructed using the year 1995 as the reference point.

Before explaining the estimation procedure, I define three groups of variables:

- Endogenous observable variables:

$$X_{it} = \left\{ log(Y_{it}),\ log(sh_{K,it}),\ log(sh_{L,it}) \right\} \ \ i = 1..N,\ \ t = 1...T$$

- Exogenous observable variables:

$$W_{it} = \{K_{it},\ L_{it}\} \ \ i = 1..N,\ \ t = 1...T$$

- Country-specific latent variables:

$$Z_i = \left\{ \rho_i,\ ,\alpha_{L,i}\ ,\alpha_{K,i},\ \lambda_{K,i},\ \lambda_{L,i} \right\} \ \ i = 1..N$$

- Time-specific latent variables:

$$F_t = \left\{ f_{L,t},\ f_{K,t} \right\} \ \ t = 1...T$$

4.1 Identification issues

In this subsection I discuss the identification of the relevant parameters of the model. For simplicity I do the analysis focusing on the labor share equation, since this equation is linear in logs. Taking first differences in equation 7 yields:

$$\Delta log(sh_{L,it}) = \rho_i\alpha_{L,i} + \rho_i\lambda_{L,i}f_{L,t} + \rho_i\triangle log\left(\frac{L_{it}}{Y_{it}}\right) + \triangle\varepsilon_{W,it} \tag{20}$$

Equation 20 is a linear panel model with country fixed effects $\rho_i\alpha_{L,i}$, a factor model structure $\rho_i\lambda_{L,i}f_{L,t}$ and random coefficients $\rho_i\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$. The country-specific latent variables (i.e., $\alpha_{L,i}$, $\lambda_{L,i}$ and ρ_i) are identified from the time series dimension, while the time-specific latent variable (i.e., $f_{L,t}$) is identified from the cross sectional dimension. An important assumption for identification is that once we control for the unobserved process of the technology by $\gamma_{L,it} = \alpha_{L,i} + \lambda_{L,i}f_{L,t}$ there is only classical measurement error in $\varepsilon_{W,it}$ that is uncorrelated with

$\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$.[22] Deviations of the growth rate of technology from the process $\gamma_{L,it} = \alpha_{L,i} + \lambda_{L,i} f_{L,t}$ are considered to be orthogonal to $\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$.

To identify ρ_i I only require sufficient time series variation in $\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$ beyond the variation in the factor $f_{L,t}$ for each country. For simplicity, consider the case where $f_{L,t}$ were observable. In this case each ρ_i will be identified from a time series regression of $\Delta log(sh_{L,it})$ on $\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$ controlling by $f_{L,t}$ and an intercept, under the usual condition $X_i'\left(I_{T_i} - F\left(F'F\right)^{-1}F'\right)X_i$ is full rank, where X_i is the $T_i x1$ vector that contains all the time series of $\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$ and F is the $T_i x1$ vector that contains the values of $f_{L,t}$ in the period for which I have available information on $\Delta log(sh_{L,it})$ and $\triangle log\left(\frac{L_{it}}{Y_{it}}\right)$ in country i.[23]

Once I identify ρ_i the identification of each of the other components requires the same conditions as in a standard factor model. In this particular case with only one unobserved factor and fixed effects, the model needs two restrictions. Under large N and large T, the cross-sectional covariance matrix of $\triangle\varepsilon_{W,it}$ or the time series covariance matrix can be of an unknown form. However, the correlation—either cross sectional or serial—must be weak, which we assume to hold.

4.2 Estimation

The estimation of the supply system in model 17-19 is challenging given the non-linearity and the amount of unobserved heterogeneity allowed, both in the cross-section and the time series dimensions. One possibility to estimate the model is to follow a fixed effects approach, that is, treating each of the country- or time- specific effects as parameters and estimate them via GMM. Another possibility is a correlated random effects approach, that is, completing the model with distributional assumptions for the country- and time-specific latent variables and computing estimates directly from the marginal likelihood. Each of these two approaches offers different advantages but also suffer from some drawbacks in a setup like my model.

Statistical properties. The fixed effects approach has the advantage of generality as it is expected to be consistent as N and T go to infinity. Nevertheless the large- N and large- T statistical properties of the fixed effects estimator in set-ups with country- and time- specific effects have only been studied in linear models with additive effects and interactive effects (Bai, 2009; Moon and Weidner, 2010*b*), and in some specific nonlinear models (Fernández-Val and Weidner, 2013). My model does not belong to any of these cases, therefore I neither know how well the estimator will behave in this particular case, nor do I know how to do inference in this set-up. On the other hand, the correlated random effects approach provides a parsimonious way to do inference by relying directly on the likelihood. However, the classical random effects estimator is based on the parametric distribution of the latent variables. If the distributional assumptions of the latent variables do not hold, the classical random effects estimator of average effects is not consistent, even when N and T go to infinity, contrary to the classical fixed effects estimator.

[22] The case of measurement error in the inputs (labor and capital) are consider in other version of the paper in a parametric way.

[23] As is emphasized in Bai (2009), since $f_{L,t}$ is not observable and is estimated, a stronger condition is required.

Computation. The fixed effects estimator of this model is computationally problematic due to the non-linearity and the high dimensionality of the model.[24] The correlated random effects approach provides a computationally tractable way of estimating the model by using Markov Chain Monte Carlo with data augmentation.

4.3 Bayesian Fixed Effects Approach

Alternatively, I consider an approach that lies in between the two approaches discussed above, known as the Bayesian fixed effects approach (BFE). The BFE also specifies distributions for the latent variables, as in the random effects approach, but they are used as priors in a Bayesian fashion. These priors are combined with data to form posterior distributions which are then used in obtaining estimates of average effects in my model. This is in contrast to the random effects approach where priors are not updated since they are taken as the truth. For example, consider the BFE estimator of the average of the elasticities of substitution across countries $M = \frac{1}{N}\sum_{i=1}^{N}\rho_i$, discussed in section 3.

$$\hat{M}_{BFE} = \int_{\rho_1}\cdots\int_{\rho_N}\left(\frac{1}{N}\sum_{i=1}^{N}\rho_i\right)p\left(\rho_1,\cdots,\rho_N \mid \mathbb{X}\right)d\rho_1\cdots d\rho_N$$

where the vector $\mathbb{X} = \{X_{11},\cdots\cdots,X_{NT},\ W_{11},\ldots,W_{NT}\}$ contains all the observable data for the cross section and the time series and $p(\rho_1,\cdots,\rho_N \mid \mathbb{X})$ is the joint posterior distribution of the elasticity of substitution given the data. Let me define the vectors $\boldsymbol{\rho^N} = \{\rho_1,\cdots,\rho_N\}$, $\boldsymbol{\alpha_K^N} = \{\alpha_{K1},\cdots\alpha_{KN}\}$, $\boldsymbol{\alpha_L^N} = \{\alpha_{L1},\cdots\alpha_{LN}\}$, $\boldsymbol{\lambda_K^N} = \{\lambda_{K1},\cdots\lambda_{KN}\}$, $\boldsymbol{\lambda_L^N} = \{\lambda_{L1},\cdots\lambda_{LN}\}$ and $\boldsymbol{f_K^T} = \{f_{K1},\cdots f_{KT}\}$, $\boldsymbol{f_L^T} = \{f_{L1},\cdots f_{LT}\}$

$$\begin{aligned} p\left(\boldsymbol{\rho^N} \mid \mathbb{X}\right) = \quad & p\left(\boldsymbol{\rho^N};\Theta_\rho\right)\int\cdots\int\left[\prod_{i=1}^{N}\prod_{t=1}^{T}p\left(X_{it},W_{it} \mid \rho_1,\rho_2,\cdots\lambda_{K1},\cdots f_{LT}\right)\right. \\ & \times p\left(\boldsymbol{\alpha_K^N};\Theta_{\alpha_K}\right)\times p\left(\boldsymbol{\alpha_L^N};\Theta_{\alpha_L}\right)\times p\left(\boldsymbol{\lambda_K^N};\Theta_{\lambda_K}\right)\times p\left(\boldsymbol{\lambda_L^N};\Theta_{\lambda_L}\right) \\ & \left.\times p\left(\boldsymbol{f_K^T};\Theta_{f_K}\right)\times p\left(\boldsymbol{f_L^T};\Theta_{f_L}\right)\right]d\lambda_{K1}\cdots dF_{LT} \end{aligned} \tag{21}$$

where $\prod_{i=1}^{N}\prod_{t=1}^{T}p\left(X_{it},W_{it} \mid \rho_1,\rho_2,\cdots\lambda_{K1},\cdots f_{LT}\right)$ is the likelihood function of the supply system model and $p\left(\boldsymbol{\rho^N};\Theta_\rho\right)$ is a joint parametric distribution for the elasticities of substitution, which is fully characterized by the parameter vector Θ_ρ. The functions $p\left(\boldsymbol{\alpha_K^N};\Theta_{\alpha_K}\right)$, $p\left(\boldsymbol{\alpha_L^N};\Theta_{\alpha_L}\right)$, $p\left(\boldsymbol{\lambda_K^N};\Theta_{\lambda_K}\right)$, $p\left(\boldsymbol{\lambda_L^N};\Theta_{\lambda_L}\right)$, $p\left(\boldsymbol{f_K^T};\Theta_{f_K}\right)$ and $p\left(\boldsymbol{f_L^T};\Theta_{f_L}\right)$ are the parametric distributions of the rest of the latent variables.

The posterior distribution in 21 takes the distributions of the latent variables as prior information and updates them with the data. The likelihood contributes to the posterior with NT elements. In the case of independent

[24] In fact, I extend the iterated OLS proposed by (Bai, 2009) to a model with random coefficients and estimate equations 18 and 19 in first differences. However, when I introduce the CES production function, I obtain numerous local minima for the optimal parameter values, as the nonlinear OLS in this setup is very sensitive to initial values.

priors in the cross section and a Markov process in the prior of the factors, the contribution of the priors will be $5N + 2T$ elements. Therefore, as the sample size increases in both dimensions, the posterior will learn from the data and the contribution of the priors will become negligible. As a result, the estimator will be consistent when both N and T go to infinity even when the priors are misspecified. The intuition follows the work by Arellano and Bonhomme (2009) for a nonlinear panel model with individual unobserved heterogeneity. They prove the consistency of the BFE of average effects when T goes to infinity in non linear models with individual specific effects, even when the parametric model of the individual effects is misspecified. It is possible to extend the same analysis for consistency in setups with only time effects when N goes to infinity, assuming a Markov process for the time effects. If the sample size is small and the priors are misspecified, there will be a bias which will disappear as the sample size increases, as in the pure fixed effects approach. The BFE estimator has similar asymptotic properties as the fixed effects estimator but may perform better in estimation in small samples, as long as the parametric distributions of the latent variables are well specified.

Distributional assumptions

The setup here is that of a middle-sized unbalanced panel where N and T are of similar magnitude. Despite the fact that the time series and the cross section are potentially informative, neither of the two dimensions is very large. For example, there are some countries with just 15 years of data. As a consequence, a fixed effects estimation approach will suffer from incidental parameter bias and finite sample variances.[25] In contrast the BFE can be expected to reduce sampling variability thanks to using informative priors.[26]

Time-specific latent variables. The factors follow an AR(1) process with a stationary Normal distribution, allowing for possible dynamics in the growth rate of the technology. As in the correlated random effect of Mundlak (1978) and Chamberlain (1982), I also condition the factors on the observable explanatory variables of the model. Given the theoretical relationship between technologies and factor endowment discussed in section 2.4, I allow for possible correlation between the factors and the cross-country average of the growth rate of physical capital and human capital.

$$\begin{aligned} f_{K_t} &\sim N\left(\beta_{0,f_K} + \beta_{1,f_K}\bar{\gamma}_{K,t} + \phi_{f_K} f_{K_{t-1}},\ \sigma^2_{f_K}\right) \\ f_{L_t} &\sim N\left(\beta_{0,f_L} + \beta_{1,f_L}\bar{\gamma}_{L,t} + \phi_{f_L} f_{L_{t-1}},\ \sigma^2_{f_L}\right) \end{aligned} \tag{22}$$

where $\bar{\gamma}_{K,t} = \frac{1}{N}\sum_i^N [logK_{it} - logK_{it-1}]$ and $\bar{\gamma}_{L,t} = \frac{1}{N}\sum_i^N [logL_{it} - logL_{it-1}]$. The hyper-parameters in $\Theta_{f_K} =$

[25] This problem is augmented by the interaction of the country- and time- specific effects and the nonlinearity of the model. Hence, the bias in the estimation of each latent variable contaminates the estimation of the other latent variables. Given that I am working with an unbalanced panel, the fixed effects estimators is affected by a bias of order $min(T_{MIN}, N_{MIN})$, where T_{MIN} is the number of time series observations of the country with the smallest sample size, whereas N_{MIN} is the number of countries for the year with less countries.

[26] In a Bayesian normal linear model set-up, the fixed effects estimator uses uninformative priors that can be thought of as a limiting case of a informative prior whose precision goes to zero.

$\left\{\beta_{0,f_K},\, \beta_{1,f_K},\, \phi_{f_K},\, \sigma^2_{f_K}\right\}$ and $\Theta_{f_L} = \left\{\beta_{0,f_L},\, \beta_{1,f_L},\, \phi_{f_L},\, \sigma^2_{f_L}\right\}$ are estimated.

Country-specific latent variables. The priors of the country specif-latent variables of the capital- and labor-augmenting technology follow a Normal distribution and are allowed to be correlated with the growth rate of physical capital and human capital, for the same reason discussed above. The prior of ρ_i follow a Normal distribution with mean $\beta_{0,\rho}$ and variance σ^2_ρ. The mean and the variance of this distribution will be estimated.

$$
\begin{aligned}
\lambda_{Li} &\sim N\left(\beta_{0,\lambda_L} + \beta_{1,\lambda_L}\bar{\gamma}_{L,i},\ \sigma^2_{\lambda_L}\right) \\
\lambda_{Ki} &\sim N\left(\beta_{0,\lambda_K} + \beta_{1,\lambda_K}\bar{\gamma}_{K,i},\ \sigma^2_{\lambda_K}\right) \\
\alpha_{Li} &\sim N\left(\beta_{0,\alpha_L} + \beta_{1,\alpha_L}\bar{\gamma}_{L,i},\ \sigma^2_{\alpha_L}\right) \\
\alpha_{Ki} &\sim N\left(\beta_{0,\alpha_K} + \beta_{1,\alpha_K}\bar{\gamma}_{K,i},\ \sigma^2_{\alpha_K}\right) \\
\rho_i &\sim N\left(\beta_{0,\rho},\ \sigma^2_\rho\right)
\end{aligned}
\tag{23}
$$

where $\bar{\gamma}_{K,i} = \frac{1}{T-1}\sum_{t=2}^{T}\left[logK_{it} - logK_{it-1}\right]$ and $\bar{\gamma}_{L,i} = \frac{1}{T-1}\sum_{t=2}^{T}\left[logL_{it} - logL_{it-1}\right]$. The hyper-parameters in $\Theta_{\lambda_L} = \left\{\beta_{0,\lambda_L},\, \beta_{1,\lambda_L},\, \phi_{\lambda_L},\, \sigma^2_{\lambda_L}\right\}$, $\Theta_{\lambda_K} = \left\{\beta_{0,\lambda_K},\, \beta_{1,\lambda_K},\, \phi_{\lambda_K},\, \sigma^2_{\lambda_K}\right\}$, $\Theta_{\alpha_L} = \left\{\beta_{0,\alpha_L},\, \beta_{1,\alpha_L},\, \phi_{\alpha_L},\, \sigma^2_{\alpha_L}\right\}$, $\Theta_{\alpha_K} = \left\{\beta_{0,\alpha_K},\, \beta_{1,\alpha_K},\, \phi_{\alpha_K},\, \sigma^2_{\alpha_K}\right\}$, $\Theta_\rho = \left\{\beta_{0,\rho},\, \beta_{1,\rho},\, \phi_\rho,\, \sigma^2_\rho\right\}$ are estimated.

The normality assumption is very useful for computation, since it allows to derive closed form solutions for the posterior distributions as it is discusses in the next section. As a robustness check I try with normal distributions without conditioning on the explanatory variables and also using uninformative priors. I also estimate equations 18 and 19 using a fixed effects approach by extending the iterated OLS algorithm proposed by Bai (2009) to a model with random coefficients.

Measurement errors. The measurement errors in equations 17-19 are assumed to follow a normal distribution $N\left(0,\ \sigma^2_Y\right)$, $N\left(0,\ \sigma^2_W\right)$ and $N\left(0,\ \sigma^2_R\right)$. I start assuming that the three measurement errors are uncorrelated across time and across countries. Additionally, I also assume independence across the disturbances in the three equations. All of these assumptions could be relaxed later, including the possibility of spatial correlation in the error terms but in a parametric way (correlation across some countries beyond the ones captured by the factors), and country specific variances for all the shocks σ^2_{Yi}, σ^2_{Wi} and σ^2_{Ri} .

5 Computation

In this section I show how to compute the posterior distributions of the latent variables of my model. The objects of interest are moments of the posterior distribution of the latent variables in $\boldsymbol{Z^N} = \{Z_1, \ldots Z_N\}$ and $\boldsymbol{F^T} = \{F_1, \ldots, F_T\}$, given the model in 17-19 and the data in $\mathbb{X}$:

$$p\left(\boldsymbol{Z^N},\boldsymbol{F^T}\mid\mathbb{X}\right) \tag{24}$$

By Bayes rule:

$$p\left(\boldsymbol{Z^N},\boldsymbol{F^T}\mid\mathbb{X}\right)\propto\mathcal{L}\left(\mathbb{X}\mid\boldsymbol{Z^N},\boldsymbol{F^T}\right)\times\mathcal{H}(\boldsymbol{Z^N})\times\mathcal{H}(\boldsymbol{F^T}) \tag{25}$$

where $\mathcal{L}\left(\mathbb{X}\mid Z_1,\ldots Z_N,\ F_1,\ldots,F_T\right)$ denotes the likelihood function of the observables variables given the latent variables. The distributions $\mathcal{H}(\boldsymbol{Z^N})$ and $\mathcal{H}(\boldsymbol{F^T})$ are the join marginal distribution of the country-specific latent variables and the time-specific latent variable, respectively. Equation 25 can be expressed in terms of the (i) supply system model 17-19 and (ii) the priors for the latent variables defined in the previous section.

$$\begin{aligned}p\left(\boldsymbol{Z^N},\boldsymbol{F^T}\mid\mathbb{X}\right) \quad\propto\quad & \prod_{i=1}^{N}\prod_{t=1}^{T}p\left(X_{it}\mid W_{it},\ Z_i,\ F_t\right)\\ & \times p\left(Z_i\mid W_i^T,\Theta_Z\right)\times p\left(F_t\mid F_{t-1},\ \boldsymbol{W_t^N},\Theta_F\right)\end{aligned}$$

where $\boldsymbol{W_t^N}=\{W_{1t},\ldots W_{Nt}\}$, $\boldsymbol{W_i^T}=\{W_{i1},\ldots W_{iT}\}$ and $\prod_{i=1}^{N}\prod_{t=1}^{T}p\left(X_{it}\mid W_{it},\ Z_i,\ F_t\right)$ is the conditional likelihood of the endogenous variables in the supply system model given the exogenous observable variables in W_{it} and the latent variables in $\{Z_i\}$ and $\{F_t\}$. The distributions $p\left(Z_i\mid W_i^T,\Theta_Z\right)$ and $p\left(F_t\mid F_{t-1},\ \boldsymbol{W_t^N},\Theta_F\right)$ are the prior distributions for the latent variables given the exogenous variables in W_{it}.[27]

In order to sample from the target distribution $p\left(\boldsymbol{Z^N},\boldsymbol{F^T}\mid\mathbb{X}\right)$ we first need to estimate the value of the hyper parameters in Θ_Z and Θ_F. Then, we can sample from the distribution of the latent variable $p\left(Z_i\mid W_i^T,\hat{\Theta}_Z\right)$ and $p\left(F_t\mid F_{t-1},\ \boldsymbol{W_t^N},\hat{\Theta}_F\right)$ and update the information in a Bayesian fashion, evaluating the values of the latent variables drawn in the likelihood $\prod_{i=1}^{N}\prod_{t=1}^{T}p\left(X_{it}\mid W_{it},\ Z_i,\ F_t\right)$. A classical approach to estimate the parameters in Θ_W and Θ_F is to maximize the average likelihood function. In my setup we have to take into account that I am working with a "double random effect model" because we have prior distributions in both dimensions, therefore we have to maximize a double log-average likelihood function:

$$\begin{aligned}\left\{\hat{\Theta}_Z,\hat{\Theta}_F\right\}\equiv\underset{\Theta_Z,\Theta_F}{\operatorname{argmax}}\Bigg\{\sum_{i=1}^{N}log\int\Bigg[\int\ldots\int\prod_{t=1}^{T}p\left(logsh_{it}^K,\,logsh_{it}^L,\,logY_{it}\mid K_{it},L_{it},Z_i,F_t\right)\\ \times\,p(F_t\mid F_{t-1},K_{t1},\ldots,K_{tN},L_{t1},\ldots L_{tN},\ \Theta_F)dF_1\ldots dF_T\Bigg]\\ \times\,p(Z_i\mid K_{i1},\ldots,K_{iT},L_{i1},\ldots L_{iT},\ \Theta_Z)dZ\Bigg\}\end{aligned} \tag{26}$$

[27] I assume that the inputs are exogenous once I have controlled for the labor- and capital- augmenting technologies using the factor model.

Solving equation 26 requires computing integrals with respect to each of the elements in Z_i and the elements in F_t. There is no closed form solution for this equation so it has to be approximated numerically. A simulation method such as importance sampling could be used to compute the integrals.

An alternative computationally efficient way of getting draws from the target distribution of the model is Markov Chain Monte Carlo (MCMC), which does not involves the maximization of the likelihood. As emphasize by Arellano and Bonhomme (2011) there is a connection between the classical random effect estimation approach and the Bayesian approach, as I am estimating the posterior distribution of the unobservables. The difference in estimation comes from the way I treat Θ_Z and Θ_F. To turn Bayesian and apply MCMC methods I define prior distributions for the hyper parameters in Θ_Z and Θ_F. [28]

5.1 MCMC and Two Step Estimation

Blocking. The MCMC deal with the simulation of high dimensional probability distributions. The idea behind this method is to sample from a given distribution, by constructing a suitable Markov chain with the property that its limiting, invariant distribution is the target distribution. Since in this application the dimension of the target distribution $p\left(\boldsymbol{Z^N}, \boldsymbol{F^T}, \Theta_Z, \Theta_F \mid \mathbb{X}\right)$ is quite large, I construct the Markov Chain simulation by sampling from the conditional posterior distribution of each latent variable:

[28] I use independent flat priors for the hyper parameter vector Θ_Z and Θ_F in order to obtain similar results for the posterior distribution as I had estimated by maximizing the double integrated likelihood in 26. The conditional posterior of Θ_Z and Θ_F will be a Normal distribution, just as in the linear Bayesian normal model.

Algorithm 1 Multiple-block MCMC

1. **Specify an initial value** $\left\{Z_1^0, \ldots, Z_N^0, F_1^0, \ldots F_T^0\right\}$
2. **Repeat for** $j = 1, 2 \ldots\ldots M$
   - Block Z_i: Repeat for $i = 1, 2, \ldots N$
     - Sub-block ρ_i
       * Sample $\rho_i^j \sim p\left(\rho_i \mid \mathbb{X},\ F_1^{j-1}, \ldots, F_T^{j-1},\ \lambda_{K,i}^{j-1},\ \lambda_{L,i}^{j-1},\ \alpha_{L,i}^{j-1},\ \alpha_{K,i}^{j-1},\ \Theta_Z^{j-1}\right)$
     - Sub-block $\lambda_{L,i}$
       * Sample $\lambda_{L,i} \sim p\left(\rho_i \mid \mathbb{X},\ F_1^{j-1}, \ldots, F_T^{j-1},\ \rho_i^j,\ \lambda_{K,i}^{j-1},,\ \alpha_{L,i}^{j-1},\ \alpha_{K,i}^{j-1},\ \Theta_Z^{j-1}\right)$
     - Sub-block for the other elements in Z_i
   - Block F_t: Repeat for $t = 2, \ldots T$
     - Sub-block $f_{L,t}$
       * Sample $f_{L,t}^j \sim p\left(f_{L,t} \mid \mathbb{X},\ f_{L,t-1}^j,\ f_{Lt+1}^{j-1}, Z_1^j, \ldots Z_N^j,\ \Theta_F^{j-1}\right)$
     - Sub-block $f_{K,t}$
       * Sample $f_{K,t}^j \sim p\left(f_{K,t} \mid \mathbb{X},\ f_{K,t-1}^j,\ f_{Kt+1}^{j-1}, Z_1^j, \ldots Z_N^j,\ \Theta_F^{j-1}\right)$
   - Block Θ_Z and Θ_F
     - Sample from $\Theta_Z^j \sim p\left(\Theta_Z \mid \mathbb{X},\ Z_1^j, \ldots Z_N^j\right)$
     - Sample from $\Theta_F^j \sim p\left(\Theta_F \mid \mathbb{X},\ F_1^j, \ldots F_N^j\right)$

In the algorithm 1, the country-specific latent variables in Z_i do not depend on the country-specific latent variables in Z_j, because given the time-latent variables and observable variables, they are independent across countries. Within each country-specific block, there are 5 sub-blocks. There is one sub-block for each of the latent variable in $Z_i = \left\{\rho_i,\ ,\alpha_{L,i}\ ,\alpha_{K,i},\ \lambda_{K,i},\ \lambda_{L,i}\right\}$.

The time-specific latent variables in F_t depends on F_{t-1} and F_{t+1} given the Markov process of the factors (see equation 22). Within each time-specific block, there are two sub-blocks for the conditional posterior distribution of each of the elements in $F_t = \left\{f_{L,t},\ f_{K,t}\right\}$.

Given that I can not derive close form solutions for each of the conditional posterior distributions, I will draw samples from each sub-block using the Metropolis Hasting algorithm (M-H).[29] In the M-H, to sample from the conditional distribution in each block I need to define (i) a starting value for the unknowns in each block and (ii) a proposal distribution for each of the blocks. In each block, I draw from the proposal distribution and compare the posterior distribution evaluated at this draw against the posterior distribution evaluated at the value of the parameter in the previous iteration.

[29]The decision to use a specific algorithm -Gibbs sampler or Metropolis Hasting (M-H)- depends on the specificity of the problem. I can use the Gibbs sampling if I can derive a close form solution for the conditional distributions in each block. This is not the case in the supply system model, since the non linearity of the CES prevents from deriving a known conditional posterior distribution in each block.

Example: sub-block ρ_i

In order to sample ρ in each iteration j from the conditional posterior distribution:

$$\rho_i^j \sim p\left(\rho_i \mid \mathbb{X}, F_1^{j-1}, \ldots, F_T^{j-1}, \lambda_{K,i}^{j-1}, \lambda_{L,i}^{j-1}, \alpha_{L,i}^{j-1}, \alpha_{K,i}^{j-1}, \Theta_Z^{j-1}\right)$$

which do not have a closed form expression, I do the following procedure:

1. **Sample from a proposal distribution:**
   $\rho_i^{'} \sim q\left(\rho_i \mid \rho_i^{j-1}, \mathbb{X}, F_1^{j-1}, \ldots, F_T^{j-1}, \lambda_{K,i}^{j-1}, \lambda_{L,i}^{j-1}, \alpha_{L,i}^{j-1}, \alpha_{K,i}^{j-1}, \Theta_Z^{j-1}\right)$**,**
2. **Calculate**
   $\alpha_1 = min\left\{\frac{p\left(\rho_i^{'}|\ldots\right)q\left(\rho_i^{'}|\rho_i^{j-1},\ldots\right)}{p\left(\rho_i^{j-1}|\ldots\right)q\left(\rho_i^{j-1}|\rho_i^{'},\ldots\right)}, 1\right\}$
3. **Set:**
   $\rho_i^j = \rho_i^{'}$ **if** $Unif\,(0,1) \leq \alpha_1$ **or** $\rho_i^j = \rho_i^{j-1}$**, otherwise.**

where $q\left(\rho_i \mid \rho_i^{j-1}, \mathbb{X}, F_1^{j-1}, \ldots, F_T^{j-1}, \lambda_{K,i}^{j-1}, \lambda_{L,i}^{j-1}, \alpha_{L,i}^{j-1}, \alpha_{K,i}^{j-1}, \Theta_Z^{j-1}\right)$ is a "proposal distribution" for ρ_i. This distribution may depend (or not) on the previous value of the "owner" of the block, (i.e., ρ_i^{j-1}), the other latent variables and the data.

Essentially, the M-H algorithm is evaluating each of the proposal values $\rho_i^{'}$ drawn from $q\,(.)$, using the conditional posterior distribution $p\,(.)$. Hence, the challenge in Metropolis Hasting is to define suitable proposal distributions to evaluate the conditional posterior distribution in each block. In order to estimate the supply system model with M-H, I need to set $5N + 2T + dim\,(\Theta_Z) + dim\,(\Theta_F)$ proposal distributions and starting values. To that end, I propose to divide the estimation in two steps and use the prices equations to define suitable proposal distributions and initial conditions for the M-H algorithm.

5.2 Two step approach

Step 1: Gibbs sampling from the price equations. The supply system model in 17-19 is an overidentified model, since all the latent variables in Z_i and F_t can be identified from either the price equations, or the nonlinear CES or both. Given that the price equations have a linear structure, I can derive closed form solutions for the conditional distribution in each block, as in the Bayesian Normal Linear Model (see Hirano (2002) and Kose, Otrok and Whiteman (2003)). The first step is to simulate from an "incomplete" model with a Gibbs sampling algorithm using only the price equations and get draws from the posterior distribution without setting any "proposal distribution". The target distribution in the first step is specified as:

$$
\begin{aligned}
p_{price}\left(\boldsymbol{Z^N}, \boldsymbol{F^T}, \Theta_Z, \Theta_F \mid \mathbb{X}\right) \quad \propto \quad & \prod_{i=1}^{N}\prod_{t=1}^{T} p\left(logsh_{it}^{K}, logsh_{it}^{L} \mid logY_{it}, W_{it}, Z_i, F_t\right) \\
& \times p\left(Z_i \mid \boldsymbol{W_i^T}, \Theta_Z\right) \times p\left(F_t \mid F_{t-1}, \boldsymbol{W_t^N}, \Theta_F\right) \\
& \times p\left(\Theta_Z\right) \times p\left(\Theta_F\right)
\end{aligned}
$$

$\prod_{i=1}^{N}\prod_{t=1}^{T} p\left(logsh_{it}^{K}, logsh_{it}^{L} \mid K_{it}, L_{it}, Z_i, F_t\right)$ is the likelihood of the model of the price equations (the labor share and capital share equations). Here I provide an example of how to construct the conditional posterior distribution of one of the sub-blocks for the Gibbs sampler algorithm. In an appendix, I derive the conditional posterior distribution for each block.

Example: Block of $f_{L,t}$

$$
\begin{aligned}
p\left(f_{L,t} \mid \mathbb{X}, f_{L,t-1}, f_{Lt+1}, \boldsymbol{Z^N}, \Theta_F\right) \quad \propto \quad & p\left(f_{L,t} \mid \boldsymbol{logsh_{L,t}^N}, \boldsymbol{logY_t^N}, \boldsymbol{W_t^N}, \boldsymbol{Z^N}\right) \\
& \times p\left(f_{L,t} \mid f_{L,t-1}, \boldsymbol{W_t^N}, \Theta_F\right) \times p\left(f_{L,t} \mid f_{L,t+1}, \boldsymbol{W_{t+1}^N}, \Theta_F\right)
\end{aligned}
$$

where $\boldsymbol{logsh_{L,t}^N} = \left\{logsh_{L,1t}, \ldots logsh_{L,Nt}\right\}$, $\boldsymbol{logY_t^N} = \{logY_{1t}, \ldots logY_{Nt}\}$. Given that the prior of $f_{L,t}$ is an $AR(1)$ process, the conditional posterior of $f_{L,t}$ will depend on the likelihood of the labor share model, the prior of $f_{L,t}$: $p\left(f_{L,t} \mid f_{L,t-1}, \boldsymbol{W_t^N}, \Theta_F\right)$ and the prior of $f_{L,t+1}$: $p\left(f_{L,t+1} \mid f_{L,t}, \boldsymbol{W_{t+1}^N}, \Theta_F\right)$ which also contains information of $f_{L,t}$. Hence, there are three equations that provide information about $f_{L,t}$

1. **Labor Share Equation**

$$
\Delta logsh_{L,it} = \rho_i\alpha_{L,i} + \rho_i\lambda_{L,i}f_{L,t} + \rho_i\triangle log\left(\frac{L_{it}}{Y_{it}}\right) + \triangle\varepsilon_{W,it}
$$

In the block of $f_{L,t}$ all the other latent variables are known:

$$
f_{L,t} \sim N\left(\left(\sum_{i=1}^{N}\rho_i^2\lambda_{L,i}^2\right)^{-1}\sum_{i=1}^{N}\left(\rho_i\lambda_{L,i}\left(\Delta logsh_{L,it} - \rho_i\alpha_{L,i} - \rho_i\triangle log\left(\frac{L_{it}}{Y_{it}}\right)\right)\right), \; 2\sigma_W^2\left(\sum_{i=1}^{N}\rho_i^2\lambda_{L,i}^2\right)^{-1}\right) \tag{27}
$$

2. **The prior distribution of f_{Lt}**

$$
f_{L,t} \sim N\left(\beta_{0,f_L} + \beta_{1,f_L}\bar{\gamma}_{L,t} + \phi_{f_L}f_{L,t-1}, \; \sigma_{f_L}^2\right)
$$

3. **The prior distribution of f_{Lt+1}**

$$f_{L,t+1} \sim N\left(\beta_{0,f_L} + \beta_{1,f_L}\bar{\gamma}_{L,t+1} + \phi_{f_L} f_{L,t},\ \sigma^2_{f_L}\right)$$

Re-arranging:

$$f_{L,t} \sim N\left(\frac{f_{L,t+1} - \beta_{0,f_L} - \beta_{1,f_L}\bar{\gamma}_{L,t+1}}{\phi_{f_L}},\ \frac{\sigma^2_{f_L}}{\phi^2_{f_L}}\right)$$

The conditional posterior is a combination of three normal distributions:

$$f_{L,t} \quad \sim \quad N\left(\mu_{f_l},\ H^{-1}_{f_L}\right) \tag{28}$$

where

$$H_{f_L} = \left(\frac{1}{2\sigma^2_W}\sum_{i=1}^{N}\rho_i^2\lambda^2_{L_i} + \frac{1}{\sigma^2_{f_L}} + \frac{\phi^2_{f_L}}{\phi^2_{f_L}}\right)$$

and

$$\mu_{f_l} = H^{-1}_{f_L}\left[\frac{1}{2\sigma^2_W}\sum_{i=1}^{N}\left(\rho_i\lambda_{L,i}\left(\Delta logsh_{L,it} - \rho_i\alpha_{L,i} - \rho_i\triangle log\left(\frac{L_{it}}{Y_{it}}\right)\right)\right)\right] + \\ H^{-1}_{f_L}\left[\frac{1}{\sigma^2_{f_L}}\cdot\left(\beta_{0,f_L} + \beta_{1,f_L}\bar{\gamma}_{L,t} + \phi_{f_L} f_{L,t-1}\right)\right] + H^{-1}_{f_L}\left[\frac{\sigma^2_{f_L}}{\phi^2_{f_L}}\frac{f_{L,t+1} - \beta_{0,f_L} - \beta_{1,f_L}\bar{\gamma}_{L,t+1}}{\phi_{f_L}}\right]$$

The conditional posterior distribution of $f_{L,t}$ in 28 can be seen as a panel data version of the standard Kalman filter. An important advantage of 28 over the time-series Kalman filter is that it does not rely solely on the parametric distribution of the latent variable, since the cross-sectional dimension is also informative. If N is large, the contribution of 27 to 28 will be more important than the prior.

Step 2: M-H from the complete supply system model. In the second step I will use the distributions simulated in the first step in order to set the proposal distributions used in the M-H in the estimation of the complete supply system model. The target distribution of the supply system is specified as

$$
\begin{aligned}
p_{supply}\left(Z_1,\ldots Z_N, F_1,\ldots,F_T,\Theta_Z,\Theta_F \mid \mathbb{X}\right) \quad \propto \quad & \prod_{i=1}^{N}\prod_{t=1}^{T} p\left(logY_{it} \mid W_{it}, Z_i,\, F_t\right) \\
& \times p\left(logsh_{it}^{K}, logsh_{it}^{L} \mid logY_{it}, W_{it}, Z_i, F_t\right) \\
& \times p\left(Z_i \mid \boldsymbol{W_i^T}, \Theta_Z\right) \times p\left(F_t \mid F_{t-1}, \boldsymbol{W_t^N}, \Theta_F\right) \\
& \times p\left(\Theta_Z\right) \times p\left(\Theta_F\right)
\end{aligned}
$$

where $\prod_{i=1}^{N}\prod_{t=1}^{T} p\left(logY_{it} \mid W_{it}, Z_i, F_t\right)$ is the likelihood of the CES production function. I consider two alternative proposal distributions that comes from the price equations:

1. The first alternative is to use directly the posterior distributions of the first step as the proposal distributions for the M-H in the second step. Doing this is exactly the same as using only the likelihood of the CES production function in the posterior distribution to evaluate the values from the proposal distribution, which comes from the price equations. The drawback of this approach is that in the case of departure from perfect competition, it could be the case that the proposals that come from the prices equations do not contain the real support of the latent variables, which implies that there will be regions of the latent variables that will not be explored by the algorithm in the second step.

2. The second possibility is to use a random walk M-H chain, the one used by Metropolis *et al.* (1953), and arguably the most popular in applications (Chib, 2001). In a random walk M-H, the candidate value for each parameter is drawn according to the process: $\psi' = \psi + \varepsilon$, where ψ' is the value of the parameter of the block draw from the random walk proposal distribution, ψ is the value of the parameter of the block in the previous iteration and ε is an innovation that follows a $N(0,\ \sigma)$. With this proposal we reduce the problem of determining a proposal distribution to one in which we only have to determine the variance of the proposal in each block. As usual one has to be careful in setting the variance of ε. If it is too large it is possible that the chain may remain stuck at a particular value for many iterations while if it is too small the chain will tend to make small moves and move inefficiently through the support of the target distribution. I will use the variance of the posterior distribution of each parameter simulated in the first step as the variance of the proposal distribution in each block.

5.3 Performance and Convergence of the algorithm

Geweke test of posterior simulations. The implementation of the algorithm discussed above requires both analytical derivation of conditional posterior distributions and computational coding. As emphasized by Geweke (2004) the data, the priors and other densities must correspond exactly to the model; the conditional distributions must be derived correctly; and the computer code that incorporates all the inputs of the algorithm must be free of error. In order to verify all these points, I implement a simple test of posterior simulations proposed by Geweke (2004) to detect coding and analytical errors in the algorithm. The idea of the test is to compare two simulation approximations of $E[g(\mathbb{Z},\ \mathbb{X})]$ by using two different ways of constructing the joint distribution $p(\mathbb{Z},\ \mathbb{X})$ of unobservables $\mathbb{Z} = \{Z_1, \ldots Z_N,\ F_1, \ldots, F_T,\ \Theta_Z,\ \Theta_F\}$ and observable variables $\mathbb{X}$. The first approximation employs the marginal-conditional simulator of the joint distribution, simulating the unobservables from the prior $p(\mathbb{Z})$ and then simulating the observables using the likelihood function of the model $p(\mathbb{X} \mid \mathbb{Z})$. The second one uses the successive-conditional simulator which simulates values for the observables using the likelihood of the model but simulates values for the unobservables using the posterior distribution of the model $p(\mathbb{Z} \mid \mathbb{X})$. If both simulators are error free, then as $M_1$and M_2 go to infinity:

$$t_{Geweke} = \left(\bar{g}^{(M_1)} - \tilde{g}^{(M_2)}\right) / \left(M_1^{-1}\hat{\sigma}_g + M_2^{-1}\hat{\tau}_g\right) \xrightarrow{d} N(0,\ 1)$$

where $\bar{g}^{(M_1)} = \frac{1}{M_1}\sum_{m=1}^{M_1} g(\mathbb{Z}(m),\ \mathbb{X}(m))$ is the sample analogue of $E[g(\mathbb{Z},\ \mathbb{X})]$, constructed using the M_1 simulations of $\mathbb{Z}$ and $\mathbb{X}$ that comes from the marginal-conditional simulator and $\tilde{g}^{(M_2)} = \frac{1}{M_2}\sum_{m=1}^{M_2} g(\mathbb{Z}(m),\ \mathbb{X}(m))$ is the sample analogue of $E[g(\mathbb{Z},\ \mathbb{X})]$, constructed using the M_2 simulations of $\mathbb{Z}$ and $\mathbb{X}$ that comes from the successive-conditional simulator. $\hat{\sigma}_g$ and $\hat{\tau}_g$ are the estimated variances of $\bar{g}^{(M_1)}$ and $\tilde{g}^{(M_2)}$, respectively. For simplicity I define the function $g(\mathbb{Z},\ \mathbb{X})$ to be equal to the sum of all the unobservables in $\mathbb{Z}$ and all the observables in $\mathbb{X}$. Using this function I get a t_{Geweke} equal to 0.12 and accept the null hypothesis of $\bar{g}^{(M_1)} = \tilde{g}^{(M_2)}$. In the calculation of the test I use one thousand iterations of each simulator $M_1 = M_2 = 10000$.

Convergence. In this subsection, I assess the performance of the sampling algorithm described above, using a carefully simulated data generating process. I use the real value of the level of capital K_{it} and labor L_{it} for a panel of 22 countries from the EU KLEMS database and I simulate values for the factor augmenting technologies and the elasticity of substitution using the factor model and the priors of the latent variables. With all these ingredients, I finally simulated the values for the dependent variables of the three equations of the supply system. Figure 1 and Figure 2 show the convergence to the stationary posterior distribution for one of the elasticity of substitutions ρ, starting from different arbitrary initial conditions.

6 Results

6.1 Data description

I use the EU KLEMS database, which is collected by the Groningen Growth and Development Center. The EU KLEMS is a cross-country database which contains industry-level measures of output, inputs and productivity for 25 European countries and Australia, Canada, Japan, Korea and the US from 1970 onwards. There are two main advantages of working with this database. First, it provides carefully constructed data on labor compensation, human capital and physical capital. The labor compensation considers the labor income of the self-employed, which is important to study the evolution of the labor share (see (Gollin, 2002)). The index of human capital takes into account heterogeneity across countries in levels of education, whereas the capital services index takes into account differences in the assets type. Second, EU KLEMS is a standardized database that allows comparability across a larger set of countries. The variables I use are: (i) aggregate real value added, (ii) an index of human capital, (iii) an index of physical capital, and (iv) labor share. Only 20 of the 30 countries have information on all the above variables. The database is an unbalanced panel with different time series length between countries. The shortest time series is 14 years, whereas the largest is 37.

6.2 Objects of interest

Elasticity of substitution. I start by commenting on the empirical estimates of the elasticity of substitution. Figure 3 shows the joint posterior distribution of the elasticity of substitution among countries. The posterior mean and median of the joint posterior distribution of the elasticity of substitution across countries is 0.92 and 0.87, respectively. The joint distribution has a standard deviation of 0.23.[30]

My framework also allows me to recover the posterior distribution of the elasticity of substitution for each country. I find considerable heterogeneity in the posterior distributions of the elasticity of substitution across countries. Table 1 shows moments of the posterior distribution for each of the 20 countries analyzed. Column 2 and 3 show the posterior mean and posterior median of all the countries in the database. Of the 20 countries used in the estimation, 14 have a posterior mean of the elasticity of substitution below one, whereas 6 have a posterior mean above one. The posterior median of the elasticity of substitution is below one for 15 countries while it is above one for the other 5 countries. The difference between the number of countries with a posterior mean below one and a posterior median below one is due to Denmark, which has a mean of 1.07 but a median of 0.99. Column 5 reports the probability that the elasticity of substitution is below one. There are 12 countries in which this probability is higher than 80%. These countries are Austria, Canada, Czech Republic, France, Germany, Hungry, Italy, Japan, Korea, Luxembourg, UK and U.S. There are 6 countries in which the probability is lower than 20%. These countries are Australia, Ireland, Netherlands, Spain, Portugal and Sweden. Finally, the probability is 71%

[30] I also estimate the model using the database of Karabarbounis and Neiman (2014), which is a very carefully constructed database on labor share for 82 countries between 1975 and 2010. However, using this database I can only estimate the prices equations, given that this database does not provide data on output, physical capital and human capital. Figure 8 shows more heterogeneity in the elasticity of substitution with a higher number of countries with an elasticity higher than one. Nevertheless, the posterior mean is still situated below one (0.95). This higher heterogeneity is due to the inclusion of developing countries which have an elasticity higher than one.

in Belgium and 50% in Denmark.

Figures 6 and 7 show the posterior distribution of the elasticity of substitution for each country. For many of the countries the posterior distribution is very informative about whether the elasticity of substitution is below or above one. For example, almost all the mass of the posterior distributions of the U.S. lies below one, while a considerable mass of the posterior distribution of Spain, lies above one. Nevertheless in some countries the posterior distribution is not so informative about whether the elasticity of substitution is below or above one, as in the case of Denmark.

Capital- and labor-augmenting technology. Columns 2 and 3 of table 2 show the average annual growth rate of the labor- and capital-augmenting technology, respectively. There is considerable heterogeneity in the growth rate of the two types of technologies. Of the 20 countries in the sample, 7 of them have experienced a positive annual growth rate of the capital-augmenting technology but a negative annual growth rate of the labor-augmenting technology. Conversely, 10 other countries have experienced a positive annual growth rate of the labor-augmenting technology but a negative annual growth rate of the capital-augmenting technology. Finally, there are 3 other countries where both the labor- and the capital-augmenting technology present a positive average annual growth rate. Column 4 shows a weighted average of the growth rate of both technologies as a measure of total efficiency. Korea and Germany are the countries with the highest growth in total efficiency, with an average annual growth rate of 2.5% and 1.3%, respectively. On the other hand, Spain and Portugal are the countries with the lowest growth in total efficiency, with an average annual growth rate of -0.2% and -0.7%, respectively. The weighted average of the factor-augmenting technologies growth rates produced by my model is very similar to the average annual growth rate of the Total Factor Productivity reported by EU KLEMS (column 5).

Table 3 shows the R-squared of a regression between the output growth rate over the common factor of the capital-augmenting technology growth rate and the common factor of the labor-augmenting technology growth rate for each country. Each regression was done inside the MCMC. Therefore, the results reported in table 3, are the posterior means of the R-squared of each country-regression produced in the chain. The result shows that the common factors of the capital-augmenting technology and the labor-augmenting technology explain a significant fraction of the fluctuations in the growth rate of output in the majority of countries in the sample. The common factors are consistent with the idea of Acemoglu and Zilbotti (2001) in which technologies are created in only a small number of countries and acquired by others.

Directed technical change. The results suggest different paths for the labor-augmenting technology and the capital-augmenting technology across countries. With these estimates of the technology process is possible to test the implications of the models of directed technical change in Acemoglu (2002) and Caselli and Coleman (2006). These models suggest that the direction of the technical change A_L/A_K depends on the relative endowment of human capital and physical capital L/K and the elasticity of substitution. In particular, these models imply a negative (positive) relation between A_L/A_K and L/K if the elasticity of substitution is lower (higher) than one. If the elasticity of substitution is lower than one, countries that experience a faster capital accumulation, are the ones that invest more in labor-augmenting technology relative to the capital-augmenting technology. In order to test this hypothesis, I carry out two different exercises:

Exercise 1: Cross-country regression. For each iteration j of the Markov Chain, I compute the OLS estimator of α and β of the following cross-country regression :

$$\bar{\gamma}^j_{A_{L,i}} - \bar{\gamma}^j_{A_{K,i}} = \alpha^j + \beta^j \rho^j_i \left(\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}\right) + \varepsilon^j_i$$

where $\bar{\gamma}^j_{A_{L,i}} - \bar{\gamma}^j_{A_{K,i}}$ is the difference between the average annual growth rate of labor augmenting technology and the average annual growth rate capital augmenting technology in country i, for the iteration j of the chain. $\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}$ is the difference between the human capital growth rate and the physical capital growth rate in country i. Table 5, shows the posterior mean of $\hat{\alpha}_{OLS}$ and $\hat{\beta}_{OLS}$. The coefficient $\hat{\beta}_{OLS}$ is positive and significant. This is in line with directed technical change models, where countries with an elasticity of substitution lower (higher) than one, i.e: $\rho < 0$ ($\rho > 0$) prefer to invest in technologies that increase the efficiency of their scarce (abundant) factor.

Exercise 2: Time-series regression. Second, for each iteration j of the Markov Chain, I compute the OLS estimator of α and β of the following time series regression of the aggregate economy:

$$f^j_{L,t} - f^j_{K,t} = \alpha^j + \beta^j \left[\frac{1}{N}\sum_{i=1}^{N} \rho^j_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right] + \varepsilon^j_t$$

where $f^j_{L,t} - f^j_{K,t}$ is the difference between the common factor of the labor augmenting technology and the common factor of the capital augmenting technology in period t, for the iteration j of the chain. $\left[\frac{1}{N}\sum_{i=1}^{N} \rho^j_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right]$ is the average across countries of the cross product of ρ_i and the difference between the observed growth rate of the human capital and the observed growth rate of the physical capital in country i for period t. Table 6 show the posterior mean of the OLS estimator of β. The coefficient is positive and significant.

Contribution in the labor share. The labor share of income could change in response to (i) channels that affect the relative factor prices, such as changes in the capital-labor ratio K/L or (ii) changes in the bias of technical change A_K/A_L. Using my model, I can assess the contribution of these two channels to the decline in the labor share for each country in the database. Column 2 of Table 4 shows the cumulative change in percentage points of the labor share. Column 3 shows the posterior mean of the cumulative change of the labor share predicted by my model (in percentage points). Column 4 shows the posterior mean of the cumulative change, in percentage points, of the labor share generated by the increase in the capital-labor ratio . Column 5 shows the posterior mean of the cumulative change in the labor share generated by a bias in the technical change. There are several findings. First, the bias in technical change has been the dominant mechanism in leading the decline in the labor share for most of the countries analyzed. In all the countries the change in the bias of the technical change has generated a decline in the labor share. For example, in the U.S. the observed labor share has fallen since 1970 from 0.670 to 0.606 in 2005, a cumulative decrease of 6.4 percentage points. My model predicts a cumulative decrease of 6.32

percentage points in which the bias in the technical change produces a decline of -8.71 whereas the increase in the capital-labor ratio produces an increase of 2.39.[31] My model predicts a similar decomposition of the decline in the labor share for Austria, Belgium, France, German, Italy and Luxemburg. On the other hand, the increase in the capital-labor ratio has generated a significant decline in the labor share in the 6 countries with elasticity of substitution higher than one (Australia, Spain, Ireland, Netherlands, Portugal and Sweden). In these countries the increase in the capital-labor ratio represents between 16% and 36% of the cumulative decline in the labor share. Moreover the increase in the capital-labor ratio has been the dominant effect in countries where the labor share has increased, or has not decreased too much, such as in Czech Republic, Japan, Korea or U.K.[32]

[31]The predictions of my model for the U.S. are in line with the results found by Oberfield and Raval (2014). They show a cumulative decline of 16.76 percentage points in the labor share for manufacturing. Of this decline, -19.96 was explained by the bias in the technical change whereas the factor prices contributes with an increase of 3.20.

[32]The evolution of the labor share explained in my model by changes in the capital-labor ratio can be also explained by a decrease in the price of investment goods as in Karabarbounis and Neiman (2014), who use a two-goods model in which there is a decline in the relative price of investment. As a result, firms shift away from labor toward capital, and with an elasticity of substitution larger than 1 the capital share increases.

7 Conclusion

In this paper I estimate in a flexible manner the parameters of the aggregate CES production function for many countries, using a panel model approach with unobserved heterogeneity. In contrast to previous studies, my framework considers country heterogeneity in the elasticity of substitution and in the growth rates of the labor- and capital- augmenting technologies. Additionally, the country specific growth rates of the labor- and capital-augmenting technologies are allowed to vary over time while retaining some commonalities across the panel via a dynamic factor model.

My framework contributes to the literature in three aspects. First, it is the first paper that estimates the elasticity of substitution for a wide range of countries, allowing for heterogeneity in the elasticity of substitution. Estimating a country specific elasticity of substitution is useful to understand which of the competing mechanisms in explaining the evolution of the labor share is most relevant in a particular country. Second, it is the first paper that estimates the supply system model using a factor model structure as a proxy for the underlying labor- and capital- augmenting technologies. The factor model captures, in a reduced-form fashion, some important features of structural models of aggregate technology that are helpful in understanding country heterogeneity and commonalities in productivities across countries. Finally, my framework provides a flexible manner of controlling for unobserved heterogeneity in the cross-section and the time series, which helps in identifying the elasticity of substitution.

The model I propose to estimate is a nonlinear panel system of equations with random coefficients (the country-specific elasticity of substitution) and unobserved factors (the growth rates of the labor- and capital-augmenting technologies). Estimation is challenging given the non-linearity and the amount of unobserved heterogeneity allowed both in the cross sectional and the time series dimensions. I complete my framework with distributional assumptions for the country- and time- specific latent variables and estimate the model using a Bayesian fixed effects approach (BFE). The BFE uses the parametric distributions of the latent variables as priors in a Bayesian fashion. These priors are combined with data to form posterior distributions which are then used in obtaining estimates of average effects in my model. This is in contrast to the random effects approach where priors are not updated since they are taken as the truth. The BFE can be expected to have similar asymptotic properties to the standard fixed effects when N and T go to infinity, but the use of priors may lead the BFE estimator to have better properties in small samples. In addition, the Bayesian fixed effects approach also provides a computationally tractable way of estimating the model by using Markov Chain Monte Carlo (MCMC).

Given the non-linearity of my model, I construct the Markov chain using the Metropolis-Hasting algorithm (M-H). The computational challenge in this algorithm is to set proposal distributions for each unobservable variable in the model (more than 200 in my model). To that end, I propose a feasible and computationally efficient procedure for obtaining posterior distributions of the parameters of the supply system model. I do this by dividing the estimation in two steps, using the price equations to define suitable proposal distributions and initial conditions for the M-H algorithm.

The results show heterogeneity in the elasticity of substitution with a mean of the cross-country distribution equal to 0.92, a median of 0.87 and a standard deviation of 0.23. My framework also allows me to recover the posterior distribution of the elasticity of substitution for each country. Of the 20 countries used in the estimation, 14 have a posterior mean of the elasticity of substitution below one, whereas 6 have a posterior mean above one. This implies that the increase in technical change has been the dominant mechanism in leading the decline in the labor share for most of the countries analyzed. However, the the capital accumulation (or the decline in the price of investment goods) has played a significant role in the labor share decline in the 6 countries with elasticity of substitution higher than one. Moreover, the capital accumulation has been the dominant effect in countries where the labor share has increased, or has not decreased too much. Finally, I find that the growth rate of the labor-augmenting technology relative to the growth rate of the capital- augmenting technology correlates negatively with the growth rate of the capital-labor ratio. These result is in line with the directed technical change models of Acemoglu and Zilbotti (2001) , Acemoglu (2002) and Caselli and Coleman (2006) in which countries invest in technologies that increase the efficiency of the scarcity factor.

References

Acemoglu, Daron. 2002. "Directed technical change." *The Review of Economic Studies* 69(4):781–809.

Acemoglu, Daron and Fabrizio Zilbotti. 2001. Productivity differences. Technical report Quarterly Journal of Economics.

Ackerberg, Daniel, Kevin Caves and Garth Frazer. 2006. "Structural identification of production functions.".

Antras, Pol. 2004. "Is the US aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution." *Contributions in Macroeconomics* 4(1).

Arellano, Manuel and Stéphane Bonhomme. 2009. "Robust priors in nonlinear panel data models." *Econometrica* 77(2):489–536.

Arellano, Manuel and Stéphane Bonhomme. 2011. "Nonlinear panel data analysis." *Annual Review of Economics* 3:395–424.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica* 77(4):1229–1279.

Caselli, Francesco. 2005. "Accounting for cross-country income differences." *Handbook of economic growth* 1:679–741.

Caselli, Francesco and Wilbur John Coleman. 2006. "The world technology frontier." *The American Economic Review* pp. 499–522.

Chamberlain, Gary. 1982. "Multivariate regression models for panel data." *Journal of Econometrics* 18(1):5–46.

Chamberlain, Gary. 1984. "Panel data." *Handbook of econometrics* 2:1247–1318.

Chib, Siddhartha. 2001. "Markov chain Monte Carlo methods: computation and inference." *Handbook of econometrics* 5:3569–3649.

de La Grandville, O and RM Solow. 2005. "Capital-labor substitution and economic growth.".

Diamond, Peter A and Daniel McFadden. 1965. "Identification of the elasticity of substitution and the bias of technical change: An impossibility theorem." *University of California, Berkeley, CA* .

Doraszelski, Ulrich and Jordi Jaumandreu. 2013. "R&D and productivity: Estimating endogenous productivity." *The Review of Economic Studies* 80(4):1338–1383.

Duffy, John and Chris Papageorgiou. 2000. "A cross-country empirical investigation of the aggregate production function specification." *Journal of Economic Growth* 5(1):87–120.

Fernández-Val, Iván and Martin Weidner. 2013. "Individual and time effects in nonlinear panel models with large N, T." *arXiv preprint arXiv:1311.7065* .

Geweke, John. 2004. "Getting it right: Joint distribution tests of posterior simulators." *Journal of the American Statistical Association* 99(467):799–804.

Gollin, Douglas. 2002. "Getting income shares right." *Journal of political Economy* 110(2):458–474.

Hall, Robert E and Charles I Jones. 1999. Why do some countries produce so much more output per worker than others? Technical report National bureau of economic research.

Herrendorf, Berthold, Christopher Herrington and Akos Valentinyi. 2013. *Sectoral technology and structural transformation.* Centre for Economic Policy Research.

Hirano, Keisuke. 2002. "Semiparametric Bayesian inference in autoregressive panel data models." *Econometrica* 70(2):781–799.

Jones, Charles I. 2005. The shape of production function and the direction of technical change. Technical report Quarterly Journal of Economics.

Karabarbounis, Loukas and Brent Neiman. 2014. "The Global Decline of the Labor Share." *The Quarterly Journal of Economics* 129(1):61–103.

Klump, Rainer and Olivier de La Grandville. 2000. "Economic growth and the elasticity of substitution: Two theorems and some suggestions." *American Economic Review* pp. 282–291.

Kose, M Ayhan, Christopher Otrok and Charles H Whiteman. 2003. "International business cycles: World, region, and country-specific factors." *American Economic Review* pp. 1216–1239.

Lagos, Ricardo. 2006. "A model of TFP." *The Review of Economic Studies* 73(4):983–1007.

León-Ledesma, Miguel A, Peter McAdam and Alpo Willman. 2010. "Identifying the elasticity of substitution with biased technical change." *The American Economic Review* 100(4):1330–1357.

Levinsohn, James and Amil Petrin. 2003. "Estimating production functions using inputs to control for unobservables." *The Review of Economic Studies* 70(2):317–341.

Moon, H and Martin Weidner. 2010*a*. "Dynamic linear panel regression models with interactive fixed effects." *manuscript, UCL* .

Moon, Hyungsik and Martin Weidner. 2010*b*. "Linear regression for panel with unknown number of factors as interactive fixed effects." *Unpublished manuscript, University of Southern California* .

Moon, Hyungsik Roger and Martin Weidner. 2014. "Linear regression for panel with unknown number of factors as interactive fixed effects." *USC-INET Research Paper* (15-03).

Mundlak, Yair. 1978. "On the pooling of time series and cross section data." *Econometrica: journal of the Econometric Society* pp. 69–85.

Oberfield, Ezra and Devesh Raval. 2014. Micro data and macro technology. Technical report Working Paper, Federal Reserve Bank of Chicago.

Piketty, Thomas and Gabriel Zucman. 2014. "Capital is back: Wealth-income ratios in rich countries, 1700-2010." *The Quarterly Journal of Economics* p. qju018.

Appendix 1: Implementation of the algorithm: Gibbs sampling from the price equations

Under the perfect competition assumption I can recover the posterior distribution of all the latent variables just working with the prices equations.

$$p\left(\mathbb{Z},\,\Theta \mid Y,\,sh^{K},\,sh^{L},\,K,\,L\right) \propto \mathcal{L}\left(sh^{K},\,sh^{L} \mid Y,\,K,\,L,\mathbb{Z}\right) p\left(\mathbb{Z}\,|K,\,L,\,\Theta\right) p\left(\Theta\right)$$

To construct the conditional posterior distribution in each block, I extend the methodology in (Hirano, 2002) to a framework with a factor model and random coefficients. With the intention of simplifying the model, by working with stationary variables and reducing the dimensionality of the problem, I work with the model in first differences[33]:

$$\triangle log\left(sh^{L}_{it}\right) = \rho_i \lambda_{Li} F_{Li} + \rho_i \triangle log\left(\frac{L_{it}}{Y_{it}}\right) + \triangle\varepsilon^{w}_{it}$$

$$\triangle log(sh^{K}_{it}) = \rho_i \lambda_{Ki} F_{Ki} + \rho_i \triangle log\left(\frac{K_{it}}{Y_{it}}\right) + \triangle\varepsilon^{r}_{it}$$

In each block, the only "unknown" is the unobservable corresponding to this block, the other parameters and latent variables will play the role of observed variables, i.e. data. Before starting with the derivation of the posterior distribution for each block, let me define, for the sake of simplicity, the vector $\mathbb{Z}_{-\alpha}$ as a vector that contains all the latent variables in $\mathbb{Z}$ but not the latent variable α and let me rename the variables of the model: $y_{L,it} \equiv \triangle log\left(sh^{L}_{it}\right)$, $y_{K,it} \equiv \triangle log(sh^{K}_{it})$, $x_{L,i1} \equiv \triangle log\left(\frac{L_{it}}{Y_{it}}\right)$, $x_{K,i1} \equiv \triangle log\left(\frac{K_{it}}{Y_{it}}\right)$, $\triangle\varepsilon_{W,it} \equiv e_{W,it}$ and $\triangle\varepsilon_{R,it} \equiv e_{R,it}$.

I use normal priors for the latent variables to obtained closed form expressions for the conditional distribution in each block and be able to draw from the model using a Gibbs sampling algorithm as in Hirano (2002); Kose, Otrok and Whiteman (2003). Given the concern of the endogeneity of the explanatory variables I will allow for correlation between the latent variables and the explanatory variables as in the correlated random effect of Mundlak (1978); Chamberlain (1984)

Note that the log of the prices equations in first difference fit perfectly in the Bai (2009) setup with the only difference of a random coefficient rather than a common one. In the next section, I will estimate the prices

[33] As a robustness check, I have also solved the model in levels.

equations, using the estimator proposed by Bai (2009) but with random coefficients in order to have a comparison between a pure fixed effect estimator and the Bayesian fixed effect.[34]

7.1 *Elasticity of Substitution:*

First I estimate the conditional distribution for each ρ_i assuming that all the other parameters are known. Each ρ_i comes from separate blocks, because given the other parameters, latent variables and observable variables they are independent across countries.

Using Bayes theorem and recalling that all the unknowns variables in this block as treated as data, the conditional distribution of ρ_i:

$$p(\rho_i \mid \mathbb{X}, \mathbb{Z}_{-\rho_i}, \Theta) \propto \mathcal{L}\left(\mathbb{X}, \mathbb{Z}_{-\rho_i}, \Theta \mid \rho_i\right) p\left(\rho_i\right)$$

$$p(\rho_i \mid \mathbb{X}, \mathbb{Z}_{-\rho_i}, \Theta) = \mathcal{L}\left(y_{L,i1} \cdots y_{L,iT} \mid \rho_i, \mathbb{Z}_{-\rho_i}, x_{L,i1} \cdots x_{L,iT}\right) \times$$

$$\mathcal{L}\left(y_{K,i1} \cdots y_{K,iT} \mid \rho_i, \mathbb{Z}_{-\rho_i}, x_{K,i1} \cdots x_{K,iT}\right) \times$$

$$p\left(\rho_i \mid \Theta, \overline{x_{Ki}}, \bar{x}_{Li}\right)$$

There are three models that give me information about ρ_i: (1) the labor share equation , (2) the capital share equation and (3) the prior. As in the optimal GMM estimation, the conditional distribution for $\rho_i$is just a Normal distribution which combines ,in an optimal way, the information from each of the three equations.

1. The labor share equation:

$$y_{L,it} = \rho_i \left(\lambda_{L,i} F_{L,t} + x_{L,it}\right) + e_{W,it}$$

In this block the only unknown is ρ_i, so we can estimate it using the time series information of country *i*. I will use the GLS estimator, because working with first difference generate a known serial correlation in the error terms.

[34] It is difficult to do the same exercise with the supply system model, because of the computational challenge involved in the estimation by "fixed effect" of the nonlinear CES with a high degree of heterogeneity. The extremum estimators are known to be difficult to compute due to highly non-convex criterion functions with many local optima

Let define the following $(T_i - 1)x1$ vectors: $W_{L,i} = \lambda_{L,i}F_L + X_{L,i}$ *and* $Y_{L,i}$, where T_i is the sample size in country i, F_L is a vector that contains the $(T_i - 1)$ growth rates of labor augmenting technology (from 2 to T), $X_{L,i}$ is a vector that contains all the time series of the variable $x_{L,it}$, the same for $Y_{L,i}$.

Following Lancaster(1997) the likelihood of the data could be written as that of a Normal distribution

$$\rho_i \sim N\left(\hat{\rho_{L,i}},\ Var(\hat{\rho_{L,i}})\right)$$

where $\hat{\rho_{L,i}} = \left(W'_{L,i}\Gamma\left(\sigma_W^2\right)^{-1} W_{L,i}\right)^{.1} W'_{L,i}\Gamma\left(\sigma_W^2\right)^{-1} Y_{L,i}$ is the GLS estimator of ρ_i and

$Var(\hat{\rho_{L,i}}) = \left(W'_{L,i}\Gamma\left(\sigma_W^2\right)^{-1} W_{L,i}\right)^{-1}$ is the variance of the GLS estimator.

2. The capital share equation:

$$y_{K,it} = \rho_i\left(\lambda_{K,i}F_{K,t} + x_{K,it}\right) + e_{R,it}$$

Following the same procedure as in the labor share equation: $\rho_i \sim N\left(\hat{\rho_{K,i}},\ Var(\hat{\rho_{K,i}})\right)$,

where $\hat{\rho_{K,i}} = \left(W'_{K,i}\Gamma\left(\sigma_K^2\right)^{-1} W_{K,i}\right)^{.1} W'_{K,i}\Gamma\left(\sigma_K^2\right)^{-1} Y_{K,i}$and

$Var(\hat{\rho_{K,i}}) = \left(W'_{K,i}\Gamma\left(\sigma_K^2\right)^{-1} W_{K,i}\right)^{-1}$

3. Prior of ρ_i

$$\rho_i \sim N\ \left(\mu_{\rho,i},\ \sigma_\rho^2\right)$$

where $\mu_{\rho,i} = \phi_\rho + \phi_{\rho,K}\overline{x_{K,i}} + \phi_{\rho,L}\overline{x_{L,i}}$

The final conditional posterior distribution of the block of ρ_i will be a weighted average of the three models:

$$\rho_i{\sim}N\left(H_{\rho_i}^{-1}\left(\frac{1}{Var(\hat{\rho_{L,i}})}\hat{\rho_{L,i}} + \frac{1}{Var(\hat{\rho_{K,i}})}\hat{\rho_{K,i}} + \frac{1}{\sigma_\rho^2}\mu_\rho\right),\ H_{\rho_i}^{-1}\right)$$

where $H_{\rho_i} = \left(\frac{1}{Var(\hat{\rho_{L,i}})} + \frac{1}{Var(\hat{\rho_{K,i}})} + \frac{1}{\sigma_\rho^2}\right) = \left(W'_{L,i}\Gamma\left(\sigma_L^2\right)^{-1} W_{L,i} + W'_{K,i}\Gamma\left(\sigma_K^2\right)^{-1} W_{K,i} + \frac{1}{\sigma_\rho^2}\right)$.

$$\rho_i \sim N\left(H_{\rho_i}^{-1}\left[\left(W'_{L,i}\Gamma\left(\sigma_L^2\right)^{-1} Y_{L,i}\right) + \left(W'_{K,i}\Gamma\left(\sigma_K^2\right)^{-1} Y_{K,i}\right) + \left(\frac{1}{\sigma_\rho^2}\right)\left(\phi_\rho + \phi_{\rho,K}\overline{x_{K,i}} + \phi_{\rho,L}\overline{x_{L,i}}\right)\right],\ H_{\rho_i}^{-1}\right)$$

7.2 *Loading of the labor augmenting parameter:*

$$p(\lambda_{L,i} \mid \mathbb{X}, \mathbb{Z}_{-\lambda_{L,i}}, \Theta) \propto \mathcal{L}\left(\mathbb{X}, \mathbb{Z}_{-\lambda_{L,i}}, \Theta \mid \lambda_{L,i}\right) p\left(\lambda_{L,i}\right)$$

$$p(\lambda_{L,i} \mid \mathbb{X}, \mathbb{Z}_{-\lambda_{L,i}}, \Theta) = \mathcal{L}\left(y_{L,i1} \cdots y_{L,iT} \mid \lambda_{L,i}, \mathbb{Z}_{-\lambda_{L,i}}, x_{L,i1} \cdots x_{L,iT}\right) \times$$

$$p\left(\lambda_{L,i} \mid \Theta, \bar{x}_{Li}\right)$$

There are two models that give me information about $\lambda_{L,i}$: (1) the labor share equation ,and (2) the prior.

1. **The labor share**

$$y_{L,it} - \rho_i x_{L,it} = \rho_i \lambda_{L,i} F_{L,t} + e_{W,it}$$

$$\lambda_{L,i} \sim N\left(\hat{\lambda_{L,i}}, Var(\hat{\lambda_{L,i}})\right)$$

where $\hat{\lambda_{L,i}} = \frac{1}{\rho_i}\left(F_L'\Gamma\left(\sigma_W^2\right)^{-1} F_L\right)^{.1} F_L'\Gamma\left(\sigma_W^2\right)^{-1} Q_{L,i}$ is the GLS estimator of $\lambda_{L,i}$ and $Q_{L,i}$ is a $(T_i - 1)\times 1$ vector that contains the all time series of the variable $y_{L,it} - \rho_i x_{L,it}$

$Var(\hat{\lambda_{L,i}}) = \frac{1}{\rho_i^2}\left(F_L'\Gamma\left(\sigma_W^2\right)^{-1} F_L\right)^{-1}$ is the variance of the GLS estimator.

2. The prior:

$$\lambda_{L,i} \sim N\left(\mu_{\lambda_{L,i}}, \sigma^2_{\lambda_L}\right)$$

where $\mu_{\lambda_{L,i}} = \phi_{0,\lambda_L} + \phi_{1,\lambda_L}\overline{x_{L,i}}$

The final conditional posterior distribution of the block of $\lambda_{L,i}$ will be a weighted average of the two models:

$$\lambda_{L,i} \sim N\left(H^{-1}_{\lambda_{L,i}}\left(\frac{1}{Var(\hat{\lambda_{L,i}})}\hat{\lambda_{L,i}} + \frac{1}{\sigma^2_{\lambda_L}}\mu_{\lambda_{L,i}}\right), H^{-1}_{\lambda_{L,i}}\right)$$

where $H_{\lambda_{L,i}} = \left(\frac{1}{Var(\hat{\lambda_{L,i}})} + \frac{1}{\sigma^2_{\lambda_L}}\right) = \left(\rho_i^2 F_L'\Gamma\left(\sigma_W^2\right)^{-1} F_L + \frac{1}{\sigma^2_{\lambda_L}}\right)$.

$$\lambda_{L,i} \sim N\left(H^{-1}_{\lambda_{L,i}}\left[\left(\rho_i F_L'\Gamma\left(\sigma_W^2\right)^{-1} Q_{L,i}\right) + \left(\frac{1}{\sigma^2_{\lambda_L}}\right)\left(\phi_{0,\lambda_L} + \phi_{1,\lambda_L}\overline{x_{L,i}}\right)\right], H^{-1}_{\lambda_{L,i}}\right)$$

7.3 *Loading of the capital augmenting parameter:*

$$p(\lambda_{L,i} \mid \mathbb{X},\, \mathbb{Z}_{-\lambda_{L,i}},\, \Theta) \propto \mathcal{L}\left(\mathbb{X},\, \mathbb{Z}_{-\lambda_{L,i}},\, \Theta \mid \lambda_{L,i}\right) p\left(\lambda_{L,i}\right)$$

$$p(\lambda_{L,i} \mid \mathbb{X},\, \mathbb{Z}_{-\lambda_{L,i}},\, \Theta) = \mathcal{L}\left(y_{L,i1} \cdots y_{L,iT} \mid \lambda_{L,i},\, \mathbb{Z}_{-\lambda_{L,i}},\, x_{L,i1} \cdots x_{L,iT}\right) \times$$

$$p\left(\lambda_{L,i} \mid \Theta,\, \bar{x}_{Li}\right)$$

There are two models that give me information about $\lambda_{K,i}$: (1) the labor share equation ,and (2) the prior.

1. **The labor share**

$$y_{K,it} - \rho_i x_{K,it} = \rho_i \lambda_{K,i} F_{K,t} + e_{R,it}$$

$$\lambda_{K,i} \sim N\left(\hat{\lambda_{K,i}},\ Var(\hat{\lambda_{K,i}})\right)$$

where $\hat{\lambda_{K,i}} = \frac{1}{\rho_i}\left(F_K'\Gamma\left(\sigma_R^2\right)^{-1}F_K\right)^{.1} F_K'\Gamma\left(\sigma_W^2\right)^{-1} Q_{K,i}$ is the GLS estimator of $\lambda_{K,i}$ and $Q_{K,i}$ is a $(T_i - 1) \times 1$ vector that contains the all time series of the variable $y_{K,it} - \rho_i x_{K,it}$

$Var(\hat{\lambda_{K,i}}) = \frac{1}{\rho_i^2}\left(F_K'\Gamma\left(\sigma_R^2\right)^{-1}F_K\right)^{-1}$ is the variance of the GLS estimator.

2. The prior:

$$\lambda_{K,i} \sim N\left(\mu_{\lambda_{K,i}},\ \sigma^2_{\lambda_K}\right)$$

where $\mu_{\lambda_{K,i}} = \phi_{0,\lambda_K} + \phi_{1,\lambda_K}\overline{x_{K,i}}$

The final conditional posterior distribution of the block of $\lambda_{K,i}$ will be a weighted average of the two models:

$$\lambda_{K,i} \sim N\left(H^{-1}_{\lambda_{K,i}}\left(\frac{1}{Var(\hat{\lambda_{K,i}})}\hat{\lambda_{K,i}} + \frac{1}{\sigma^2_{\lambda_K}}\mu_{\lambda_{K,i}}\right),\ H^{-1}_{\lambda_{K,i}}\right)$$

where $H_{\lambda_{K,i}} = \left(\frac{1}{Var(\hat{\lambda_{K,i}})} + \frac{1}{\sigma^2_{\lambda_K}}\right) = \left(\rho_i^2 F_K'\Gamma\left(\sigma_R^2\right)^{-1}F_K + \frac{1}{\sigma^2_{\lambda_K}}\right).$

$$\lambda_{K,i} \sim N\left(H^{-1}_{\lambda_{K,i}}\left[\left(\rho_i F_K'\Gamma\left(\sigma_R^2\right)^{-1}Q_{K,i}\right) + \left(\frac{1}{\sigma^2_{\lambda_K}}\right)\left(\phi_{0,\lambda_K} + \phi_{1,\lambda_K}\overline{x_{K,i}}\right)\right],\ H^{-1}_{\lambda_{K,i}}\right)$$

7.4 *Growth rate of the labor augmenting technical factor:*

When estimating the posterior distribution of F_{Kt} and F_{Lt} we have to take into account that they follow a Markov process.

$$p(F_{L,t} \mid \mathbb{X}, \mathbb{Z}_{-F_{L,t}}, \Theta) \propto \mathcal{L}\left(\mathbb{X}, \mathbb{Z}_{-F_{L,t}}, \Theta \mid \lambda_{L,i}\right) p\left(F_{L,t}\right)$$

$$p(F_{L,t} \mid \mathbb{X}, \mathbb{Z}_{-F_{L,t}}, \Theta) = \mathcal{L}\left(y_{L,1t} \cdots y_{L,Nt} \mid F_{L,t}, \mathbb{Z}_{-F_{L,t}}, x_{L,1t} \cdots x_{L,Nt}\right) \times$$

$$p\left(F_{L,t} \mid \Theta, \bar{x}_{Lt}, F_{L,t-1}\right) \times p\left(F_{L,t+1} \mid \Theta, \bar{x}_{Lt+1}, F_{L,t}\right)$$

Given that the prior is an AR(1) process, the conditional posterior of the capital augmenting technology in period t will depend on the likelihood of the capital share model, the prior of $F_{Kt} : p\left(F_{L,t} \mid \Theta, \bar{x}_{Lt}, F_{L,t-1}\right)$ and the prior of $F_{Kt+1} : p\left(F_{L,t+1} \mid \Theta, \bar{x}_{Lt+1}, F_{L,t}\right)$ which also contains information of F_{Kt}.

1. Labor Share Equation

$$y_{L,it} - \rho_i x_{L,it} = \rho_i \lambda_{L,i} F_{L,t} + e_{W,it}$$

$$F_{L,t} \sim N\left(\left(\sum_{i=1}^{N} \rho_i^2 \lambda_{L,i}^2\right)^{-1} \sum_{i=1}^{N} \left(\rho_i \lambda_{L,i}\left(y_{L,it} - \rho_i x_{L,it}\right)\right),\ 2\sigma_W^2 \left(\sum_{i=1}^{N} \rho_i^2 \lambda_{L,i}^2\right)^{-1}\right)$$

2. The prior distribution of F_{Lt}

$F_{L,t} \sim N\left(\phi_{0,F_L} + \phi_{1,F_L}\bar{x}_{Lt} + \phi_{2,F_L}F_{L,t-1},\ \sigma_{F_L}^2\right)$

3. The prior distribution of F_{Lt+1}

$p\left(F_{L,t+1} \mid \Theta, \bar{x}_{Lt+1}, F_{L,t}\right) = N\left(\phi_{0,F_L} + \phi_{1,F_L}\bar{x}_{Lt+1} + \phi_{2,F_L}F_{L,t}^{j},\ \sigma_{F_L}^2\right) \to p\left(F_{L,t} \mid \Theta, \bar{x}_{Lt+1}, F_{L,t+1}\right)$

$F_{L,t} \sim N\left(.\frac{F_{L,t+1} - \phi_{0,F_L} + \phi_{1,F_L}\bar{x}_{Lt+1}}{\phi_{2,F_L}},\ \frac{\sigma_{F_L}^2}{\phi_{2,F_L}^2}\right)$

Again the conditional posterior is a combination of three normal distributions

$$F_{L,t} \sim N\left(\mu_{F_L}, H_{F_L}^{-1}\right)$$

where

$$\mu_{F_L} = H_{F_L}^{-1}\Big[\frac{1}{2\sigma_W^2}\sum_{i=1}^{N}\left(\rho_i\lambda_{L,i}\left(y_{L,it}-\rho_i x_{L,it}\right)\right)+\frac{1}{\sigma_{F_L}^2}\cdot\left(\phi_{0,F_L}+\phi_{1,F_L}\bar{x}_{Lt}+\phi_{2,F_L}F_{L,t-1}\right)$$
$$+\frac{\phi_{2,F_L}^2}{\sigma_{F_L}^2}\frac{F_{L,t+1}-\phi_{0,F_L}+\phi_{1,F_L}\bar{x}_{Lt+1}}{\phi_{2,F_L}}\Big]$$

and

$$H_{F_L} = \left(\frac{1}{2\sigma_L^2}\sum_{i=1}^{N}\rho_i^2\lambda_{L_i}^2+\frac{1}{\sigma_{F_L}^2}+\frac{\phi_{2,F_L}^2}{\sigma_{F_L}^2}\right)$$

7.5 *Growth rate of the capital augmenting technical factor:*

Follow the same procedure as the growth rate of the labor augmenting technical factor.

$$F_{K,t}\sim N\left(\mu_{F_K},H_{F_K}^{-1}\right)$$

where

$$\mu_{F_K} = H_{F_K}^{-1}\Big[\frac{1}{2\sigma_W^2}\sum_{i=1}^{N}\left(\rho_i\lambda_{K,i}\left(y_{K,it}-\rho_i x_{K,it}\right)\right)+\frac{1}{\sigma_{F_K}^2}\cdot\left(\phi_{0,F_K}+\phi_{1,F_K}\bar{x}_{Kt}+\phi_{2,F_K}F_{K,t-1}\right)$$
$$+\frac{\phi_{2,F_K}^2}{\sigma_{F_K}^2}\frac{F_{K,t+1}-\phi_{0,F_K}+\phi_{1,F_K}\bar{x}_{Kt+1}}{\phi_{2,F_K}}\Big]$$

and

$$H_{F_K} = \left(\frac{1}{2\sigma_K^2}\sum_{i=1}^{N}\rho_i^2\lambda_{K_i}^2+\frac{1}{\sigma_{F_K}^2}+\frac{\phi_{2,F_K}^2}{\sigma_{F_K}^2}\right)$$

7.6 *Hyper-parameters.*

In order to estimate the posterior distribution for the remaining parameters, I will assume a flat prior distribution, thus, the conditional posterior of these parameters will be a Normal distribution, just as in the linear Bayesian normal model.

Appendix 2: Tables and Figures

Figure 1: **Convergence to the stationary distribution of a particular ρ_i**

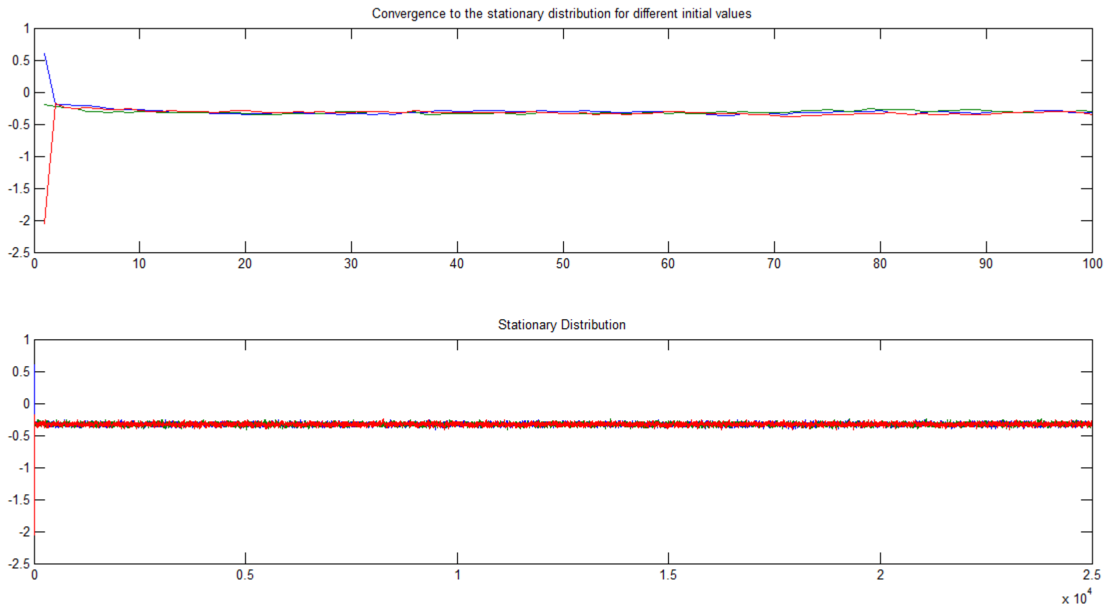


Figure 2: **Stationary Distribution of a particular ρ_i**

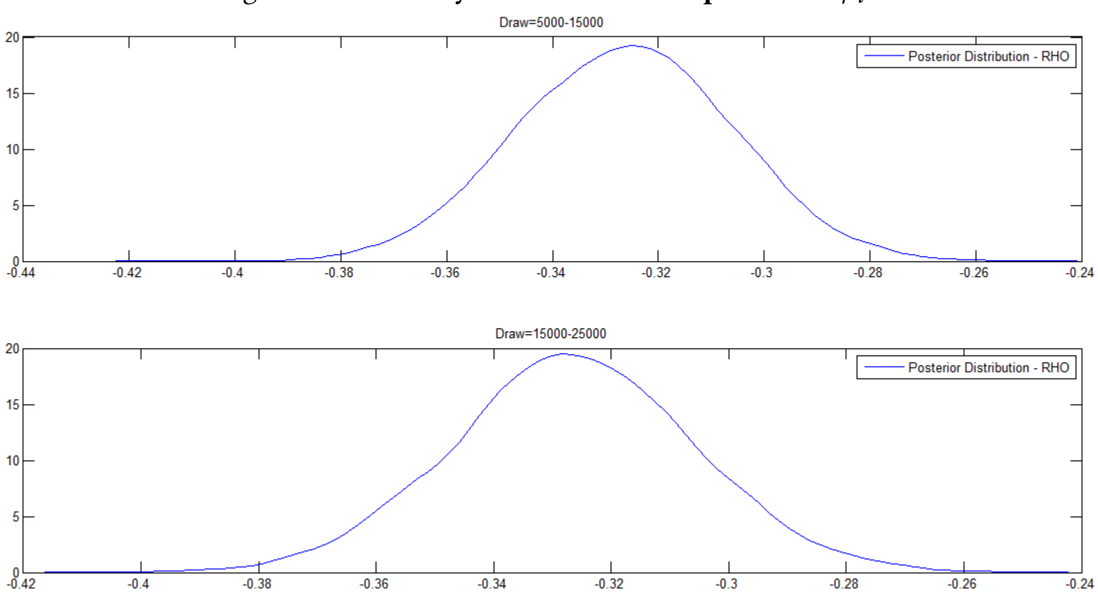
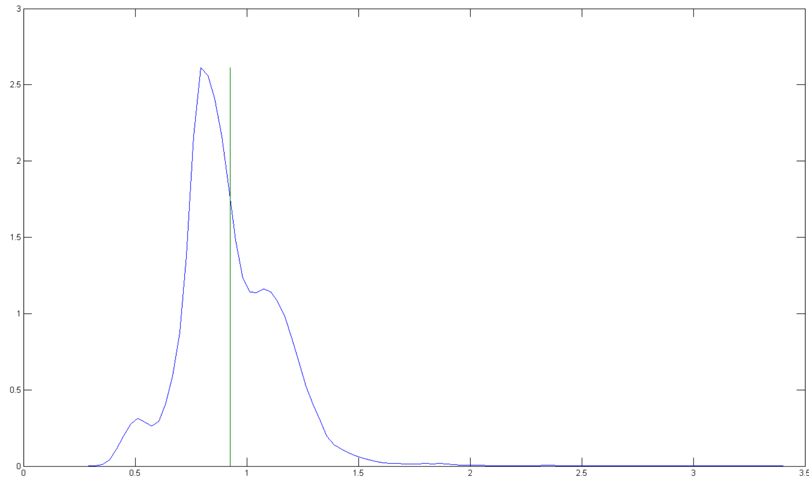
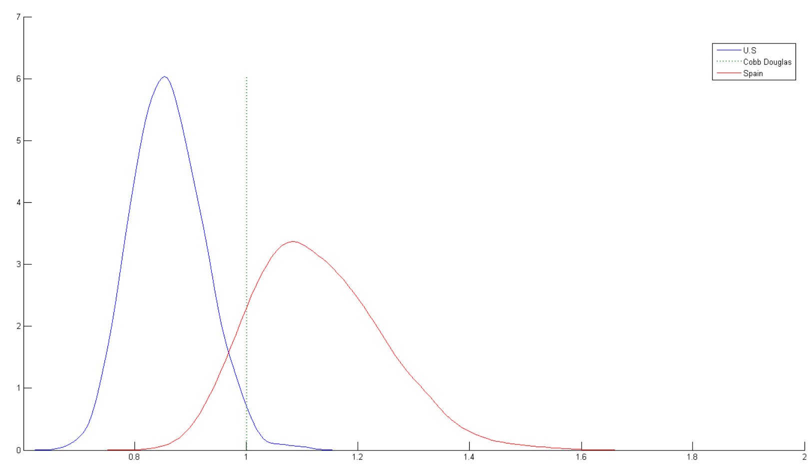


Figure 3: Distribution of the elasticity of substitution across countries



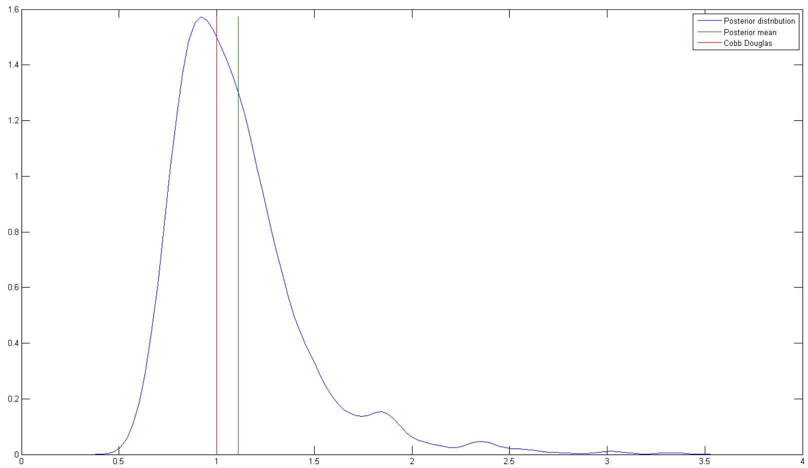
Note: The graph is the joint posterior distribution of all the countries in the sample. The straight line indicates the mean of the distribution, which is 0.92. The distribution of the elasticity lies both below and above one.

Figure 4: Posterior distribution of the elasticity of substitution



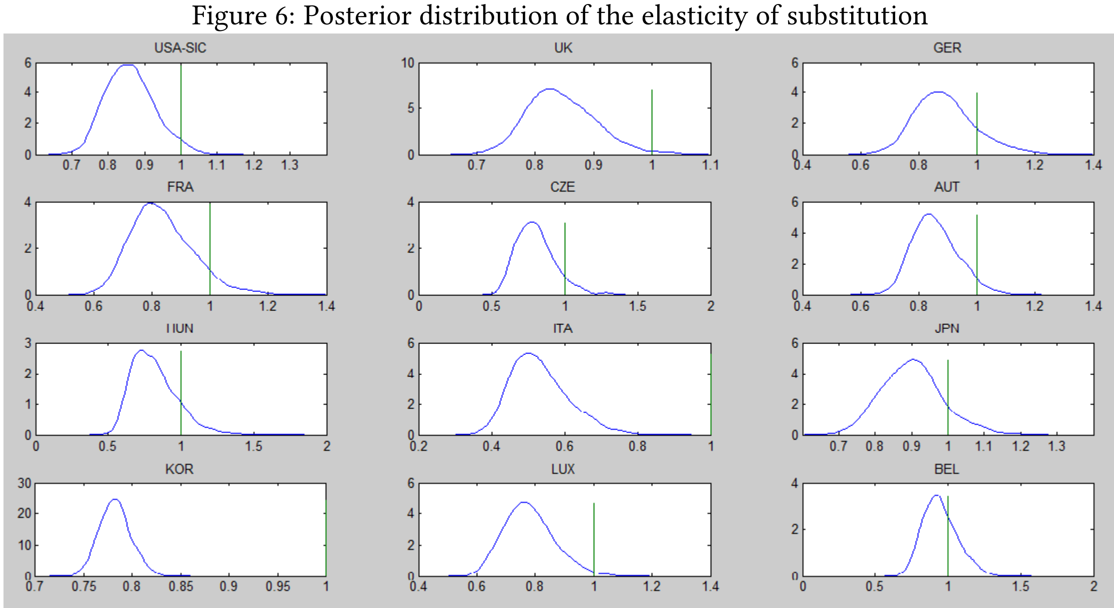
Note: The graph shows the posterior distribution of the elasticity of substitution for the U.S. (the blue line) and for Spain (the red line). The dotted green line indicates the Cobb-Douglas case when the elasticity of substitution is one. Almost all the mass of the posterior distribution in the U.S. lies below one. Conversely, most of the mass of the posterior distribution for Spain is situated above one.

Figure 5: Posterior distribution of Denmark



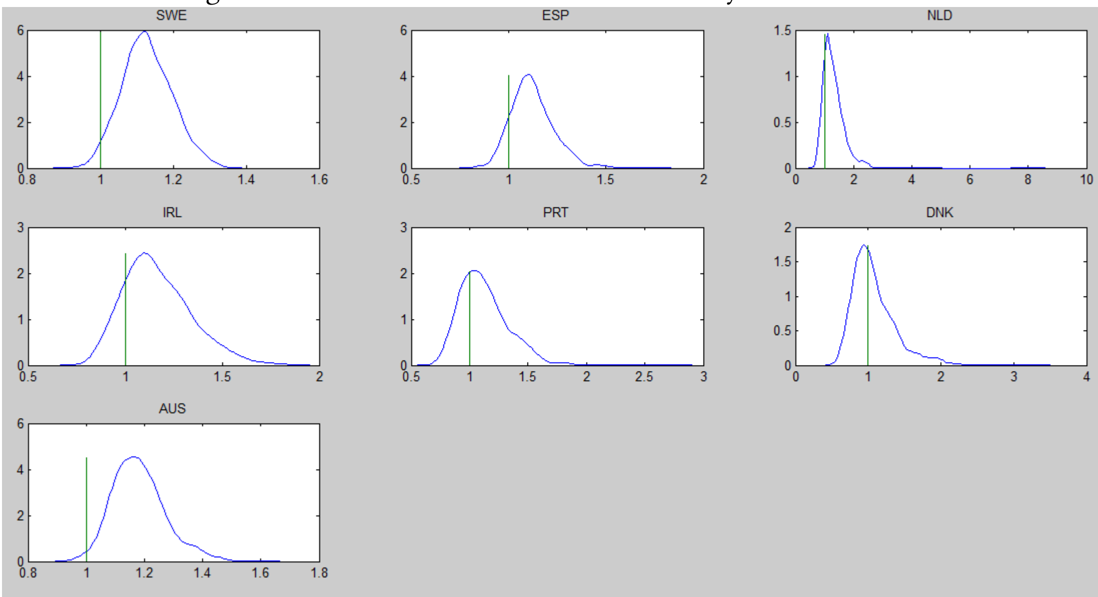
Note: The graph shows the posterior distribution of the elasticity of substitution in Denmark. The green line indicates the posterior mean of the distribution and the red line is the Cobb-Douglas case, when the elasticity of substitution is equal to one. The posterior distribution is not so informative about whether the elasticity of substitution is lower or higher than one since there is a similar mass of the distribution located below and above one. The probability of the elasticity of substitution of being less than one is 51%.

Figure 6: Posterior distribution of the elasticity of substitution



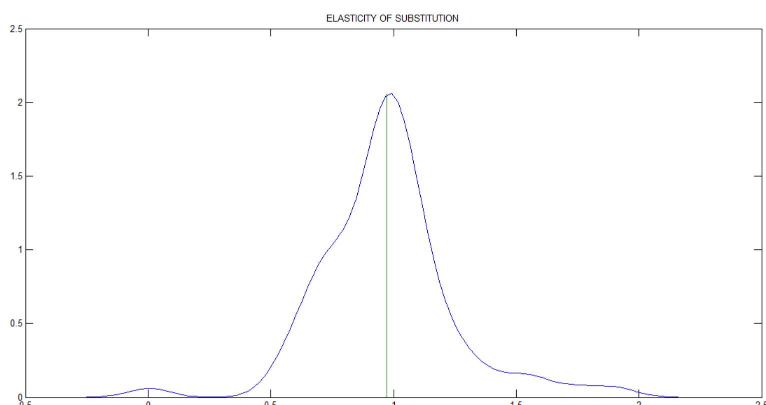
Note: The graph shows the posterior distribution of each country in the sample. The straight line indicates the Cobb-Douglas case, when the elasticity of substitution is one.

Figure 7: Posterior distribution of the elasticity of substitution



Note: The graph shows the posterior distribution of each country in the sample. The straight line indicates the Cobb-Douglas case, when the elasticity of substitution is one.

Figure 8: Distribution of the elasticity of substitution across countries. Karabarbounis and Neiman database



Note: I perform a similar exercise with the database of Karabarbounis and Neiman (2014). This is a very carefully constructed database on labor share for 82 countries between 1975 and 2010. However, using this database I can only estimate the prices equations, given that this database does not provide data on output, physical capital and human capital. Figure 8 shows more heterogeneity in the elasticity of substitution with a more number of countries with elasticities higher than one. Nevertheless, the posterior mean is still situated below one (0.95). This higher heterogeneity is due to the inclusion of developing countries that tends to have an elasticity higher than one.

Table 1: Elasticity of Substitution

Country	Mean	Median	Std	Prob(ES<1)
AUS	1.18	1.17	0.09	0.8%
AUT	0.84	0.84	0.07	96%
BEL	0.94	0.93	0.12	71%
CZE	0.80	0.78	0.14	92%
DNK	1.07	0.99	0.35	51%
ESP	1.13	1.12	0.11	10%
FRA	0.83	0.82	0.10	93%
GER	0.88	0.87	0.10	88%
HUN	0.80	0.77	0.15	91%
IRL	1.16	1.14	0.17	19%
ITA	0.63	0.61	0.08	100%
JPN	0.90	0.89	0.09	88%
KOR	0.78	0.78	0.02	100%
LUX	0.78	0.77	0.09	98%
NLD	1.37	1.22	2.18	18%
PRT	1.12	1.10	0.22	30%
SWE	1.16	1.13	0.07	0.5%
UK	0.84	0.84	0.06	99%
U.S	0.86	0.85	0.07	97%
Joint distribution	**0.92**	**0.87**	**0.23**	**40%**

Note: The second column shows the posterior mean of the elasticity of substitution for each country. The third column shows the posterior median of the elasticity of substitution for each country. The fourth column shows the standard deviation of the posterior distribution for each country. The last column is the probability of the elasticity of substitution being lower than one. This probability is the empirical probability calculated from the posterior distribution of the elasticity of substitution in each country.

Table 2: Average annual growth rate of the technology 1970-2005

Country	**Labor-Augmenting** $\gamma_{A_{L,i}}$	**Capital-Augmenting** $\gamma_{A_{K,i}}$	**Weighted Productivity** $(1-\delta)\gamma_{A_{L,i}} + \delta\gamma_{A_{K,i}}$	**TFP** EU KLEMS
AUS	-0.3%	2.1%	0.5%	0.5%
AUT	4,6%	-7,1%	1.1%	0.9%
BEL	-3.2%	6.6%	-0.1%	0.1%
CZE	-0.1%	0.4%	0.1%	0.1%
DNK	0.5%	0.1%	0.4%	0.3%
ESP	-1.1%	1.7%	-0.2%	-0.1%
FRA	4.6%	-8.1%	1.0%	0.9%
GER	3.5%	-3.8%	1.3%	1.2%
HUN	2.2%	3.8%	2.7%	2.7%
IRL	0.6%	2.7%	1.3%	-
ITA	1.4%	-1.9%	0.4%	0.5%
JPN	2.7%	-1.7%	0.9%	0.9%
KOR	3.6%	-0.8%	2.5%	2.7%
LUX	2.3%	-3.8%	0.0%	-0.1%
NLD	-0.1%	1.7%	0.4%	0.4%
PRT	-1.4%	1.3%	-0.7%	-
SWE	-2.3%	6.5%	0.1%	0.9%
UK	1.2%	-2.0%	0.3%	0.3%
U.S	2.8%	-3.4%	0.5%	0.7%

Note: Columns 2 and 3 of table 2 show the posterior means of the average annual growth rate of the labor- and capital-augmenting technology, respectively. Column 4 shows the posterior mean of a weighted average of the growth rate of both technologies. Column 5 shows the Total Factor Productivity reported by EU KLEMS .

Table 3: Country regressions of the growth rate of the output on common factors of the capital and labor augmenting technology

Country	R-squared
AUS	0.12
AUT	0.10
BEL	0.36
CZE	0.70
DNK	0.32
ESP	0.26
FRA	0.24
GER	0.33
HUN	0.59
IRL	0.17
ITA	0.69
JPN	0.20
KOR	0.74
LUX	0.32
NLD	0.34
PRT	0.30
SWE	0.48
UK	0.44
USA	0.34

Note: The table shows the posterior mean of the R-squared for each country. The R-squared is calculated from a regression of the output growth rate over the common factor of the labor-augmenting technology growth rate and the common factor of the capital-augmenting technology growth rate.

Table 4: Contribution to the labor share change.

Country	Period	Labor Share		Cumulative Contribution	
		Real	Model	Capital-labor ratio $\left(\frac{K}{L}\right)$	Bias technical change $\left(\frac{A_K}{A_L}\right)$
AUT	1980-2005	-8.98	-7.97	1.97	-9.95
BEL	1980-2005	-4.92	-4.63	1.36	-5.99
CZE	1995-2005	2.85	2,29	3.47	-1.18
DNK	1980-2005	-3.89	-3.75	-0.18	-3.57
ESP	1980-2005	-4.86	-4.58	-1.28	-3.30
FRA	1980-2005	-9.17	-8.16	1.84	-10.00
GER	1970-2005	-5.60	-5.30	3.87	-9.18
HUN	1995-2005	0.39	0.04	-0.10	0.14
IRL	1988-2005	-4.81	-4.87	-1.51	-3.35
ITA	1970-2005	-6.91	-6.53	18.46	-24.99
JPN	1973-2005	-0.35	-0.39	4.07	-4.46
KOR	1977-2005	0.81	0.72	6.13	-5.41
LUX	1992-2005	-2.42	-2.74	0.43	-3.18
NLD	1979-2005	-10.09	-8.88	-1.71	-7.16
PRT	1992-2005	-6.50	-6.08	-1.19	-4.90
SWE	1993-2005	-1.74	-2.99	-1.05	-1.94
UK	1970-2005	2.23	2.36	5.32	-2.97
U.S	1970-2005	-6.40	-6.32	2.39	-8.71

Note: Cumulative changes are in percentage points and represent the change in percentage points of the labor share between the period indicated by column 2. The values reported in column 3,4, and 5 come from the MCMC algorithm. Column 3 is the posterior mean of the labor share generated by the model. Column 4 is the posterior mean of the cumulative contribution of the changes in the capital-labor ratio to the labor share. Column 5 is the posterior mean of the cumulative contribution of the bias in technical change to the labor share.

Table 5: Directed technical change: Cross-country regression

	Mean	Std
$\hat{\alpha}_{OLS}$	0.013***	0.003
$\hat{\beta}_{OLS}$	0.400***	0.076
R-squared	0.200	0.010

Note: The table show the posterior mean and the posterior standard deviation of $\hat{\alpha}_{OLS}$, $\hat{\beta}_{OLS}$ and the R-squared of the following regression: $\bar{\gamma}_{A_{L,i}} - \bar{\gamma}_{A_{K,i}} = \alpha + \beta\rho_i\left(\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}\right) + \varepsilon_i$. The dependent variable $\bar{\gamma}_{A_{L,i}} - \bar{\gamma}_{A_{K,i}}$ is the posterior mean of the difference between the average annual growth rate of labor augmenting technology and the average annual growth rate capital augmenting technology in country i. The regressor $\bar{\gamma}_{L,i} - \bar{\gamma}_{K,i}$ is the difference between the human capital growth rate and the physical capital growth rate in country i. This regressions is calculated for each iteration of the MCMC.

Table 6: Directed technical change: Cross-country regression

	Mean	Std
$\hat{\alpha}_{OLS}$	0.010***	0.001
$\hat{\beta}_{OLS}$	0.500***	0.027
R-squared	0.300	0.030

Note: The table show the posterior mean and the posterior standard deviation of $\hat{\alpha}_{OLS}$, $\hat{\beta}_{OLS}$ and the R-squared of the following regression: $f_{L,t} - f_{K,t} = \alpha + \beta \left[\frac{1}{N}\sum_{i=1}^{N} \rho_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right] + \varepsilon_t$. The dependent variable $f^j_{L,t} - f^j_{K,t}$ is the difference between the common factor of the labor- augmenting technology and the common factor of the capital- augmenting technology in period t. The regressor $\left[\frac{1}{N}\sum_{i=1}^{N} \rho^j_i \left(\gamma_{L,it} - \gamma_{K,it}\right)\right]$ is the average across countries of the cross product of ρ_i and the difference between the observed growth rate of the human capital and the observed growth rate of the physical capital in country i for period t. This regressions is calculated for each iteration of the MCMC.