Cross-Section Dependence and Latent Heterogeneity to Evaluate the Impact of Human Capital on Country Performance: a Robust Nonparametric Frontier Model

Camilla Mastromarco* Léopold Simar§

January 29, 2016

Abstract: Human Capital has been recognized as the most important force behind economic growth of countries. However, the effect of this important growth factor on economic growth remains ambiguous due to endogeneity and latent heterogeneity. By using a dataset of 40 countries over 1970-2007, we estimate the global frontier and explore the channels under which human capital and time affect the production process and its components: impact on the attainable production set (input-output space), and the impact on the distribution of efficiencies. We extend existing methodological tools - robust frontier in non parametric location-scale models - to examine these interrelationships. We use a flexible nonparametric two-step approach on conditional efficiencies to eliminate the dependence of production inputs/outputs on common factors. We emphasize the usefulness of "pre-whitened" inputs/outputs to obtain more reliable measure of productivity and efficiency to better investigate the impact of human capital on the catching-up productivity process. Then, we take into account the problem of unobserved heterogeneity and endogeneity in the analysis of the influence of human capital on the production process by extending the instrumental nonparametric approach proposed by Simar, Vanhems and Van Keilegom (JoE 2015) to account also for cross section and time dependence.

JEL: C14, C13, C33, D24, O47.

Keywords: Robust Frontier, Nonparametric Location-Scale, , Human Capital, Latent Heterogeneity and Endogeneity, Heterogeneous Panels, Time-Varying Efficiency Measures, Environmental Factors.

---

*Dipartimento di Scienze dell'Economia, Universitá degli Studi del Salento. Centro ECOTEKNE, Via per Monteroni - 73100 LECCE - Italy. Phone +39832298779, fax +390832298757, E-mail: camilla.mastromarco@unisalento.it.

§Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain. Voie du Roman Pays 20 B - 1348 Louvain-La-Neuve - Belgium. Email: leopold.simar@uclouvain.be. Research support from IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged.

1 Introduction

US economic slowdown in the second half of the nineties and the slowdown following 2001 - more marked in Europe than in US -, leading many to question the recipe for endogenous self-sustained growth. The understanding of the sources of growth may mirror the larger debate between the neoclassical and new growth theories, but economists generally agree that this recent economic decline has largely been caused by the weak growth in TFP, i.e., that part of the rise in productivity which is neither due to the increase in capital nor to the rise in the labour.

The "new" growth theory of Lucas (1988), Romer (1990a) and Barro (1990) and Barro (1997) has human capital playing an important role in productivity growth because human capital can help in explaining an economy's capacity to absorb new technologies (Abromovitz 1986, Cohen and Levinthal 1989, Kneller 2005, Kneller and Stevens 2006). For example, Benhabib and Spiegel (1994)' empirical study suggests that human capital plays a role in economic growth by helping in the adoption of technology from abroad and in creating the appropriate domestic technology. Hence, according to these studies, human capital is the most important force behind economic growth of countries.

Moreover, in response to the question of how technology diffusion affects economic growth, there has been an emerging empirical literature examining the nexus between the channels of technology diffusion and human capital in promoting economic growth. The evidence on this issue is mixed as seen in the differing conclusions of Miller and Upadhyay (2000, 2002) and Olofsdotter (1998). However, a consistent feature of all the empirical studies on this issue is the use of a cross- country regressions framework on a sample of developed and developing countries. Cross-country regressions cannot control for the unobservable heterogeneity which can arise, for example, from the different institutions in the various countries. Furthermore, Rodriguez (2006) argued that policy analysis within the growth-regression framework can carry considerable risks from the misspecification bias that come from using such a specification when it is not valid.

Therefore, the effect of this important growth factor - human capital - on economic growth remain ambiguous due to endogeneity and latent heterogeneity. In order to avoid the pitfalls of these parametric cross-regression studies we use an alternative empirical methodology, robust frontier in non parametric location-scale models to estimate the global frontier of 40 countries over 1970-2007, to answer our research question of whether a countries' productivity is affected by their existing levels of human capital.

We also control for the effect of openness on productivity. In particular we include in our analysis FDI as the most important openness channel for technology diffusion. FDI leads to increases in productivity by spurring competition and transferring technology. New foreign competition arrivals provide domestic firms an incentive to use existing resources more efficiently which increases their productivity. Consequently, foreign firms have to invest even more in order to keep up with their technological advantage (Glass and Saggi 1998). FDI can also increase productivity through the transfer of technology. This occurs with the adoption of new technology brought by foreign multinational companies, imports of high-technology inputs, and the skills acquired by the local labour force as they are educated and trained by the foreign firms (for the empirical evidence on the effect of FDI on productivity see Mastromarco and Simar 2015).[1]

In addition, by using nonparametric frontier methodology we can also study the channels under which human capital affects the production process and its components: impact on the attainable production set (input-output space), and the impact on the distribution of efficiencies. We take into account the problem of unobserved heterogeneity and endogeneity in the analysis of the influence of human capital on the production process by extending the instrumental nonparametric approach proposed by Simar et al. (2015) to account also for cross section and time dependence.

We use a flexible nonparametric two-step approach on conditional efficiencies to eliminate the dependence of production inputs/outputs on common and external factors. We emphasize the usefulness of "pre-whitened" inputs/outputs to eliminate cross-section dependence and observed heterogeneity and to obtain more reliable measure of productivity and efficiency to better investigate the impact of human capital on the catching-up productivity process. Then, we take into account the problem of unobserved heterogeneity and endogeneity in the analysis of the influence of human capital on the production process by extending the instrumental nonparametric approach proposed by Simar et al. (2015) to account also for cross section and time dependence.

2 The Methodology

We apply Simar et al. (2015) methodology and consider a Data Generating Process (DGP) characterizing the production process in the presence of observed environmental factors and unobserved and latent heterogeneity, we extend their models to a dynamic framework to allow

[1]Using these arguments, Borensztein et al. (1998), De Mello (1999) and Xu (2000) conclude that FDI increases an economys productive efficiency. Javorcik (2004) argues that FDI can also raise productivity growth through vertical spillovers rather than horizontal spillovers.

the introduction of the time dimension and cross sectional dependence (CSD). Consider a generic input vector $X \in \mathbb{R}^p_+$, a generic output vector $Y \in \mathbb{R}_+$ and we will denote by $Z \in \mathbb{R}^r$ the generic vector of environmental variables (FDI in our study). Since we are in a context of panel data, our sample will be denoted by (X_{it}, Y_{it}, Z_{it}), with $i = 1, \dots, n$ being the firm index and $t = 1, \dots, s$ the time index. Moreover, one of production factors, human capital "X^H" in our case, suffers from endogeneity and latent heterogeneity, because it is correlated with the environmental variables affecting the heterogeneity in the production process which is not observed.

To better investigate the influence of globalization factors (e.g., technological shocks and financial crises) on the economic performance of countries under analysis, we envelop the effect of CSD on the production process (**?**). Hence, we assume that the production process is function of unobserved time-varying factors. As proposed by Pesaran (2006), Bai (2009) we will consider $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ as proxy for the unobserved nonlinear and complex trending patterns associated with globalisation and the business-cycle.[2]

2.1 Frontier models in presence of observable external factors and cross sectional dependence

We start summarizing the setup of frontier models in presence of observable heterogenous factors Z and global factors (cross sectional dependence) F_t (see Mastromarco and Simar 2015) to then extend our framework at the case of unobservable heterogenous variables.

The production process is a process generating inputs $X \in \mathbb{R}^p_+$ and outputs $Y \in \mathbb{R}_+$; we define the unconditional (marginal) attainable set of feasible combinations of inputs and outputs as $\Psi = \{(x, y) \in \mathbb{R}^{p+1}_+ | x \text{ can produce } y\}$.

When we want to condition the frontier analysis to some environmental factors (Z, F_t), as is our setup here, we define the attainable set $\Psi^{z,f_t} \subset \mathbb{R}^{p+1}_+$ as the support of the conditional probability (Cazals et al. 2002):

$$H_{X,Y|Z,F_t}(x, y|z, f_t) = \text{Prob}\left(X \leq x, Y \geq y \,|\, Z = z, F_t = f_t\right). \tag{1}$$

Accordingly, and following Daraio and Simar (2005), when the output is univariate, the conditional frontier function at input x, facing conditions z and f_t (in particular at time t),

[2]Here we use the standard notation where a dot in a subscript, means that we averaged over this index.

is defined as[3]

$$\tau(x,z,f_t) = \sup\{y|S_{Y|X,Z,F_t}(y|x,z,f_t) > 0\}, \tag{2}$$

where $S_{Y|X,Z,F_t}(y|x,z,f_t) = \text{Prob}(Y \geq y|X \leq x, Z = z, F_t = f_t)$ (note the difference in the conditioning for X, the inputs, and for Z and F_t, the environmental and global factors; to respect the vfor Z but not for X). Again we can report the Farrell-Debreu conditional efficiency scores as

$$\lambda\,(x,y|z,f_t) = \tau(x,z,f_t)/y = \sup\{\lambda|S_{Y|X,Z,F_t}(\lambda y|x,z,f_t) > 0\}. \tag{3}$$

A nonparametric estimator of the conditional survival function $S_{Y|X,Z,F_t}(y|x,z,f_t)$ could be obtained by using standard smoothing methods where a bandwidth h has to be determined for each component of (Z,F_t) (as e.g. in Badin et al., 2010). In summary, these nonparametric estimators are consistent with rate $n^{1/(p+1)}$ and Weibull limiting distribution for the unconditional FDH (see Park et al., 2000). For the conditional case, we have similar results where n is replaced by nh^d where d is the dimension of all the conditioning variables (Z,F_t) (see Jeong et al., 2010). So the rates of convergence of the conditional estimators are deteriorated by the dimension d.

In most of the empirical examples, a naive application of these nonparametric techniques may be problematic because real samples contain in general some anomalous data. In that case, the estimated frontier is fully determined by these outliers or extreme data points and the measurement of inefficiencies are totally unrealistic. Whereas most of the practitioners use a rule of thumb for outliers elimination, better approaches have been proposed in the frontier literature (Cazals et al., 2002; Daouia and Simar, 2007) to keep all the observations in the sample but to replace the frontier of the empirical distribution by (conditional) quantiles or by the expectation of the minimum (or maximum) of a subsample of the data. This latter method defines the order-m frontier that we will use here. To be short, the partial output-frontier of order-m is defined for any integer m and for an input x, as the expected value of the maximum of the output of m units drawn at random from the populations of firms using less inputs than x. Formally

$$\tau_m(x) = \mathbb{E}\left[\max(Y_1,\ldots,Y_m)\right], \tag{4}$$

where the Y_j are independently distributed as $S_{Y|X}(\cdot|X \leq x)$. The same applies for the

[3]We only focus the presentation on the output orientation version of the estimators, the same could be done for any other orientation (input, hyperbolic, directional distance).

conditional order-m frontier $\tau_m(x, z, f_t)$ where the Y_j are distributed as $S_{Y|X,Z,F_t}(\cdot|X \leq x, Z = z, F_t = f_t)$. Nonparametric estimators are obtained by plugging the nonparametric estimators of the survival functions in (4).

If m increases and converges to ∞, it has been shown (see Cazals et al., 2002) that the order-m frontier and its estimator converge to the full frontier, but for a finite m, the frontier will not envelop all the data points and so is much more robust than the FDH to outliers and extreme data points (see e.g. Daouia and Gijbels, 2011, for the analysis of these estimators from a theory of robustness perspective). Another advantage of these estimators is that they achieve the parametric rate of convergence $\sqrt{n}$ and that they have a normal limiting distribution.

To clean the effect of global factors F_t, and i.e. cross section dependence and eternal observed factor FDI, we will rather follow the approach suggested in Florens et al. (2014) which avoids direct estimation of the conditional survival function $S_{Y|X,Z,F_t}(y|x, z, f_t)$. As pointed by Florens et al., the procedure is less impacted by the curse of dimensionality (of the conditioning variables Z, F_t) and requires smoothing in these variables in the center of the data cloud and so avoiding smoothing at the frontier where typically the data are rather sparse and estimators are more sensitive to outliers. Moreover the inclusion of time factor $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ enables us to eliminate the common time factor effect, in a very flexible nonparametric location-scale model.[4]

We thus assume that the data are generated by the following nonparametric location-scale regression model

$$\begin{cases} X_{it} &= \mu_x(Z_{it}, F_t) + \sigma_x(Z_{it}, F_t)\varepsilon_{x,it} \\ Y_{it} &= \mu_y(Z_{it}, F_t) + \sigma_y(Z_{it}, F_t)\varepsilon_{y,it} \end{cases}, \tag{5}$$

where μ_x, σ_x and ε_x have each p components and, for ease of notations, the product of vectors is componentwise. So the first equation in (5) represents p relations, one for each component of X. We assume that each element of ε_x and ε_y have mean zero and standard deviation equal to 1. The model also assume that $(\varepsilon_x, \varepsilon_y)$ is independent of (Z, F_t).

This model allows us to capture for any (z, f_t), for each input, $j = 1, \ldots, p$ and for the output, the locations $\mu_x^{(j)}(z, f_t) = \mathbb{E}\left[(X^{(j)}|Z = z, F_t = f_t)\right]$, $\mu_y(z, f_t) = \mathbb{E}\left[(Y|Z = z, F_t = f_t)\right]$ and the scale effects $\sigma_x^{(j),2}(z, f_t) = \mathbb{V}\left[(X^{(j)}|Z = z, F_t = f_t)\right]$, $\sigma_y^2(z, t) = \mathbb{V}\left[(Y|Z = z, F_t = f_t)\right]$

[4]The statistical properties of the resulting frontier estimators are established in Florens et al. (2014). In particularly, they demonstrate the consistency for the full-frontier FDH estimator and $\sqrt{n}$-consistency and asymptotic normality for the robust order-m frontiers. Moreover, they prove that, if the functions μ_ℓ and σ_ℓ for $\ell = 1, 2$, are smooth enough, the FDH estimator would keep its usual nonparmetric rate of convergence i.e. $n^{1/(p+1)}$.

of the environmental and common factors on the production plans.[5]

As explained in Florens et al. (2014), ε_x and ε_y can be interpreted as "pure" inputs and output, because due to the independence between the vector $(\varepsilon_x, \varepsilon_y)$ and (Z, F_t), they can be viewed as "whitened" versions of X and Y respectively. Since no particular assumption is made on the distribution of $(\varepsilon_x, \varepsilon_y)$, the model remains basically nonparametric. Note also that in the case where (Z, F_t) would be independent of all the inputs X and of the output Y, the functions μ_ℓ and σ_ℓ would be constant for $\ell = x, y$ and $(\varepsilon_x, \varepsilon_y)$ would simply be a standardized version of the original inputs and output.

To estimate the production frontier we follow the method in two stages proposed by Florens et al. (2014). In the first stage we estimate model (5) by using some usual non-parametric techniques (e.g. local constant or local linear): (i) estimation of the location functions $\mu_\ell(z_{it}, f_t)$ and (ii) estimation of the variance functions $\sigma^2_\ell(z_{it}, f_t)$ by regressing the square residuals, resulting from the location regression, on (z, f_t). For the location we use local linear and for the variance local constant to avoid negative values of the estimated variances. From this first analysis we obtain the residuals

$$\widehat{\varepsilon}_{x,it} = \frac{X_{it} - \widehat{\mu}_x(Z_{it}, F_t)}{\widehat{\sigma}_x(Z_{it}, F_t)}, \tag{6}$$

$$\widehat{\varepsilon}_{y,it} = \frac{Y_{it} - \widehat{\mu}_y(Z_{it}, F_t)}{\widehat{\sigma}_y(Z_{it}, F_t)}, \tag{7}$$

where for ease of notation, a ratio of two vectors has to be understood component wise. These are the whitened inputs and output obtained by eliminating the influence of the external and other environmental variables as common factors. In practice we will need to test the independence between $(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it})$ and (Z_{it}, F_t), i.e. the independence of whitened inputs and output from the external and global effects to validate the location-scale model (see Florens et al. 2014, for a bootstrap based testing procedure).

2.2 Latent Heterogeneity and Endogeneity

In this paper, for estimating the conditional measures we take also into account the problem of latent heterogeneity and endogeneity - due to omitted relevant variables - and follow the approach suggested in Simar et al. (2015). In particularly, we decompose our input vector $X = \left(X^K, X^L, X^H \in \mathbb{R}^p_+\right)$ and we assume that human capital, X^H, is linked to the unobserved variable V. The neglect of a latent factor V, or unobserved global factors F_t

[5] Hereafter, for a vector a, $a^{(j)}$ denotes its j^{th} component.

(hence cross section dependence) or an observed one Z will cause an endogeneity problem (see Simar et al. 2015).

We will solve the problem of the latent factor V by assuming that the relationship between our input $X^H$and V is through the instrumental variable W:

$$X^H = \phi(W, V) \tag{8}$$

where W is an observed variable correlated with X^H and independent of V. As explained by Simar et al. (2015), under the assumption of monotonicity of ϕ with repsect to V and uniform distribution of V on $[0, 1]$, V can be identified by conditional distribution of X^H given the instrument W:

$$V = F_{X^H|W} \tag{9}$$

The latent variable V captures the part of the observable variable X^H independent from the instrument W. In our empirical application, human capital H (measured as average years of education in the population) impacts the productivity of a country. However, there may exist unobserved characteristics of the countries that may influence both the existing level of education and the productivity, as the quality of institutions. To take into account this latent heterogeneity we use life expectancy as instrumental variable W. The estimated V captures the variation of human capital which does not depend on life expectancy. This estimated unobservable heterogeneity, correlated with human capital, might be interpreted as absorptive capability, defined as the potential to master new knowledge embodied in innovation, which acts as free disposal input. The innovation has similar attributes as a public good, indeed it is non rival and non exclusive. Given these features, there exists the problem to protect the property rights of innovations which would guarantee profits to the innovators and stimulate new efforts to innovate. Hence, it is very likely that, this unobservable factor is capturing the quality of institutions (e.g. very high in US, very low in Mexico) and the difference in property rights systems among countries.

We estimate the pure frontier and pure efficiency by conditioning on unobserved ($\widehat{V}$) variables whose predicted value for each observation is given by (see Simar et al. 2015):

$$\widehat{V}_i = \widehat{F}_{X^H|W} = \frac{\sum_{j=1}^{n} \mathbb{1}(X_{H,J} \leq X_{H,i}) K_{h_w}(W_i - W_j)}{\sum_{j=1}^{n} K_{h_w}(W_i - W_j)} \tag{10}$$

Define the attainable conditional set of pure inputs and output $(\varepsilon_x, \varepsilon_y)$ as

$$\Psi_\varepsilon^{\hat{v}} = \left\{(e_x, e_y|\hat{v}) \in \mathbb{R}^{p+1}|H_{\varepsilon_x,\varepsilon_y|\widehat{V}}(e_x, e_y|\hat{v}) = \text{Prob}(\varepsilon_x \leq e_x, \varepsilon_y \geq e_y|\ \widehat{V}) > 0\right\}.$$

The nonparametric FDH estimator is obtained by plugging the empirical estimators $\widehat{H}_{\varepsilon_x,\varepsilon_y}(e_x, e_y|\hat{v})$ obtained with the observed residuals defined in (6) and (7). As shown in Florens et al. (2014), replacing the unobserved $(\varepsilon_x, \varepsilon_y)$ by their empirical counterparts $(\widehat{\varepsilon}_x, \widehat{\varepsilon}_y)$ does not change the usual statistical properties of frontier estimators. So we have the consistency for the full-frontier FDH estimator and $\sqrt{n}$-consistency and asymptotic normality for the robust order-m frontiers. It is conjectured in Florens et al. (2014), that if the functions μ_ℓ and σ_ℓ for $\ell = x, y$, are smooth enough, the conditional FDH estimator would keep its usual nonparametric rate of convergence i.e. $n^{1/(p+1)}$.

Accordingly, the conditional pure output oriented technical efficiency of a production plan $(\varepsilon_{x,it}, \varepsilon_{y,it})$ facing conditions $\hat{v}$, is defined by measuring the distance of a particular point $(\varepsilon_{x,it}, \varepsilon_{y,it})$ to the conditional efficient frontier. Since the pure inputs and output are centered on zero, they may have negative values and so radial distances are inappropriate. We should rather use directional distances defined for a particular unit (e_x, e_y) as

$$\delta(e_x, e_y|\hat{v}; d_x, d_y) = \sup\{\gamma|H_{\varepsilon_x,\varepsilon_y|\widehat{V}}(e_x - \gamma d_x, e_y + \gamma d_y|\widehat{V} = \hat{v}) > 0\}, \tag{11}$$

where $d_x \in \mathbb{R}_+^p$ and $d_y \in \mathbb{R}_+$ are the chosen direction. In our case here we choose an output orientation so that $d_x = 0$ and we can choose $d_y = 1$, for more general cases, see Simar and Vanhems (2012) (if only some elements of $d_x = 0$ see Daraio and Simar, 2014 for practical computations). So, for this particular output direction and in the case of univariate output we follow here, the optimal production frontier can be described at any value of the pure input $e_x \in \mathbb{R}^p$, by the function

$$\varphi(e_x|\hat{v}) = \sup\{e_y|H_{\varepsilon_x,\varepsilon_y|\widehat{V}}(e_x, e_y|\hat{v}) > 0\}, \tag{12}$$

so that the distance to the frontier of a point (e_x, e_y), in the output direction, is directly given by $\delta(e_x, e_y|\hat{v}; 0, 1) = \varphi(e_x|\hat{v}) - e_y$. Then, for each units in the sample, the "pure" efficiency estimator is obtained through

$$\widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}|\hat{v}; 0, 1) = \widehat{\varphi}(\widehat{\varepsilon}_{x,it}|\hat{v}) - \widehat{\varepsilon}_{y,it}, \tag{13}$$

where $\widehat{\varphi}(\cdot)$ is the conditional FDH estimator of the pure efficient frontier in the output direction. It is simply obtained as

$$\begin{aligned}\widehat{\varphi}(e_x|v) &= \sup\{e_y|\widehat{H}_{\varepsilon_x,\varepsilon_y|\widehat{V}}(e_x,e_y|v)>0\} \\ &= \max_{\{(i,t)|\widehat{\varepsilon_{x,it}}\le e_x\},|\widehat{V}-v|\le h_v} \widehat{\varepsilon}_{y,it}. \end{aligned} \tag{14}$$

Similar expressions can be derived for the order-m efficiency estimator. As explained above, the order-m frontier at an input value e_x, is the expected value of the maximum of the outputs of m units drawn at random in the population of units such that $\varepsilon_{x,it} \le e_x$ and $\widehat{V} = \hat{v}$. The nonparametric estimator is obtained by looking to its empirical version:

$$\widehat{\varphi}_m(e_x\hat{v}) = \widehat{E}\left[\max\left(\varepsilon_{y,1t},\ldots,\varepsilon_{y,mt}\right)|\widehat{V}=\hat{v}\right], \tag{15}$$

where the $\varepsilon_{y,it}$ are drawn from the empirical conditional survival function $\widehat{S}_{\varepsilon_y|\varepsilon_x,\widehat{V}}(e_y|\widehat{\varepsilon}_{x,it} \le e_x, \widehat{V} = \hat{v})$. This can be computed by Monte-Carlo approximation or by solving a univariate numerical integral (for practical details see Simar and Vanhems 2012, Daraio and Simar 2014).

2.3 Effect of $\widehat{V}$ on the production process

To investigate the impact of the latent variable $\widehat{V}$ on the production process, we compare the conditional measures $\widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}|\hat{v}; 0, 1)$ with the unconditional measures $\widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}; 0, 1)$. We follow the procedure described in details in Daraio and Simar (2014) where they adapt the methodology of Bădin et al. (2012) for radial oriented distances to the direction distances case. The procedure allows to disentangle the potential effects of $\widehat{V}$ on the boundary (shift of the frontier) and on the distribution of the inefficiencies.

The first effect can be investigated by considering the difference of unconditional to conditional efficiency scores, which are measures relative to the full frontier of respectively, the unconditional and the conditional attainable sets. So we have

$$R_O(e_x, e_y|\widehat{V}) = \widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}) - \widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}|\hat{v}). \tag{16}$$

By construction, for the output orientation, $R_O(e_x, e_y|\hat{v}) \ge 0$ (the conditional efficient boundary is below the unconditional one). Looking to these differences as a function of $\widehat{V}$ allows to investigate the effect of innovation and quality of institutions of a country on this potential shift.[6]

A global tendency of the differences to decrease with the conditioning variables indicates a positive effect (the conditional efficient frontier moves up to the unconditional one when the variables increase, i.e. the variables act as freely available inputs) and unfavorable in the opposite case (the conditional efficient boundary moves away form the marginal one when the variables increase, the variables act as undesirable outputs).

As illustrated in Daraio and Simar (2007), some extreme or outlying data points may hide the real effect of $\widehat{V}$, so it is suggested to do the same analysis with our order-m frontier, with large values of m to get robust estimates of the full frontier (we discuss in the application how to select m for this purpose). In this case, the difference to be analyzed are given by

$$R_{O,m}(e_x, e_y|\widehat{V}) = \widehat{\delta}_m(e_x, e_y) - \widehat{\delta}_m(e_x, e_y|\hat{v}). \tag{17}$$

As pointed in Bădin et al. (2012), the full frontier differences, or their robust version with large values of m, indicate only the influence of the latent variable $\widehat{V}$ on the shape of the frontier, whereas the partial frontiers for small values of m, characterizes behavior of the shift more in the center of the distribution of efficiencies, inside the attainable sets. For instance if $m = 1$, the order-m frontier turns out to be an average production function and the differences (17) would analyze the shift of the mean of the distribution of the inefficiencies. Some potential shifting effect already observed with (16) could be enhanced (or reduced) if the effect is different with the differences (17).

But, as for the full differences above, a tendency of $R_{O,m}(e_x, e_y|\widehat{V})$ to increase with the conditioning variables indicates a negative effect of these variables on the distribution of the efficiencies (the conditional distribution is less concentrated to its upper boundary when the conditioning variables increase) and the opposite in the case of a favourable effect. If this effect is similar to the one shown with the differences with full frontier, we can conclude that we have a shift of the frontier while keeping the same distribution of the efficiencies when the conditioning variable $\widehat{V}$ change; if the effect with the partial frontiers is more important than for the full frontier, this indicates that in addition to a shift of the frontier, we have also an effect on the distribution of the efficiencies. Moreover, we follow the procedure and algorithm illustrated in Daraio and Simar (2014) for testing the significance of the $\widehat{V}$ variable on the average efficiency scores.

[6] As suggested in our previous paper (Mastromarco and Simar 2015), the effect could be different for different values of X (possibility of interactions), and, hence, the analysis of these differences has to be done for fixed levels of the inputs x.

3 Empirical Application

3.1 The data and the variables

Our non parametric approach in constructing the worldwide production frontier does not require the specification of the production functional form, and also limit the problem of 'curse of dimensionality' at the second stage of our methodology. In addition, we provide an analysis which is robust to extreme or outlying data points that might hide some features of the production process. We consider the simplest production model with only four macroeconomic variables: aggregate output and three aggregate inputs (labour, capital and human capital).

Human capital is dealt differently from the other two more conventional production factors (capital K and labour L). Indeed, human capital may affect the production process in two way: it may influence the the attainable set: $\Psi^z = \{(x, y) \mid Z = z,\ x \text{ can produce } y\}$, as a production input (Mankiw et al. 1992), or the efficiency distribution as productivity factor (Lucas 1988, Romer 1990a).

The dataset is collected over the period, 1970-2007 (38 years) for a total of 40 countries using data from the Penn; 26 are developed OECD countries (Australia, Austria, Belgium, Canada, Chile, Hong Kong, Denmark, Finland, France, Germany, Greece, Ireland, Israel, Italy, Japan, Korea, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Turkey, the United Kingdom and the United States) and 14 are developing countries (Argentina, Bolivia, Côte d'Ivoire, Dominican Republic, Honduras, Jamaica, Kenya, Malawi, Morocco, Philippines, Thailand, Venezuela, Zambia, Zimbabwe).[7]

We use data from the Penn World Tables (version 8) where output is the real gross domestic product $RGDP$ measured in million US dollars at 2005 constant prices. For labor input, we use the number of workers EMP. Capital stock K which is our chosen input is then measured in million US dollars at 2005 constant prices. All three variables are rescaled to get a standard deviation of 1 and then transformed in logarithms before estimation.

For human capital we use the variable H defined as "index of human capital per person, based on years of schooling (Barro and Lee 2012) and returns to education (Psacharopoulos 1994)". The purpose of this study is to re-exam the growth effect of human capital. We thus acknowledge the endogenous growth models of Lucas (1988) and Romer (1990a) that use a theoretical framework where persistent economic growth is conditional on the accumulation

[7]The choice of countries depends on data availability, the constraint variable is human capital. Developed and developing countries are classified following the World Bank (2007) classification.

of human capital.[8] The empirical growth literature emphasises the latent heterogeneity and endogeneity of human capital in the economic process, due to unobserved characteristics, which may impact both the level of education and the level of output. This problem leads often to counter-intuitive results due to the difficulty to properly assess the impact of this important economic growth driver on country's output. Indeed, many empirical papers find not significant or even negative impact of human capital on economic growth.[9] To take into account the unobserved heterogeneity and endogeneity of human capital we use life expectancy as instrument. This variable (from World Bank Indicators) is measured as "life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life".

For globalization factor we identify one of the most important channels: FDI inflows, measured as net inflows of foreign direct investment, which are then transformed as a ratio to GDP.[10] This external variable FDI might suffer from endogeneity bias. The endogeneity caused by reverse causality is still an open issue in the empirical studies investigating the relationship between total factor productivity (TFP) and FDI. We explicitly address this issue by eliminating in the first stage the dependence of the FDI on the production process. Furthermore, the global economy becomes increasingly integrated, all the individual countries are likely to be exposed more to global shocks. As explained above we follow Pesaran (2006) and Bai (2009) and consider $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ as proxy for these common factors.

[8]The new endogenous growth theories (Aghion and Howitt 1992); Romer (1990a) describe human capital as the engine of growth through innovation. Grossman and Helpman (1991) show that the skill composition of the labor force matters for the amount of innovation in the economy. In particular, they obtain that an increase in the stock of skilled labor is growth-enhancing while an increase in the stock of unskilled labor can be growth-depressing. Benhabib and Spiegel (1994), Tallman and Wang (1994) find that the channel through which human capital positively affects output is through the efficiency enhancing effect. Recent contributions emphasise the different roles that different types of human capital may play in either backward or advanced economies (Caselli and Coleman, 2006), and the distinction between innovation activities and adoption of existing technologies from the (world) technology frontier (Acemoglu et al., 2006). In this context, low-skilled human capital appears better suited to the adoption of technology in low-income countries, while skilled human capital has a growth enhancing impact which increases with the level of development (Caselli and Coleman 2006, Vandenbussche et al. 2006).

[9]Romer (1990b), Benhabib and Spiegel (1994) and Barro (1997) find a significantly positive effect of schooling levels on output growth, while Cohen and Soto (2001) find no link. Temple (1999) and De La Fuente and Domenech (2001) find a significantly positive correlation between improvements in education and growth, while Barro and Sala-i-Martin (1995), Caselli et al. (1996), and Pritchett (2001) find no effect of schooling improvements on growth. Topel (1999) and Krueger and Lindahl (2001) find both education level and improvement effects on growth.

[10]FDI is sourced from the World Bank World Development Indicators and Unctad, all the other data from PWT 8. The observation period is selected by the data availability.

3.2 Effect of human capital on the production process

In our framework we consider that there is an unobserved heterogeneity, as quality of institutions (e.g. different property right systems) and different absorptive capability among countries, which determines the level of innovation and, hence, it affects our production data generation process. We assume that these latent factors are linked to one of our production factor and, specifically, to human capital variable.

To identify this non-observable factor which is related to the level of human capital, we use life expectation as instrument. Indeed, the human capital is linked to the life expectation but there exists a non-observable factor which may also influence the level of human capital in a country as quality of institutions. Therefore, by following Simar et al. (2015) we use life expectancy as an instrument W to identify the latent factor. We calculate the predicted value $\widehat{V}$ for each observation given by equation 10 as explained in section 2.2. We then estimate the conditional output-oriented efficiency, conditional on observed and non-observed factors.

The latent factor $\widehat{V}$ proxies for the aggregate effect of human capital not explained by life expectation - our instrument W -. In our macroeconomic contest, this component would be related, for example, to the influence of institutions as difference in property rights systems or absorptive capability, defined as the potential to master new knowledge, among countries. Figure 1 shows the distributions and Table 1 summarizes descriptive statistics for the latent factor for 40 countries. The highest values over the sample period are registered by USA, New Zealand and Australia, whereas France, Portugal, Italy and Spain stand at the bottom of the ranking. These low values are mainly due to the fact that, in these countries, the life expectancy is very high. We may infer that these countries benefit from the increase in human capital more in terms of highest life expectancy than in better quality of their institutions.

To explore if our predicted latent variable are somehow linked to absorptive capability, defined as the potential to master new knowledge and, hence, to the level of innovation in a country, in Tables 2 and 3 we report the same statistics for number of patents (for thousand people) and number of researchers in R&D (for million people) for the available countries and periods. The USA, New Zealand and Australia rank at the top of the distribution for the number of patents and researchers in R&D (except USA that, for this last indicator, displays an average value). This evidence seems to confirm that our latent variable is somehow connected to the level of innovation. To give a visual impression of the change in latent factor over time, average value for each year is displayed in Figure 2 for United States, New Zeeland, Australia, Morocco, Spain, Belgium and Italy.

To take into account the impact on the production process of observed and non-observed common factors, we follow the two-stage estimation procedure described above which enables us, in the first stage, to better capture the impacts of global shocks (such as FDI, trade policy and cycle fluctuations) and, hence, CSD on the world production frontier and technical efficiency. By applying the estimation of the models (5) to our transformed data, we obtain by equations (6) and (7) the "pure" versions of our inputs , $\widehat{\epsilon}_x$ (capital, labour and human capital) and of the output $\widehat{\epsilon}_y$ (GDP). Before looking for frontier estimates we have to verify if the "pure" inputs-output $(\varepsilon_x, \varepsilon_y)$ are independent of the conditioning variables (F_t, Z). Table 4 reports the correlations between $\varepsilon_x, \varepsilon_y$ and time factors $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ and FDI. They are very small indicating that our first stage location-scale model has cleaned most of the effects of these variables and confirming that the influence of FDI and the cross section dependence has been removed from our data.

The estimation of the world production frontier then follows in the second stage. The full frontier estimate is the FDH of the preceding cloud of points and was defined above as $\widehat{\varphi}(e_x|\hat{v})$. For the robust version of the full frontier, we select $m = 500$ (this leaves 33% of data points above the corresponding order-m frontier).

To assess the influence of latent variable $\widehat{V}$ defined as "the part of the human capital which is not related to the life expectancy" on the production process, we investigates the differences of conditional and unconditional efficiency measures for full and partial frontier as discussed in Section 2.3. The results of the potential effects of the latent variable $\widehat{V}$ are shown in Figure 3 for the full differences and the diffferences (17) for $m = 500$ (robust version of the full differences) and in Figure 4 for $m = 1$ to assess the influence of $\widehat{V}$ on the average of the inefficient distribution.

The main messages of these pictures is as follows. To investigate the effect of $\widehat{V}$ on the shift of the frontier, we have to analyze the differences for the full frontier, and its robust version (Figure 3). First we see that the order-m results are very similar to the full frontier results, this confirms, than in our data set, the most outlying points are not too influential for global analysis. Second, we see for left and right panels a linear shape with the level of the differences very low and near to 1 that, in our setup, we can interpret that we have not shifts of the frontiers when human capital increases. So, $\widehat{V}$ does not act on the shift of the boundary. Hence, from this evidence, $\widehat{V}$ appears not to play an important role in accelerating the technological change (shifts in the frontier).

The Figure 4 allows, when compared to the Figures 3, to identify some changes in the distribution of the inefficiencies due to $\widehat{V}$. Globally, we can see some changes in the shape of the clouds of points and we observe a clear decreasing trend of differences with respect to $\widehat{V}$. The level of these differences is far from 1. Combining this result with the previous on the shift of the technology, we could interpret as the fact that $\widehat{V}$, which captures absorptive capacity, thus the capability to assimilate innovations, induces catching-up to the production frontier, but not necessarily shifts of it. This seems to suggest that countries benefit from new technology only when they have the ability to exploit it. Countries can switch to a better technology if they accumulate the technology-specific expertise (Greenwood and Jovanovic 2001, Helpman and Rangel 1999).

Efficiency is the most important growth component for convergence analysis of countries that are below the technological frontier because it reflects "the process of imitation and transmission of existing knowledge" (Romer 1986). Quah (1997), Mankiw et al. (1992), Barro and Sala-i-Martin (1995) argue that slow convergence in the level of output per worker is caused by slow technological catch-up. This latent variable which captures the part of human capital not linked to life expectancy, i.e. might capture absorptive capability, might increase efficiency and, hence, convergence. This occurs with the adoption of foreign technology through technology licensing or technology purchase, imports of high technology capital goods, and the skills acquired by the local labour force (Borensztein et al. 1998, De Mello 1999, Xu 2000).

Our findings support the convergence evolution of output among countries with respect to human capital. Our latent variable which captures the part of human capital not linked to life expectancy, i.e. might measure absorptive capability of a country, thus the capability to assimilate innovations, influences efficiency distribution and it acts as a transmission channel to diffuse technology. Hence, it induces catching-up to the production frontier, but not necessarily shifts of it.

4 Conclusion

The recent economic slowdown first in USA during late nineties and then in Europe in 2001 leads the economists to question the recipe for endogenous self-sustained economic growth. Economic growth literature emphasizes the importance of human capital in spurring pructivity growth. Moreover, the productivity analysis recognizes the importance of considering the spillover effects of global shocks and business cycles due to increasing globalization and interconnection among countries.

So far all studies analysing effect of human capital on productivity of countries have produced ambiguous empirical results due to endogeneity and latent heterogeneity. Many empirical studies have been on the stream of parametric modelling which suffers of misspecification problems when the data generating process is unknown, as usual in the applied studies.

In order to avoid the pitfalls of these parametric cross-regression studies we propose an alternative empirical methodology, robust frontier in non parametric location-scale models for accommodating simultaneously the problem of model specification uncertainty, potential endogeneity and cross-section dependence in modelling technical efficiency in frontier models.

The neglect of a latent factor, or unobserved global factors (hence cross section dependence) or an observed one will cause an endogeneity problem (Simar et al. 2015). We combine two different methodology, the two-step procedure advanced by Florens et al. (2014), which enables us to deal with both observed heterogeneity and cross section dependence by combining location scale model and conditional efficiency estimation and Simar et al. (2015) which handles the endogeneity due to unobserved heterogeneity.

Our non parametric approach to estimate conditional efficiency does not require any parametric assumption regarding technology or efficiency term. Moreover, the assumption of complete homogeneity of considered units is not needed. Therefore the economic units under investigation, can potentially consist of different groups of population governed by different distributional laws of the generation of input-output mix and on efficiency. This is an advantage in our sample formed by developed and developing countries which most likely have different distributions of efficiency scores.

Moreover our frontier model enables us to see whether the effect of environmental/global variables on productivity occurs via technology change or efficiency. We can then quantify the impact of environmental/global factors on efficiency levels and make inferences about the contributions of these variables in affecting efficiency.

The "new" growth theory of Romer (1990a), Lucas (1988) and Barro (1997) has human capital playing an important role in growth because human capital can help in explaining an economy's capacity to absorb new technologies (Abromovitz 1986, Cohen and Levinthal 1989, Kneller 2005, Kneller and Stevens 2006, Mastromarco and Ghosh 2009). Our paper extends previous studies on similar topics by investigating this channel in full nonparametric framework which avoids some restrictive and often unverifiable prior assumptions on functional relationships and distributions.

We focus on the effect of human capital on economic performance of 40 countries over the period 1970-2007. In a cross-country framework, production inefficiencies can be identified as the distance of the individual country's production from the frontier as proxied by the maximum output of the reference country (regarded as an empirical counterpart of an optimal production boundary). Hence, efficiency improvement will represent productivity catch-up via technology diffusion because inefficiencies generally reflect a sluggish adoption of new technologies (Ahn and Sickles 2000).

Our findings prove that human capital plays an important role in accelerating the technological catch-up (increase in the efficiency) but not on the technological changes (shifts in the frontier). This result seems to confirm the theoretical hypothesis that countries benefit from new technology (technological catch-up) only when they have the ability to exploit it, hence only when they have high level of absorptive capability. Countries can switch to a better technology if they accumulate the technology-specific expertise (Greenwood and Jovanovic 2001, Helpman and Rangel 1999).

References

Abromovitz, M.: 1986, Catching up, forging ahead, and falling behind, *Journal of Economic History* **46**, 386–406.

Aghion, P. and Howitt, P.: 1992, A model of growth through creative destruction, *Econometrica* **60**, 323–351.

Ahn, S. C. and Sickles, R. C.: 2000, Estimation of long-run inefficiency levels: A dynamic frontier approach, *Econometric Reviews* **19**, 461 – 492.

Bai, J.: 2009, Panel data models with interactive fixed effects, *Econometrica* **77**, 1229–1279.

Barro, R.: 1990, Government spending in a simple model of endogenous growth, *Journal of Political Economy* **98**, S103–S125.

Barro, R. J.: 1997, *Determinants of Economic Growth*, The MIT Press, Cambridge, MA.

Barro, R. J. and Lee, J.: 2012, A new data set of educational attainment in the world, 1950-2010, *Journal of Development Economics* **104**, 184–198.

Barro, R. J. and Sala-i-Martin, X.: 1995, *Economic Growth*, McGraw-Hill, New York.

Benhabib, J. and Spiegel, M.: 1994, The role of human capital in economic development. Evidence from aggregate cross-country data, *Journal of Monetary Economics* **34**, 143–173.

Borensztein, E., De Gregorio, J. and Lee, L.-W.: 1998, How does foreign direct investment affect economic growth?, *Journal of International Economics* **45**, 115–135.

Bădin, L., Daraio, C. and Simar, L.: 2012, How to measure the impact of environmental factors in a nonparametric production model?, *European Journal of Operational Research* **223**, 818–833.

Caselli, F. and Coleman, W. J.: 2006, The world technology frontier, *American Economic Review* **96**, 499–522.

Caselli, F., Esquivel, G. and Lefort, F.: 1996, Reopening the convergence debate: A new look at cross-country growth empirics, *Journal of Economic Growth* **1**, 363–389.

Cazals, C., Florens, J. and Simar, L.: 2002, Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* **106**, 1–25.

Cohen, W. and Levinthal, D.: 1989, Innovation and learning: Two faces of R&D, *Economic Journal* **107**, 139–149.

Daraio, C. and Simar, L.: 2005, Introducing environmental variables in nonparametric frontier models: A probabilistic approach, *Journal of Productivity Analysis* **24**, 93–121.

Daraio, C. and Simar, L.: 2007, Conditional nonparametric frontier models for convex and non convex technologies: A unifying approach, *Journal of Productivity Analysis* **28**, 13–321.

Daraio, C. and Simar, L.: 2014, Directional distances and their robust versions: Computational and testing issues, *European Journal of Operational Research* **237**, 358–369.

De La Fuente, A. and Domenech, R.: 2001, Schooling data, technical diffusion, and the neoclassical model, *American Economic Review Papers and Proceedings* **90**, 323–327.

De Mello, L. R. J.: 1999, Foreign direct investment-led growth: Evidence from time series and panel data, *Oxford Economic Papers* **51**, 133–151.

Florens, J., Simar, L. and van Keilegom, I.: 2014, Frontier estimation in nonparametric location-scale models, *Journal of Econometrics* **178**, 456–470.

Glass, A. J. and Saggi, K.: 1998, International technological transfer and technology gap, *Journal of Development Economics* **55**, 369–398.

Greenwood, J. and Jovanovic, B.: 2001, Accounting for growth, *in* T. U. o. C. P. NBER (ed.), *New Directions in Productivity Analysis*, NBER, pp. 179–222.

Grossman, G. and Helpman, E.: 1991, Quality ladders in the theory of growth, *Review of Economic Studies* **58**, 43–61.

Helpman, E. and Rangel, A.: 1999, Adjusting to a new technology: Experience and training,, *Journal of Economic Growth* **4**, 359–383.

Javorcik, B. S.: 2004, Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages, *American Economic Review* **94**(4), 605–627.

Kneller, R.: 2005, Frontier technology, absorptive capacity and distance, *Oxford Bulletin of Economics and Statistics* **67**, 1–24.

Kneller, R. and Stevens, P. A.: 2006, Frontier technology and absorptive capacity: Evidence from OECD manufacturing industries, *Oxford Bulletin of Economics and Statistics* **68**, 1–21.

Krueger, A. and Lindahl, M.: 2001, Education for growth: Why and for whom?, *Journal of Economic Literature* **39**, 1101–1136.

Lucas, R. E.: 1988, On the mechanics of economic development, *Journal of Monetary Economics* **22**, 3–42.

Mankiw, G., Romer, D. and Weil, D.: 1992, A contribution to the empirics of economic growth, *Quarterly Journal of Economics* **107**, 407–437.

Mastromarco, C. and Ghosh, S.: 2009, Foreign capital, human capital, and efficiency: A stochastic frontier analysis for developing countries, *World Development* **37**, 489 – 502.

Mastromarco, C. and Simar, L.: 2015, Effect of fdi and time on catching-up: New insights from a conditional nonparametric frontier analysis, *Journal of Applied Econometrics* **30**, 826–847.

Miller, S. M. and Upadhyay, M. P.: 2000, The effects of openness, trade orientation, and human capital on total factor productivity, *Journal of Development Economics* **63**, 399–423.

Miller, S. M. and Upadhyay, M. P.: 2002, Total factor productivity, human capital and outward orientation: Differences by stage of development and geographic regions. Department of Economics Working Paper Number 2002-03, University of Connecticut.

Olofsdotter, K.: 1998, Foreign direct investment, country capabilities and economic growth, *Weltwirtschaftliches Archiv* **134**, 534–547.

Pesaran, M. H.: 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* **74 (4)**, 967–1012.

Pritchett, L.: 2001, Where has all the education gone?, *World Bank Economic Review* pp. 367–391.

Psacharopoulos, G.: 1994, Returns to investment in education: A global update., *World Development* **22**, 1325–1343.

Quah, D. T.: 1997, Empirics for growth and distribution: Stratification, polarization, and convergence clubs, *Journal of Economic Growth* **2**, 27–59.

Rodriguez, F.: 2006, Cleaning up the kitchen sink: On the consequences of the linearity assumption for cross-country growth empirics. Mimeo, Wesleyan University, Middletown, CT.

Romer, P. M.: 1986, Increasing returns and long-run growth, *Journal of Political Economy* **94**, 1002–1037.

Romer, P. M.: 1990a, Endogenous technological change, *Journal of Political Economy* **98**, S71–S102.

Romer, P. M.: 1990b, Human capital and growth: Theory and evidence, *Carnegie-Rochester Conference Series on Public Policy* **32**, 251–286.

Simar, L. and Vanhems, A.: 2012, Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* **166**, 342–354.

Simar, L., Vanhems, A. and van Keilegom, I.: 2015, Unobserved heterogeneity and endogeneity in nonparametric frontier estimation, *Journal of Econometrics* p. forthcoming.

Tallman, E. and Wang, P.: 1994, Human capital and endogenous growth: Evidence from taiwan, *Journal of Monetary Economics* **34**, 101–124.

Temple, J.: 1999, The new growth evidence, *Journal of Economic Literature* **XXXVII**, 112–156.

Topel, R. H.: 1999, Labor markets and economic growth, *in* O. Ashenfelter and D. Card (eds), *The Handbook of Labor Economics*, Elsevier, The Netherlands: North Holland, Amsterdam, pp. 2943–2984.

Vandenbussche, J., Aghion, P. and Meghir, C.: 2006, Growth, distance to frontier and composition of human capital, *Journal of Economic Growth* **11**, 97–127.

Xu, B.: 2000, Multinational enterprises, technology diffusion, and host country productivity growth, *Journal of Development Economics* **62**, 477–493.

Country	Mean	Std. Dev.	Change (%)
USA	0.969	0.021	0.001
NZL	0.953	0.028	0.001
AUS	0.890	0.020	-0.001
BOL	0.887	0.087	0.011
KOR	0.861	0.065	0.004
IRL	0.827	0.023	-0.001
PHL	0.774	0.084	0.009
ISR	0.765	0.041	-0.003
CAN	0.739	0.066	0.005
ZWE	0.703	0.227	0.015
NOR	0.694	0.171	0.018
ARG	0.690	0.077	-0.001
ZMB	0.662	0.188	0.039
KEN	0.661	0.203	0.049
DNK	0.629	0.067	-0.004
NLD	0.598	0.075	0.005
BEL	0.592	0.039	-0.008
SWE	0.582	0.110	0.024
JAM	0.558	0.258	0.034
CHL	0.537	0.125	-0.009
JPN	0.511	0.105	0.007
DOM	0.456	0.088	0.009
FIN	0.429	0.110	-0.014
TUR	0.396	0.102	-0.014
CIV	0.341	0.240	0.073
AUT	0.319	0.124	-0.024
MWI	0.315	0.169	0.056
HND	0.312	0.089	0.009
MEX	0.309	0.120	0.038
GER	0.297	0.287	0.103
HKG	0.290	0.078	0.005
GBR	0.285	0.125	-0.032
GRC	0.273	0.081	-0.001
THA	0.232	0.068	0.017
VEN	0.188	0.129	-0.001
FRA	0.141	0.116	0.126
PRT	0.111	0.045	-0.057
ITA	0.103	0.043	-0.013
ESP	0.084	0.112	0.104
MAR	0.050	0.025	0.025

Table 1: *Estimated latent factor $\widehat{V}$ of 40 countries over 1970 till 2007: mean and standard deviation over time and change in % from 1970 to 2007.*

Country	Mean	Std. Dev.	Change (%)
JPN	2.858	0.456	0.015
KOR	1.655	1.296	0.101
NZL	1.478	0.365	0.004
NOR	1.188	0.266	-0.035
CAN	1.092	0.136	0.000
AUS	1.086	0.124	0.001
USA	0.905	0.387	0.039
HKG	0.900	0.761	0.067
ISR	0.892	0.150	0.004
FIN	0.733	0.327	-0.032
GER	0.635	0.091	0.005
DNK	0.600	0.411	-0.043
IRL	0.555	0.399	-0.066
SWE	0.541	0.210	-0.047
GBR	0.508	0.090	-0.023
AUT	0.389	0.151	-0.035
FRA	0.303	0.059	-0.021
NLD	0.213	0.085	-0.036
ITA	0.174	0.034	-0.020
BEL	0.153	0.130	-0.069
ARG	0.131	0.028	-0.013
ESP	0.122	0.083	-0.045
CHL	0.121	0.059	0.016
PRT	0.110	0.115	-0.056
GRC	0.110	0.116	-0.057
MEX	0.097	0.034	0.010
VEN	0.096	0.027	-0.056
THA	0.052	0.033	0.068
JAM	0.036	0.013	-0.018
PHL	0.033	0.007	-0.010
TUR	0.026	0.014	0.032
DOM	0.025	0.006	0.062
ZWE	0.025	0.009	-0.080
MAR	0.018	0.008	-0.005
HND	0.016	0.015	-0.128
BOL	0.016	0.006	-0.047
ZMB	0.007	0.006	-0.113
MWI	0.005	0.005	-0.140
KEN	0.004	0.002	-0.010
CIV			

Table 2: *Number of Patents for 1000 habitants of 40 countries over 1980 till 2011: mean and standard deviation over time and change in % from 1980 to 2011. Data Source: Source: WIPO statistics database. Last updated: March 2015. Own calculations.*

Country	Mean	Std. Dev.	Change (%)
FIN	7189.900	781.960	0.022
ISR	6602.300		
SWE	5229.100	535.300	0.013
JPN	5170.800	134.190	0.002
DNK	5158.400	1326.100	0.048
NOR	4869.000	568.360	0.026
CAN	3932.900	586.550	0.025
AUS	3755.300	404.880	0.036
AUT	3712.300	689.770	0.059
USA	3637.000	270.850	0.015
KOR	3609.800	1289.400	0.059
GBR	3540.200	651.480	0.028
GER	3354.400	431.490	0.024
FRA	3223.700	437.020	0.024
BEL	3186.400	432.360	0.028
NZL	3028.200	595.580	0.063
NLD	2897.900	352.160	0.023
IRL	2659.900	546.880	0.040
PRT	2559.600	1270.200	0.077
ESP	2254.900	575.330	0.042
HKG	2053.000	725.960	0.076
GRC	1660.300	412.920	0.080
ITA	1402.500	242.950	0.016
ARG	863.910	194.280	0.038
MAR	715.410	89.933	0.053
TUR	543.340	228.760	0.071
CHL	322.550	28.689	-0.020
MEX	297.660	73.264	0.035
THA	234.230	104.130	0.103
KEN	141.610	121.400	0.551
VEN	134.930	72.127	0.115
BOL	108.310	37.229	0.101
ZWE	95.101		
PHL	76.524	4.946	0.031
CIV	72.956		
ZMB	50.958	5.023	-0.033
MWI	39.173	13.561	0.225
DOM			
HND			
JAM			

Table 3: *Researchers in R&D (per million people) of 40 countries over 1996 till 2012: mean and standard deviation over time and change in % from 1980 to 2011. Data Source: World Development Indicators (2015). Own calculations.*

	eps_L	eps_K	eps_H	eps_Y
	Pearson correlations			
t	-0.0115	-0.0522	-0.0470	-0.0521
$L_{\cdot t}$	-0.0279	0.0142	-0.0026	-0.0063
$K_{\cdot t}$	-0.0369	0.0212	-0.0171	0.0089
$H_{\cdot t}$	-0.0096	0.0086	-0.0067	0.0112
$Y_{\cdot t}$	-0.0353	0.0187	-0.0142	0.0104
FDI	0.0081	-0.0052	-0.0251	0.0066
	Spearman rank correlations			
t	-0.0127	-0.0690	-0.0476	-0.0533
$L_{\cdot t}$	-0.0278	0.0229	0.0070	-0.0095
$K_{\cdot t}$	-0.0454	0.0308	-0.0193	0.0117
$H_{\cdot t}$	-0.0132	0.0119	-0.0121	0.0131
$Y_{\cdot t}$	-0.0379	0.0302	-0.0154	0.0140
FDI	0.0273	-0.0168	-0.0229	0.0223
	Kendall correlations			
t	-0.0082	-0.0452	-0.0320	-0.0348
$L_{\cdot t}$	-0.0192	0.0151	0.0049	-0.0064
$K_{\cdot t}$	-0.0316	0.0204	-0.0126	0.0081
$H_{\cdot t}$	-0.0090	0.0072	-0.0080	0.0090
$Y_{\cdot t}$	-0.0266	0.0202	-0.0101	0.0096
FDI	0.0185	-0.0104	-0.0156	0.0146

Table 4: *Correlation between eps_L, eps_K, eps_H, eps_Y and the factor $F_t = t, L_{\cdot t}, K_{\cdot t}, H_{\cdot t}, Y_{\cdot t}$ and FDI*

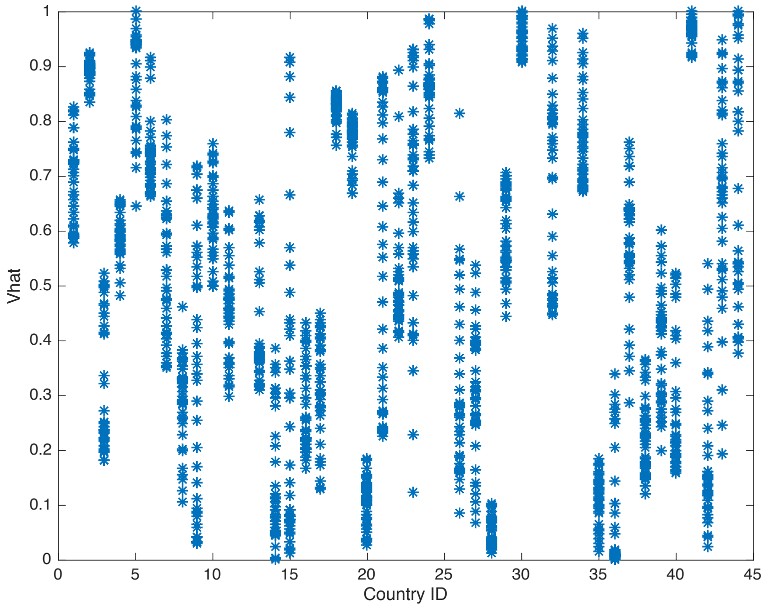

Figure 1: *Distribution of latent factor* $\widehat{V}$ *for the countries under analysis.*

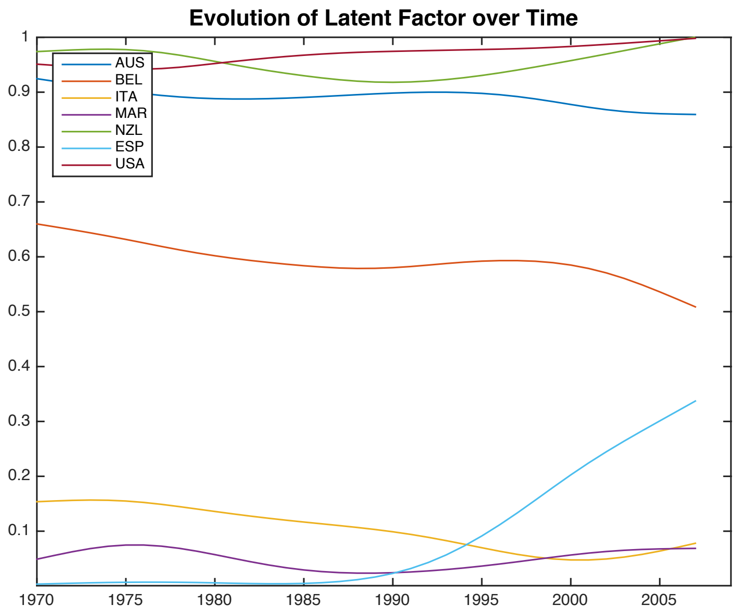

Figure 2: *Evolution over time of latent factor $\widehat{V}$ for Australia (AUS), Belgium (BEL), Italy (ITA), Morocco (MAR), New Zealand (NZL), Spain (ESP) and United States (USA).*

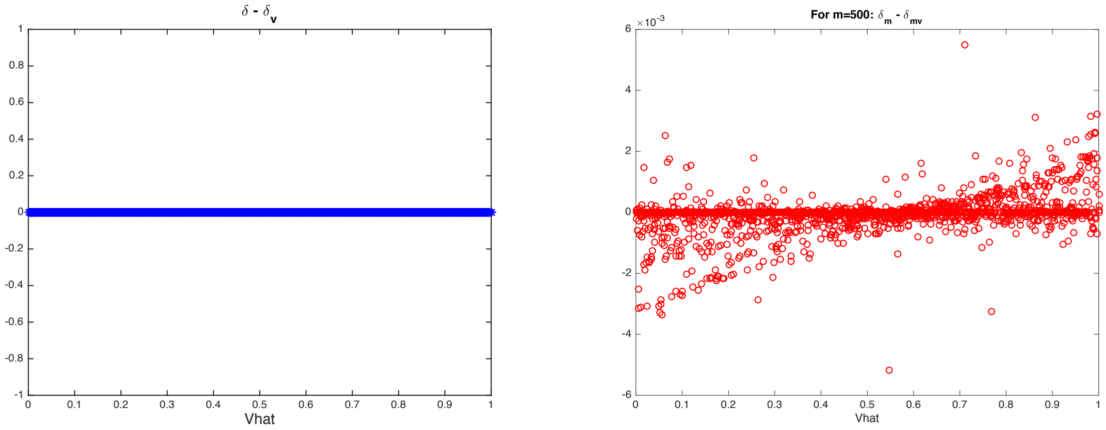

Figure 3: *Estimated differences of marginal and conditional efficiency of full frontier (left panel) and order-m frontier (right panel). m=500 and the sample size n=1520*

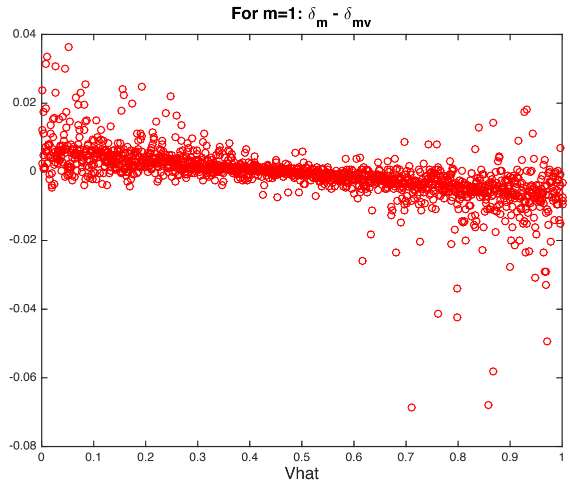

Figure 4: *Estimated differences of marginal and conditional efficiency of order-$m = 1$ frontier*