I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/15

Effect of FDI and Time on Catching-up: New Insights
from a Conditional Nonparametric Frontier Analysis

MASTROMARCO, C. and L. SIMAR

Effect of FDI and Time on Catching-up: New Insights from a Conditional Nonparametric Frontier Analysis

Camilla Mastromarco* Léopold Simar†

March 12, 2013

Abstract: Productivity catching-up among countries has been an important topic of interest for many researchers. In particular, one is willing to see how the productivity of countries evolves over time and how exogenous factors may influence the process. Many studies have analyzed the effect of Foreign Direct Investment (FDI) on catching-up over time, many by using restrictive parametric models. We use an appropriate nonparametric two-step approach on conditional efficiencies to investigate how FDI and time affect the process. By using a dataset of 44 countries over 1970-2007, we explore the channels under which FDI fosters productivity by disentangling the impact of this factor on 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 - conditional nonparametric efficiency measures - to examine these interrelationships. We emphasize the usefulness of smoothing over time to better analyze the potential dynamic influence of FDI on efficiency. We find that both FDI and time play an important role as influencing efficiency distribution and affecting, to a smaller extent, the production set. This effect of FDI does not seem to vary much over time. By the second stage nonparametric regression of the conditional efficiencies over FDI and time we identify clearly effect of time and FDI on conditional efficiency and we determine idiosyncratic efficiency, which represents "Solow residual", measured by looking to the unexplained part of the conditional efficiencies. By doing so, we are able to rank the productivity of countries at each time period, or on the average, when facing different environmental conditions.

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

Keywords: Nonparametric Frontier in Heterogeneous Panels, Time-Varying Conditional Efficiency Measures and Environmental Factors.

*Dipartimento di Scienze Economiche e Matematico-Statistiche, Università degli Studi del Salento. Centro ECOTEKNE, Via per Monteroni - 73100 LECCE - Italy. Ph. +39832298779, fax +390832298757, 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

Productivity catching-up (efficiency) among countries has been an important topic of interest for many researchers. In particular, one is willing to see how the productivity of countries evolves over time and how exogenous factors may influence the process.

Total factor productivity has been recognized as the most important driver behind economic growth (Prescott 1998, Caselli 2005, Parente and Prescott 2005). Starting with Färe et al. (1994), many studies use a method based on Malmquist productivity indexes (Caves et al. 1982) and decompose productivity growth into components attributable to technological change, technological catch-up and input accumulation by linking the literature on convergence and the efficient frontier. These studies go beyond the standard growth accounting method and propose a decomposition of output growth based on efficiency frontier, and hence can avoid the caveats in the assumptions made in using the growth accounting approach (Caselli 2005).¹

Growth accounting decomposes increases in output into two parts. One is explained by input changes and the other, calculated as a residual, as technical change. Interpretation of the residual as technical change is reasonable only if all countries are producing on their frontier. The strength of the frontier model is that the residual can be decomposed into technical change and efficiency. Efficiency measures describe the deviation from the best practice technology. Frontier methodology allows the analysis of the factors that affect technical efficiency. Productivity growth is the net change in output due to changes in efficiency and technical change. Therefore, efficiency is a component of productivity. The main advantage of the frontier approach, used here, is that it allows this important distinction.

The DEA method of estimating efficiency has gained popularity because of several advantages over other methods. The main advantage is that, being non parametric, this method does not require any parametric assumption regarding technology or the inefficiency term. Moreover, the assumption of complete homogeneity of considered economic units is not needed. Therefore the economic units under investigation, can potentially consist of different groups of populations governed by different distributional laws of the generation of input-output mix and on efficiency. This means, that, if the sample is formed by developed and developing countries, as in our case, these groups of countries can have different distributions of efficiency scores.

In this paper, we focus on the role of Foreign Direct Investment (FDI) in spurring technological catching-up among countries. FDI leads to increases in productivity by stimulating competition and transferring technology. New foreign competition arrivals provide domestic

¹Growth accounting approach requires strong assumption on production function form (Cobb-Douglas) and on returns to scale (constant).

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. Using these arguments, Borensztein et al. (1998), De Mello (1999), Xu (2000), Mastromarco and Ghosh (2009) and Iyer et al. (2008) conclude that FDI increases an economy's productive efficiency. Javorcik (2004) argues that FDI can also raise productivity growth through vertical spillovers rather than horizontal spillovers.

In a cross-country framework, production inefficiencies can be identified as the distance of the individual production from the frontier estimated by the maximum output of the reference country regarded as the empirical counterpart of an optimal boundary of the production set. Inefficiencies generally reflect a sluggish adoption of new technologies, and thus efficiency improvement will represent productivity catch-up via technology diffusion. An exact dynamic mechanism as to how technology efficiency and globalization factors such as FDI relate to each other remains ambiguous. Particularly at the center of theoretical debates, lies the issue as to how best to model dynamic adjustments.

Frontier methodology, pioneered by Aigner et al. (1977) and by Meeusen and van den Broeck (1977), allows to estimate economic efficiency of various economic systems (firms, industries, countries, regions, etc.) and, moreover, to analyze the factors that may affect the production process as production factors (shape of the frontier), technological factors (shift of the frontier) or efficiency factors (distance from the frontier) (Kumbhakar and Lovell 2000, Bădin et al. 2012). Frontier models tend to focus on estimating the long-run equilibrium relationship between output and production factors without explicitly modeling dynamic adjustments towards an equilibrium (Kumbhakar and Lovell 2000). In general, the partial adjustment process cannot be appropriately encapsulated by a standard frontier model based on an implicit assumption of complete adjustment because the adjustment from the current to the desired future input use is far from perfect due to time delays, delivery lags and installation costs. This neglect may result in a misleading conclusion that an inter-temporally efficient producer can be classified as being inefficient.

Inclusion of the short-run dynamics is likely to be relevant to a stochastic frontier approach as analysed in Ahn and Sickles (2000), Tsionas (2006). These authors provide a dynamic efficiency model which allows inefficiency levels to be serially correlated. The difference among them is that Ahn and Sickles (2000) use generalized methods of moments (GMM) to estimate their model, whereas Tsionas (2006) applies Bayesian inference procedures based on Markov

chain Monte Carlo (MCMC) techniques. Both these papers propose parametric methods which require strong assumptions on the specification of the production function and distributional assumptions on the error components. If these assumptions are not supported by the data under analysis, the estimates of efficiencies and parameters will be biased.

So far most of the studies analyzing effect of external factors on productivity of countries have been in this stream of parametric modeling, see e.g. Kneller and Stevens (2006), Iyer et al. (2008). Kneller and Stevens (2006) use parametric stochastic frontier analysis (SFA) to study the efficiency effects of human capital and R&D undertaken by an industry on technological efficiency for a sample of OECD industries at the industry level from 1973 - 1981. They find that human capital affects a country's efficiency while R&D has an insignificant effect on efficiency. Iyer et al. (2008) use SFA to study OECD countries from 1982 to 2000 and find that trade and foreign direct investment (FDI) inflows increase efficiency whilst FDI outflows increase inefficiency. The SFA framework is also used by Mastromarco and Ghosh (2009) who examine the efficiency effects of FDI and human capital for developing countries from 1981 to 2000 and find that FDI imported capital goods and imported R&D are important channels for improving efficiency, as is human capital accumulation. Finally, a recent study by Wijeweera et al. (2010) looks at FDI flows for 45 countries from 1997 to 2004 and finds that FDI inflows have a positive impact on economic growth in the presence of a highly skilled labour force.

However, the parametric approach suffers of misspecification problems when the data generating process is unknown, as usual in the applied studies, and the nonparametric methods often give the most reliable results. The purpose of this paper is to provide fully nonparametric estimators of production frontiers and time variant technical efficiency in a dynamic framework which allows external factors to affect technical efficiency.²

Econometric studies on macroeconomic data using nonparametric approaches (like FDH or DEA) is not new. Kumar and Russell (2002) employed standard DEA (data envelopment analysis) to investigate international macroeconomic convergence of 57 industrialized and developing countries. They decompose labour productivity into components attributable to technological change, technological catch-up (efficiency) and physical capital accumulation. They conclude that technological catch-up did not contribute to convergence in income per capita across countries and that capital accumulation was the primary driving force behind growth in income per capita during the period 1965-1990. Henderson and Russell (2005) add new insights to this analysis by including human capital accumulation into the decomposition and show the importance of this factor as driver of economic growth. Following these studies and using same data as Henderson and Russell (2005) and bootstrapping techniques,

²The efficiency frontier literature defines environmental or external factors, those variables which might affect the production process but which are not under the direct control of the production unit.

Henderson and Zelenyuk (2007) find that efficiency scores were significantly different between developed and developing countries.

Nonparametric and parametric frontier studies try to model the effect of environmental factors on production process and applying one stage or two-stage approaches to capture the heterogeneity caused by these factors (e.g. Henderson and Russell 2005, Henderson and Zelenyuk 2007, Iyer et al. 2008, Kneller and Stevens 2006). Unfortunately, these approaches often are flawed by some restrictive and often unverifiable prior assumptions (see Simar and Wilson 2007, 2011).

Following recent developments in nonparametric frontier literature (Cazals et al. 2002, Daraio and Simar 2005, Bădin et al. 2012) we suggest in this paper a more appropriate nonparametric two-step approach to investigate dynamic transmission mechanism through which environmental factors (FDI and time) may affect the catching-up of countries. By using a dataset of 44 countries over 1970-2007, we explore the channels under which FDI fosters productivity by disentangling the impact of this factor on 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 - conditional nonparametric efficiency measures and location-scale regression models - to examine these interrelationships. By the appropriate second stage nonparametric regression of the conditional efficiencies over FDI and time we determine idiosyncratic efficiency measures by looking to the residuals, that can be viewed as the remaining unexplained part of the conditional efficiencies and i.e. the “Solow residual”. The analysis of these residuals allows to rank the different countries from less to most efficient and to investigate its evolution over time.

The paper is organized as follows. Section 2 presents the methodology. Specifically, section 2.1 explains our proposed measure of time dependent conditional efficiency scores and the nonparametric estimation. Section 2.2 describes the ratios of conditional and unconditional efficiency measures as exploratory tools to assess the effect of explanatory variables on production process. Section 2.3 discusses the location scale model. Section 3 illustrates the empirical application and section 4 summarizes the main findings and concludes the paper.

2 The Methodology

2.1 First stage: Time Dependent Conditional Efficiency Scores

Following Cazals et al. (2002), Daraio and Simar (2005) we consider a Data Generating Process (DGP) characterizing the production process in the presence of environmental factors and we extend their models to a dynamic framework to allow the introduction of the time dimension.

Consider a generic input vector $X \in \mathbb{R}_+^p$, a generic output vector $Y \in \mathbb{R}_+^q$ and we will denote by $Z \in \mathbb{R}^d$ the generic vector of environmental variables (FDI in our study). First the unconditional (marginal) attainable set of feasible combinations of inputs and outputs, $\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} | x \text{ can produce } y\}$ can be characterized by $\Psi = \{(x, y) | H_{X,Y}(x, y) > 0\}$, where $H_{X,Y}(x, y) = \text{Prob}(X \leq x, Y \geq y)$. So Ψ is the support of the joint random variable (X, Y) .

Then, with some abuse of notation, we will consider the time T as a conditioning variable and we will define for each time period t , Ψ_t , the attainable set at time t as the set of combinations of inputs and outputs feasible at time t . Now $\Psi_t \subset \mathbb{R}_+^{p+q}$ is the support of (X, Y) at time t , whose distribution is completely determined by

$$H_{X,Y}^t(x, y) = \text{Prob}(X \leq x, Y \geq y | T = t),$$

which is the probability of being dominated for a production plan (x, y) at time t . Finally, when considering the presence of additional external factors Z , the attainable set is defined as $\Psi_t^z \subseteq \Psi_t \subset \mathbb{R}_+^{p+q}$ defined as the support of the conditional probability

$$H_{X,Y|Z}^t(x, y|z) = \text{Prob}(X \leq x, Y \geq y | Z = z, T = t).$$

For efficiency measures, we only focus the presentation on the output orientation and on DEA version of the estimators (Daraio and Simar 2007b), the same could be done for any other orientation (input, hyperbolic, directional distance) and for FDH version of the estimators (Daraio and Simar 2007a, Bădin et al. 2010, 2012, van Keilegom and Wilson 2012, Simar and Vanhems 2012, Simar, Vanhems and Wilson 2012).

In this framework, assuming free disposability of inputs and outputs, the unconditional (marginal) output oriented Farrell-Debreu technical efficiency of a production plan (x, y) , is defined as

$$\lambda(x, y) = \sup\{\lambda | (x, \lambda y) \in \Psi\} = \sup\{\lambda | H_{X,Y}(x, \lambda y) > 0\}. \quad (1)$$

For the output orientation, it will be useful to use the decomposition $H_{X,Y}(x, y) = S_{Y|X}(y|x)F_X(x)$ where $S_{Y|X}(y|x) = \text{Prob}(Y \geq y | X \leq x)$ is a nonstandard conditional survival function (non-standard because we condition on $X \leq x$) and $F_X(x)$ is the usual marginal cdf of X . So, for any x in the support of X , we have equivalently

$$\lambda(x, y) = \sup\{\lambda | S_{Y|X}(\lambda y|x) > 0\}. \quad (2)$$

Accordingly, the conditional output oriented technical efficiency of a production plan $(x, y) \in$

Ψ_t^z , at time t facing conditions z , is defined in (Daraio and Simar 2005) as

$$\begin{aligned}\lambda_t(x, y|z) &= \sup \{ \lambda | (x, \lambda y) \in \Psi_t^z \} = \sup \{ \lambda | H_{Y,X|Z}^t(x, \lambda y|z) > 0 \} \\ &= \sup \{ \lambda | S_{Y|X,Z}^t(\lambda y|x, z) > 0 \}.\end{aligned}\tag{3}$$

where $S_{Y|X,Z}^t(y|x, z) = \text{Prob}(Y \geq y | X \leq x, Z = z, T = t)$ (note the difference in the conditioning for X , the inputs, and for Z , the environmental factor).

Nonparametric Estimation of Efficiencies

Suppose we have a panel of data $(x_{i,t}, y_{i,t}, z_{i,t})$ for $i = 1, \dots, n$ and $t = 1, \dots, s$, the unconditional and conditional attainable sets can be estimated. Assuming that the true attainable sets are convex and under free disposability of inputs and outputs, the DEA estimators can be written as (Daraio and Simar 2007b)

$$\begin{aligned}\hat{\Psi}_{DEA} = \{ (x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+^q | y \leq \sum_{j=(i,t)} \gamma_j y_j; x \geq \sum_{j=(i,t)} \gamma_j x_j; \\ \gamma \geq 0 \text{ s.t. } \sum_{j=(i,t)} \gamma_j = 1 \},\end{aligned}\tag{4}$$

where the summations are over all the combinations $j = (i, t)$ with $i = 1, \dots, n$ and $t = 1, \dots, s$. Similarly, at time t and facing the condition $Z = z$ we have

$$\begin{aligned}\hat{\Psi}_{t,DEA}^z = \{ (x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+^q | y \leq \sum_{j \in \mathcal{J}(z,t)} \gamma_j y_j; x \geq \sum_{j \in \mathcal{J}(z,t)} \gamma_j x_j; \\ \gamma \geq 0 \text{ s.t. } \sum_{j \in \mathcal{J}(z,t)} \gamma_j = 1 \}\end{aligned}\tag{5}$$

where $\mathcal{J}(z, t) = \{ j = (i, v) | z - h_z < z_{i,v} < z + h_z; t - h_t < v < t + h_t \}$; h_z and h_t are bandwidths of appropriate size selected by data-driven methods. These $\mathcal{J}(z, t)$ describe the localizing procedure to estimate the conditional DEA estimates, they determine the data points in a neighborhood of (z, t) that will be used to compute the local DEA estimate.

Daraio and Simar (2005, 2007a) and Bădin et al. (2010) discuss in detail how to choose the appropriate bandwidths. They are determined by the estimation of conditional distributions $S_{Y|X,Z}^t(y|x, z)$, where we condition on $X \leq x$, time $T = t$ and a particular value of $Z = z$ and here standard tools from Hall, Racine and Li (2004), Li and Racine (2008) can be adapted. Of course here only the variables (t, z) require smoothing and appropriate bandwidths, since

we have

$$\hat{S}_{Y|X,Z}^t(y|x, z) = \frac{\sum_{j=(i,v)} \mathbb{I}(x_j \leq x, y_j \geq y) K_{h_z}(z_j - z) K_{h_t}(v - t)}{\sum_{j=(i,v)} \mathbb{I}(x_j \leq x) K_{h_z}(z_j - z) K_{h_t}(v - t)},$$

where the function are kernels with compact support (see Bădin et al. (2010) for technical details). Optimal bandwidths can be selected by Least Squares Cross-Validation (LSCV) or by Maximum Likelihood Cross-Validation, which are asymptotically equivalent, see e.g. Li and Racine (2007). Note also that the variable T is discrete and discrete kernels could be used for time, but as pointed in Li and Racine (2007), usual continuous kernels are more appropriate when the discrete variables can take many different values, as is the case in our empirical analysis. Note that, as pointed in Hall et al. (2004), by using this data-driven method, in the case where one of the conditioning variable (T or Z) is irrelevant, the selected bandwidth will take large values, as they should, to eliminate their practical impact by conditioning (they should indeed converge to ∞ as the sample size increases).

The estimators of the output efficiency score is then obtained by plugging $\hat{\Psi}_{DEA}$ for Ψ in (1) and $\hat{\Psi}_{t,DEA}^z$ for Ψ_t^z in (3).³ In practice, they are computed by solving the linear programs

$$\begin{aligned} \hat{\lambda}^{DEA}(x, y) = \sup\{ & \lambda | \lambda y \leq \sum_{j=(i,t)} \gamma_j y_j; x \geq \sum_{j=(i,t)} \gamma_j x_j, \\ & \gamma \geq 0 \text{ s.t. } \sum_{j=(i,t)} \gamma_j = 1 \} \end{aligned} \quad (6)$$

For the conditional efficiency measure at time t , we have

$$\begin{aligned} \hat{\lambda}_t^{DEA}(x, y|z) = \sup\{ & \lambda | \lambda y \leq \sum_{j \in \mathcal{J}(z,t)} \gamma_j y_j; x \geq \sum_{j \in \mathcal{J}(z,t)} \gamma_j x_j, \\ & \gamma \geq 0 \text{ s.t. } \sum_{j \in \mathcal{J}(z,t)} \gamma_j = 1 \} \end{aligned} \quad (7)$$

The statistical properties of these nonparametric estimators are well known, see Park et al. (2000) for the FDH efficiency scores, Kneip et al. (2008) for the DEA scores and Jeong et al. (2010) for their conditional versions. To summarize, these estimators are consistent and converge to some complicated limiting distributions to be approximated by bootstrap techniques. The rates of convergence are corrupted by the number of inputs and outputs, illustrating the curse of dimensionality of most of nonparametric methods (see Simar and Wilson 2013, for a recent survey).

³The FDH estimators, that do not assume convexity of the attainable sets, are obtained by replacing the distribution H_{XY} and $H_{XY|Z}$ by usual nonparametric estimators in the definitions (1) and (3) respectively.

2.2 Effects on the Production Process: Exploratory Tools

As explained and described in Bădin et al. (2012) we can disentangle the potential effects of conditioning variables to identify effects on the boundary (shift of the frontier) and effects on the distribution of the inefficiencies. The first effect can be investigated by considering the ratios of conditional to unconditional efficiency measures, which are measures relative to the full frontier of respectively, the conditional and the unconditional attainable sets.

$$R_O(x, y|z, t) = \frac{\lambda_t(x, y|z)}{\lambda(x, y)}. \quad (8)$$

In our setup, we are mainly interested on the effects of T and Z on these ratios, and not on the potential effects of X . By construction, for the output orientation, $R_O(x, y|z, t) \leq 1$ (the conditional efficient boundary is below the unconditional one) and $R_O(x, y|z, t) = 1$ if and only if, at time t and with conditions z , there is no shift of the efficient boundary of the two attainable sets. Looking to these ratios as a function of t and z allows to investigate the effect of time T and of the environmental factor Z on this potential shift. A global tendency of the ratios to increase with the conditioning variables indicates a favorable 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 from the marginal one when the variables increase, the variables act as undesirable outputs).

The effect of T and of Z on the distribution of the inefficiencies can be done, as suggested in Bădin et al. (2012), by looking rather to order- α quantiles efficiency measures. The full frontier above correspond to an extreme quantile (the maximum achievable output), but we will look here to more central quantiles. We can chose the most central, i.e. the median. The order- α quantile efficiency measures have been introduced in Daouia and Simar (2007). They are defined for any $\alpha \in (0, 1)$ as follows (for the unconditional and conditional cases respectively)

$$\lambda(x, y) = \sup\{\lambda|S_Y(\lambda y|x) > 1 - \alpha\}. \quad (9)$$

$$\lambda_{t,\alpha}(x, y|z) = \sup\{\lambda|S_{Y|X,Z}^t(\lambda y|x, z) > 1 - \alpha\}, \quad (10)$$

so we are not characterizing the full support of Y under the conditioning, but a less extreme quantile (unless $\alpha = 1$). Here for our purpose of analyzing the impact of t and z on the distribution of efficiencies, we are not interested in extreme quantiles (like, say, $\alpha = 0.95$) but

rather to the median, by choosing $\alpha = 0.50$. The ratios to be analyzed are now

$$R_{O,\alpha}(x, y|z, t) = \frac{\lambda_{\alpha,t}(x, y|z)}{\lambda_{\alpha}(x, y)}. \quad (11)$$

As explained in Bădin et al. (2012), the ratios are not bounded by 1, because the order- α efficiency scores are not bounded by 1. The latter equal to 1 if and only if (x, y) is on the α -frontier, bigger than 1, if they are below and smaller than 1 if they are above (roughly 50%, when $\alpha = 0.50$). But, as for the full ratios above, a tendency of $R_{O,\alpha}(x, y|z, t)$ to increase with the conditioning variables indicates a favorable effect of these variables on the distribution of the efficiencies (the conditional distribution is more concentrated to its upper boundary when the conditioning variables increase) and the opposite in the case of a unfavorable effect. If this effect is similar to the one shown with the ratios 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 variables T and Z change; if the effect with the medians 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 (see Figure 10, in Bădin et al. (2012) for a detailed explanation of the different possible scenarios).

In practice we use nonparametric estimators of the efficiency scores and we explore the effect of T and Z by looking to the behavior of $\hat{R}_O(x, y|z, t)$ and $\hat{R}_{O,\alpha}(x, y|z, t)$ as a function of T and Z . Of course here, due to the statistical uncertainty, only exploratory analysis will be available because the estimates are not equal to the true values used for defining the ratios in (8) and (11). In the empirical analysis below, we have chosen the most flexible and robust FDH estimators, that do not impose convexity of the attainable sets. For analyzing boundaries, it is better to avoid to impose assumptions that are not empirically verified.⁴

2.3 Second Stage: Regression of the Conditional Efficiency Scores

By conditioning on time and on Z in the first stage we are able to analyze the behavior of $\lambda_t(x, y|z)$ over both time and Z , taking into account the dependence in the DGP on both variables (T, Z) and therefore improving the precision of our estimates. We have in fact described three ways for describing the attainable sets: Ψ , Ψ_t for each t and Ψ_t^z for each value of the pair (z, t) . The general relation between these sets have been described in Simar and Wilson (2007, 2011), we have $\Psi = \bigcup_t \Psi_t = \bigcup_{t,z} \Psi_t^z$.

Then several scenarii can happen. We can have time independence of the attainable sets

⁴So the analysis described here allows to identify potential effects and this, in a full nonparametric setup; but more refined analysis should then be conducted to quantify these effects more precisely. So far, the latter requires additional parametric assumptions but additional research is needed for nonparametric extensions.

(the separability condition with respect to time) which corresponds to the case where for all t , $\Psi_t \equiv \Psi$. If we have full independence for both T and Z (full separability condition), then $\Psi_t^z \equiv \Psi$ for all (z, t) , i.e. neither time nor Z influence the shape (the boundary) of the attainable set. In fact we expect often that the separability condition does not hold, for instance, in case of technical progress (or regress), the frontier of the attainable set will change when t changes.

As explained in Simar and Wilson (2007, 2011), when the separability condition is not verified, the marginal measures of efficiencies, computed relatively to boundary of the marginal attainable set Ψ , i.e. $\lambda(x, y)$, have no particular economic meaning because each observed production plan will be benchmarked against an unattainable frontier at any particular time t and facing conditions z . So, a regression of $\lambda(X, Y)$ on (T, Z) , as suggested in traditional two-stage approaches, could provide meaningless results. This is the reason why, as suggested by Bădin et al. (2012), it is more meaningful to regress, in the second stage, the conditional efficiency scores $\lambda_t(x, y|z)$ on the relevant explanatory variables (t, z) . In this way we can capture in the efficiency scores, the shift of the frontier, technological changes, which may cause moves of the frontier over time or with the values of Z .

Therefore, in practice, in the second stage, we will analyze the average behavior of $\lambda_t(X, Y|z)$ for the time period t and conditionally on $Z = z$. The regression aims to analyze the behavior of $\mathbb{E}[\lambda_t(X, Y|Z = z)]$ as function of z and t . We use, as in Bădin et al. (2012), the flexible location-scale regression model:

$$\lambda_t(X, Y|z) = \mu(z, t) + \sigma(z, t)\epsilon \quad (12)$$

where we assume that $\mathbb{E}[\epsilon|z, t] = 0$ and $\mathbb{V}[\epsilon|z, t] = 1$. This model allows us to capture the location $\mu(z, t) = \mathbb{E}[\lambda_t(X, Y|Z = z)]$ and the scale effect $\sigma^2(z, t) = \mathbb{V}[\lambda_t(X, Y|Z = z)]$.

The error term for a given particular production plan (x_t, y_t, z_t) at time t is calculated as

$$\epsilon_{x_t, y_t, z_t, t} = \frac{\lambda_t(x_t, y_t|z_t) - \mu(z_t, t)}{\sigma(z_t, t)}. \quad (13)$$

This represents the unexplained part of the conditional efficiency measures and can be interpreted as the idiosyncratic part of the efficiency, called “pure efficiency” in Bădin et al. (2012), i.e. an efficiency scores cleaned by the external effects (here time and Z). Since we include the time as regressor, that enables us to eliminate the common time factor effect, as economic cycles, in a very flexible and robust way (the location-scale model). Our pure efficiency measure provides a better indicator to assess the economic performance of production units over time and allows the ranking of production units facing different environmental factors at different

time periods.⁵ In our macroeconomic framework the “pure efficiency” represents a measure of Solow residual.

Nonparametric estimation of the location-scale model

When applied to a panel of data $(x_{i,t}, y_{i,t}, z_{i,t})$ for $i = 1, \dots, n$ and $t = 1, \dots, s$, we will replace the unknown $\lambda_t(x_{i,t}, y_{i,t}|z_{i,t})$ by $\hat{\lambda}_t(x_{i,t}, y_{i,t}|z_{i,t})$ (we will use conditional DEA estimates defined above) and the estimation of the location-scale model will be done by nonparametric regression techniques (we will use in our empirical study local constant smoothing techniques). This will provide $\hat{\mu}(z_{i,t}, t)$ and $\hat{\sigma}^2(z_{i,t}, t)$. So the idiosyncratic part of the conditional efficiency will be estimated for $i = 1, \dots, n$ and $t = 1, \dots, s$ by

$$\hat{\epsilon}_{i,t} = \frac{\hat{\lambda}_t(x_{i,t}, y_{i,t}|z_{i,t}) - \hat{\mu}(z_{i,t}, t)}{\hat{\sigma}(z_{i,t}, t)}. \quad (14)$$

Nonparametric location-scale models have been studied by many authors, see e.g. Akritas and van Keilegom (2001). In our empirical case, as recommended in (Bădin et al. 2012), we estimate rather a location scale model for $\log \hat{\lambda}_t(x_{i,t}, y_{i,t}|z_{i,t})$ (the log transform is natural when analyzing ratios). So we obtain $\hat{\mu}(z_{i,t}, t)$ by regressing $\log \hat{\lambda}_t(x_{i,t}, y_{i,t}|z_{i,t})$ on (t, z) and $\hat{\sigma}^2(z_{i,t}, t)$ by regressing the squared residuals of the preceding regression on (t, z) . In both cases, we use nonparametric local constant models (to guarantee positive estimates of both functions) and bandwidths determined by usual LSCV (Pagan and Ullah 1999). So at the end the estimates of the idiosyncratic part of the inefficiencies will be given by $\exp(\hat{\epsilon}_{i,t})$.

3 Empirical Application

The dataset is collected over the period, 1970-2007 (38 years) for a total of 44 countries; 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 18 are developing countries (Argentina, Bolivia, Côte d’Ivoire, Dominican Republic, Ecuador, Honduras, Jamaica, Kenya, Madagascar, Malawi, Morocco, Nigeria, Panama, Philippines, Thailand, Venezuela, Zambia, Zimbabwe).⁶

GDP is measured in million US dollars at the 2005 price and labour measured as total employment in thousands. Capital is measured in millions US dollars at 2005 and constructed

⁵Note that here, these measure are standardized with mean zero and variance one.

⁶The choice of countries depends on data availability. Developed and developing countries are classified following the World Bank (2007) classification. See Appendix for data description.

using the perpetual inventory method (PIM).⁷ All three variables are logged and normalized with respect to the median - to ensure homogeneity assumption in inputs and outputs - before estimation. 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.⁸

First it is interesting to look at the optimal bandwidths which have been evaluated for each data points (due to the individual condition $X \leq x_{(i,t)}$). Table 1 summarizes some basic statistics on these values. We see for instance that the median of the values for h_t is around 3.4, so the evaluation of the conditional survival function (and of the depending conditional efficiency scores) have been selected over a window of almost 7 years around the current date. Since the full sample covers 38 years, we can expect the identification on some effect of time on the production process. The same conclusion applies for FDI, since in the chosen units the mean of FDI is around 2 and the standard deviation around 4, whereas the median bandwidth for h_z is around 1.3. Localizing in (t, z) certainly matters.

From this first evidence we can get insights on the dynamics of technological change (shifts of the frontier) and FDI on the production process. In particular, the time effect of technological change may provide a reasonable indicator of the phase of business cycle. The common technology of the countries under investigation appears to have a dominant cycle of 7 years i.e. the Juglar cycle.⁹ The impacts of FDI on production process requires a time period of 2-3 years to be fully absorbed due to adjustments costs (Tybout 1992, Coe and Helpman 1995).

To assess the influence of time FDI on the production process, we first investigate the ratios of conditional and unconditional efficiency measures for full and partial frontier. Figure 1 exhibits the pictures of all the ratios from a marginal point of view (as a marginal function of Time and as a marginal function of FDI). The 3-dimensional pictures are of course also available but they did not provide any additional insights. In particular, they did not show any visible interaction between the effect of time and FDI. They are not reproduced to save space.

⁷PIM is necessitated by the lack of capital stock data across all the countries. For an individual country, the capital stock is constructed as $K_t = K_{t-1}(1 - \theta) + I_t$, where I_t is investment and θ the rate of depreciation assumed to be 6% (e.g., Hall and Jones, 1999; Iyer *et al.*, 2008). Repair and maintenance are assumed to keep the physical production capabilities of an asset constant during its lifetime. Initial capital stocks are constructed, assuming that capital and output grow at the same rate. Specifically, for country with investment data beginning in 1970, we set the initial stock, $K_{1970} = K_{1970}/(g + \theta)$, where g is the 10-year output growth rate from 1970 to 1980. Estimated capital stock includes both residential and non-residential capital.

⁸Capital is sourced from PWT 6.2, labour from OECD Labour Force Statistics; GDP, trade and FDI from the World Bank World Development Indicators and Unctad. The observation period is selected by the data availability.

⁹The traditional business cycle ranges are 3-5 years (Kitchin cycle), 5-7 years and 7-10 years (Juglar cycle).

The full frontier ratios (two top panels) don't show a clear effect of Time and a slight positive effect of Z . We might have some shift of the frontier when FDI increases, but with lot of heterogeneity for small values of Z . Turning to the center of the distribution ($\alpha = 0.5$, looking to the effect on the median of the distribution of Y given that $X \leq x$), Figure 1 (two bottom panels) displays a clear favourable effect of Time (when time increases, the distribution of Y given $X \leq x$ is more concentrated to its upper boundary) and a confirmation of a favourable effect of FDI which is very similar to the effect detected for the shift of the frontier (full frontier). So, FDI act mainly on the shift of the boundary and less on the distribution of inefficiencies. Hence, from this evidence, FDI appear to weakly affect efficiency (technological catch-up) and play a more important role in accelerating the technological change (shifts in the frontier). This result seems to confirm the theoretical hypothesis that FDI leads to increase in productivity by spurring competition: foreign firms have to invest even more in innovation in order to keep up with their technological advantage (Glass and Saggi 1998). To a lesser extent, FDI can also increase efficiency. 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 as they are trained by the foreign firms (Borensztein et al. 1998, De Mello 1999, Xu 2000).

Given that the separability assumption is not verified for all Z variables we analyze the effect of FDI and Time on conditional efficiency scores. We estimate this effect using a flexible local constant estimator, see Figure 2 where location effect, $\hat{\mu}$, is represented (remember we have chosen the log-scale for $\hat{\lambda}$). We see clearly the positive average effects of both Time and FDI on the conditional efficiency (a decrease of $\log \lambda$ indicates an increase of the efficiencies, the optimum being zero in the log-scale). The effect of FDI seems much more important than the effect of Time, except for small values of FDI. This finding suggests that FDI need to reach a certain level to spur efficiency externalities. The effect of Time is much less important for large values of FDI, corroborating the hypothesis of decreasing adjustment costs and time delays.

The two bottom panels of Figure 2 allows to see the marginal effects. These pictures are points on the surface of the top panel, evaluated at observed pairs (t, z) , seen from the perspective of Time and of FDI respectively. Figure 3 shows scale effect: $\hat{\sigma}$. The picture illustrates the high heteroskedasticity of the conditional efficiency scores for small values of FDI. These results highlight the importance of taking into account the time dynamics in investigating the impact of FDI on efficiency. The effect of FDI on conditional efficiency is positive. This finding together with the previous ones underline the importance of time delay in assessing the influence of FDI on efficiency. Knowledge embodied in FDI requires time to be fully transferred for productivity externalities (Cohen and Levinthal 1989, Tybout 1992,

Coe and Helpman 1995). The positive effect of the time on conditional efficiency reveals the presence of a slowly process of technological catch-up dependent on the openness of the country (captured by the presence of multinationals).

Figure 4 displays Idiosyncratic Efficiencies. Histogram of the “pure” efficiencies $\hat{\epsilon}_{i,t}$ (top panel) looks like a truncated normal, with more concentration below zero (more observation are below, this means in output orientation, are more efficient than the average). The bottom panels displays the residuals $\hat{\epsilon}_{i,t}$ against FDI and Time respectively. We cannot see clear remaining dependence, indicating that the location scale model has cleaned most of the effects of time and FDI on the conditional efficiencies. This is confirmed by looking to the correlation between $\hat{\epsilon}_{i,t}$ and FDI and Time respectively given in Table 2. They are all very small. Clearly the $\hat{\epsilon}_{i,t}$, $i = 1, \dots, n$ and $t = 1, \dots, s$ represent the remaining unexplained part of the conditional efficiencies, that we can interpret here as the idiosyncratic component of the efficiencies. This idiosyncratic component is nothing that the residual, somewhat like the “Solow residual”, that proxies for the aggregate effect of various factors not included, other than technology, production inputs or conditional efficiency. In our macroeconomic contest, this component would be related e.g. to the influence of quality of local and international institutions on each country. To help the interpretation we calculate the Solow residual as $\widehat{solow}_{i,t} = \exp(-(\hat{\epsilon}_{i,t} - \min_{i,t}(\hat{\epsilon}_{i,t})))$ which varies between 0 and 1 (1 being efficient).

Table 3 summarises descriptive statistics for Solow Residual for the 44 countries. We find that the most efficient countries over the sample period are USA, Morocco, Norway and Canada, while the least efficient are Zambia, Malawi and Jamaica. Positive performance observed in Morocco is also documented in other studies (see Koop et al. 2000). From the end of 1990s, Morocco recovers from economic stagnation and, since then, has been on a steady path of economic growth, with sound macroeconomic management and sustained growth in non-agricultural sectors. Recent global economic slowdown has only slightly affected Morocco’s economy without crushing it.¹⁰ To give a visual impression of the change of Solow Residual over time, average idiosyncratic efficiency for each year is displayed in Figure 5 for USA, Norway, Zambia, Malawi, Belgium and Italy.

4 Conclusion

This paper aims to explore dynamic transmission mechanism through which environmental factors may affect production process. By following recent literature in nonparametric production frontier model (Cazals et al. 2002, Daraio and Simar 2005, Bădin et al. 2012) we propose a two-step dynamic approach to analyse the channels through which external factors

¹⁰Country Brief 2012, The World Bank Group, web site <http://www.worldbank.org/afr/>.

may affect the economic performance of production unit. Our approach by clearly distinguishing between the effects of environmental variables on the production technology (shift of the frontier) and on efficiency distribution (movement away/towards the frontier) is able to disentangle the effect of external variables on the production process in a dynamic set-up.

We focus on the effect of FDI on economic performance of 44 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).

In this paper we aim to assess the diffusion dynamics of a technology (frontier) in the countries under analysis with respect to FDI. We then intend to redress an important policy issue of whether the protection-oriented policy will hamper the production efficiency through limiting FDI by explicitly analysing dynamic interactions between efficiency and openness factor FDI.

Our empirical evidence reveals that FDI influence productivity through different channels by increasing efficiency and enhancing technological changes. Our results confirm that knowledge embodied in FDI is transferred for efficiency externalities (movement away/towards the frontier) and for technology externalities (shift of the frontier) (Cohen and Levinthal 1989). Moreover, our results reveal that the knowledge embodied in FDI requires time to be fully transferred for productivity externalities (Cohen and Levinthal 1989, Tybout 1992, Coe and Helpman 1995). Hence, our findings support the studies highlighting that lowering barriers to entry of foreign goods and investments have exerted a significantly positive effects on productivity through efficiency and technology gains, e.g. Borensztein et al. (1998), Cameron et al. (2005).

h_j	$\min(h_j)$	$Q_1(h_j)$	$Q_2(h_j)$	$Q_3(h_j)$	$\max(h_j)$
h_{TIME}	2.420	3.049	3.432	4.218	39.530
h_{FDI}	0.971	1.136	1.293	1.690	175.216

Table 1: *Distributions of the optimal bandwidths h_t for TIME and h_z for FDI.*

	$\text{corr}(\epsilon, \text{Time})$	$\text{corr}(\epsilon, \text{FDI})$
Pearson Correlations	-0.037678	-0.027453
Spearman Rank Correlations	-0.036907	-0.074757
Kendall Correlations	-0.024749	-0.049002

Table 2: *Correlations between residuals $\hat{\epsilon}$ and TIME and FDI.*

Figure 1: The two top panels represent the full ratios $\widehat{R}_O(x,y|z,t)$ as a marginal function of Time and FDI. The two bottom panels are the ratios $\widehat{R}_{O,\alpha}(x,y|z,t)$ for $\alpha = 0.5$, so for the median-order efficiencies.

Figure 2: *Estimated location $\hat{\mu}(z_{i,t}, t)$ as a function of Time and FDI, and their two marginal views. Here we used the $\log \hat{\lambda}_t(x_{i,t}, y_{i,t} | z_{i,t})$ as dependent variable.*

Figure 3: *Estimated scale function $\hat{\sigma}(z_{i,t}, t)$ as a function of Time and FDI.*

Figure 4: *Top panel, the empirical distribution of $\hat{\epsilon}_{i,t}$. In the bottom panels, their plot versus Time and FDI respectively.*

Figure 5: *Time varying idiosyncratic Efficiency for Belgium (BEL), Italy (ITA), Malawi(MWI), Norway (NOR), United States (USA) and Zambia (ZMB). Notice also that we use the Hodrick and Prescott (1996) filter to smooth the time paths with a smoothing weight, 100.*

Table 3: Idiosyncratic efficiency (Solow Residual) of 44 countries over 1970–2007 (average over time)

	Mean	Standard Deviation	Change (%) 1970 to 2007
USA	0.885	0.069	-0.014
MAR	0.725	0.126	-0.152
NOR	0.637	0.155	0.110
CAN	0.626	0.170	-0.137
CIV	0.598	0.176	0.081
HKG	0.582	0.229	-0.036
NLD	0.574	0.184	0.045
GBR	0.505	0.119	-0.058
DOM	0.486	0.138	0.243
BEL	0.465	0.161	-0.209
AUT	0.434	0.135	0.681
MEX	0.421	0.155	-0.124
AUS	0.411	0.137	-0.064
TUR	0.387	0.141	-0.146
ISR	0.355	0.086	0.498
FRA	0.351	0.093	0.812
MDG	0.330	0.264	0.010
ITA	0.329	0.090	0.040
ESP	0.311	0.126	-0.067
ARG	0.304	0.156	0.171
KOR	0.297	0.173	-0.295
IRL	0.295	0.161	-0.520
NGA	0.283	0.328	0.148
DNK	0.281	0.064	-0.052
VEN	0.278	0.220	0.081
ZWE	0.276	0.253	-0.131
GER	0.271	0.055	0.060
NZL	0.255	0.091	0.035
SWE	0.253	0.127	0.307
GRC	0.251	0.118	0.073
JPN	0.231	0.120	-0.193
BOL	0.210	0.188	0.149
PHL	0.208	0.141	0.579
PRT	0.190	0.060	-0.105
KEN	0.184	0.168	-0.159
PAN	0.183	0.133	0.158
FIN	0.132	0.095	0.140
CHL	0.128	0.267	-0.176
ECU	0.124	0.079	0.131
HND	0.100	0.040	0.131
THA	0.058	0.064	0.049
JAM	0.040	0.105	-0.133
MWI	0.022	0.032	0.065
ZMB	0.021	0.083	1.496

Notes: Mean, standard deviation and percentage change of Solow residual see Eq. (14) .

Table 4: Appendix: Descriptive statistics for variables used in estimations

		Y (in 000's)	K (in 000's)	L	FDI
ARG	mean	379180000	1274900000	12850000	1.4167
	sd	871760000	207680000	3034500	1.5449
AUS	mean	415640000	1555200000	8026500	1.7520
	sd	159150000	615850000	1593700	1.5583
AUT	mean	195200000	824870000	3515300	1.1496
	sd	52361000	265050000	388390	1.5485
BEL	mean	236490000	972600000	4178200	8.6938
	sd	58824000	297560000	221000	17.5400
BOL	mean	21411000	44883000	2607600	3.0835
	sd	6067100	6243600	849660	3.5117
CAN	mean	706680000	2262800000	13606000	2.0653
	sd	237660000	951840000	2696000	1.8676
CHL	mean	138330000	525510000	4782900	3.1032
	sd	73909000	111040000	1244700	2.8861
HKG	mean	140720000	417220000	2768800	7.5684
	sd	77491000	280480000	606300	8.0952
CIV	mean	32026000	32955000	4442600	1.3192
	sd	10319000	8073500	1580700	0.8489
DNK	mean	124580000	473630000	2740300	1.4252
	sd	31549000	148190000	170360	1.9889
DOM	mean	39913000	92584000	2446900	2.0851
	sd	20601000	56963000	858870	1.4595
ECU	mean	50745000	183880000	3592100	2.1204
	sd	16244000	53429000	1353400	1.9810
FIN	mean	104550000	533060000	2501700	1.4321
	sd	29492000	126570000	133430	2.1967
FRA	mean	1303700000	5095700000	25130000	1.4042
	sd	319960000	1495000000	1947900	1.3762
GER	mean	1868400000	8672300000	37513000	0.8327
	sd	429980000	1790100000	2520000	1.7548
GRC	mean	181320000	755660000	3926500	0.8026
	sd	46151000	204270000	484490	0.4093
HND	mean	14977000	45482000	1633600	1.8503
	sd	5354000	23795000	622920	2.0105
IRL	mean	73417000	257340000	1438400	3.3962
	sd	41785000	128020000	276090	7.4600
ISR	mean	86562000	322980000	1717700	1.4909
	sd	39126000	157460000	579880	2.0182
ITA	mean	1225300000	5464100000	23253000	0.5493
	sd	297150000	1556100000	1118900	0.4928
JAM	mean	17898000	97325000	1067400	3.4814
	sd	2976900	17557000	145270	3.5170
JPN	mean	2826900000	13295000000	61200000	0.0646
	sd	791430000	6055200000	5736900	0.1080
KEN	mean	43813000	60071000	9930000	0.4251
	sd	16383000	18514000	3813100	0.4496
KOR	mean	512200000	1912000000	17595000	0.5835
	sd	338480000	1721700000	4368300	0.8334
MDG	mean	11515000	11462000	5424300	0.9993
	sd	2987800	481770	1852300	1.8814
MWI	mean	8654600	16887000	3747700	1.0145
	sd	3792300	4243900	1064600	1.0361
MEX	mean	738800000	2263300000	27891000	1.7632
	sd	244250000	920030000	9961000	1.1549
MAR	mean	105520000	194670000	7859900	1.2311
	sd	39233000	100810000	2470100	1.6120
NLD	mean	371190000	1465900000	6604200	3.3973
	sd	104820000	401880000	1345600	4.1371
NZL	mean	65060000	240540000	1616100	2.2355
	sd	17809000	65959000	331750	1.6353
NGA	mean	146890000	208720000	29206000	2.7955
	sd	60131000	58166000	7634400	2.3936
NOR	mean	133620000	616490000	2089300	1.1894
	sd	46672000	177370000	246410	1.0544
PAN	mean	14408000	45357000	925720	4.6708
	sd	6415800	16176000	301210	6.0100
PHL	mean	226270000	505540000	25105000	1.1908
	sd	84344000	205920000	7762000	0.8394
PRT	mean	140680000	549490000	4680900	1.6504
	sd	46829000	252890000	564240	1.5543
ESP	mean	738650000	2879400000	15706000	1.7916
	sd	245360000	1215100000	2104600	1.4367
SWE	mean	195780000	766000000	4365300	2.5289
	sd	44265000	146210000	201110	4.3117
THA	mean	289320000	1074600000	28915000	1.9167
	sd	170210000	762140000	6307900	1.6539
TUR	mean	297300000	713460000	20395000	0.6718
	sd	126450000	416280000	3377700	1.1200
GBR	mean	1236000000	3796300000	27737000	2.4637
	sd	346190000	1108400000	1442400	2.0416
USA	mean	7811600000	23994000000	122110000	0.8514
	sd	2712800000	10196000000	20462000	0.7542
VEN	mean	207080000	768170000	6875300	1.1321
	sd	43705000	87495000	2694600	1.8393
ZMB	mean	11756000	39339000	2853200	3.1635
	sd	3315500	5203200	947300	2.8625
ZWE	mean	38348000	78386000	3719300	0.8757
	sd	11506000	20281000	1055000	1.3953

Notes: Y , K and L denote output, capital (measured in 000's) and labour. FDI are inflows measured as ratios of GDP .

References

- Ahn, S. C. and Sickles, R. C.: 2000, Estimation of long-run inefficiency levels: A dynamic frontier approach, *Econometric Reviews* **19**, 461 – 492.
- Aigner, D., Lovell, C. and Schmidt, P.: 1977, Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* **6**, 21–37.
- Akritis, M.G. and van Keilegom, I.: 2001, Nonparametric estimation of the residual distribution *Scandinavian Journal of Statistics* , **28**, 549–568.
- 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.: 2010, Optimal bandwidth selection for conditional efficiency measures: A data-driven approach, *European Journal of Operational Research* **201**, 633–640.
- 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.
- Cameron, G., Proudman, J. and Redding, S.: 2005, Technological convergence, R&D, trade and productivity growth, *European Economic Review* **49**, 775–807.
- Caselli, F.: 2005, Accounting for cross-country income differences, in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Elsevier Press, North-Holland: Amsterdam, pp. 555–677.
- Caves, D. W., Christensen, L. R. and Diewert, W.: 1982, The economic theory of index numbers and the measurement of input, output, and productivity, *Econometrica* **50**, 1393–1414.
- Cazals, C., Florens, J. and Simar, L.: 2002, Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* **106**, 1–25.
- Coe, D. T. and Helpman, E.: 1995, International R&D spillovers, *European Economic Review* **39**, 859–887.
- Cohen, W. and Levinthal, D.: 1989, Innovation and learning: Two faces of R&D, *Economic Journal* **107**, 139–149.

- Daouia, A. and Simar, L.: 2007, Nonparametric Efficiency Analysis: A Multivariate Conditional Quantile Approach, *Journal of Econometrics* **140**, 375–400.
- 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.: 2007a, *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*, Springer, New York.
- Daraio, C. and Simar, L.: 2007b, Conditional nonparametric frontier models for convex and non convex technologies: A unifying approach, *Journal of Productivity Analysis* **28**, 13–321.
- De Mello, L. R. J.: 1999, Foreign direct investment-led growth: Evidence from time series and panel data, *Oxford Economic Papers* **51**, 133–151.
- Färe, R., Grosskopf, S., Norris, M. and Zhang, Z.: 1994, Productivity growth, technical progress, and efficiency change in industrialized countries, *American Economic Review* **84**, 66–83.
- Glass, A. J. and Saggi, K.: 1998, International technological transfer and technology gap, *Journal of Development Economics* **55**, 369–398.
- Hall, P. and Racine, J.S. and Li, Q.: 2004, Cross-Validation and the Estimation of Conditional Probability Densities, *Journal of the American Statistical Association* **99**, 1015–1026.
- Henderson, D. J. and Russell, R. R.: 2005, Human capital and convergence: A production-frontier approach, *International Economic Review* **46**, 1167–1205.
- Henderson, D. J. and Zelenyuk, V.: 2007, Testing for (efficiency) catching-up, *Southern Economic Journal* **73**, 1003–1019.
- Iyer, K., Rambaldi, A. and Tang, K.: 2008, Efficiency externalities of trade and alternative forms of foreign investment in OECD countries, *Journal of Applied Econometrics* **23**, 749–766.
- 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.
- Jeong, S., Park, B. and Simar, L.: 2010, Nonparametric conditional efficiency measures: asymptotic properties, *Annals of Operational Research* **173**, 105–122.

- van Keilegom, I. and Wilson, P. W.: 2012, *Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift for Lèopold Simar*, Springer.
- Koop, G. and Osiewalski, J. and Steel, M.: 2000, Modelling the Sources of Output Growth in a Panel of Countries, *Journal of Business and Economic Statistics* **3**, 284-299.
- Kneip, A., Simar, L. and Wilson, P.: 2008, Asymptotics and consistent bootstraps for dea estimators in non-parametric frontier models, *Econometric Theory* **24**, 1663-1697.
- 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.
- Kumar, S. and Russell, R.: 2002, Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence, *American Economic Review* **92**, 527–548.
- Kumbhakar, S. and Lovell, C.: 2000, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.
- Li, Q. and Racine, J.: 2007, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Li, Q. and Racine, J.: 2008, Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data, *Journal of Business and Economic Statistics* **26**, 423–434.
- Mastromarco, C. and Ghosh, S.: 2009, Foreign capital, human capital, and efficiency: A stochastic frontier analysis for developing countries, *World Development* **37**, 489 – 502.
- Meeusen, W. and van den Broeck, J.: 1977, Efficiency estimation from cobb-douglas production functions with composed error, *International Economic Review* **18**, 435–444.
- Pagan, A. and Ullah, A.: 1999, *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- Parente, S. L. and Prescott, E. C.: 2005, A unified theory of the evolution of international income levels, in P. Aghion and S. N. Durlauf (eds), *Handbook of Economic Growth*, Elsevier B.V., Netherlands, U.S.A., U.K., pp. 1371–1416.
- Park, B., Simar, L. and Weiner, C.: 2000, The fdh estimator for productivity efficiency scores: Asymptotic properties, *Econometric Theory* **16**, 855-877.

- Prescott, E.: 1998, Business cycle research: Methods and problems. Federal Research Bank of Minneapolis Working Paper no 590.
- Simar, L. and Vanhems, A.: 2012, Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* **166**, 342–354.
- Simar, L. and Vanhems, A. and Wilson, P. W.: 2012, Statistical Inference for DEA Estimators of Directional Distances, *European Journal of Operational Research* **220**, 853–864.
- Simar, L. and Wilson, P. W.: 2007, Estimation and inference in two-stage, semi-parametric models of production processes, *Journal of Econometrics* **136**, 31–64.
- Simar, L. and Wilson, P. W.: 2011, Two-stage dea: Caveat emptor, *Journal of Productivity Analysis* **36**, 205–218.
- Simar, L. and Wilson, P. W.: 2013, *Estimation and inference in nonparametric frontier models: Recent developments and perspectives*, in press *Foundations and Trends in Econometrics*.
- Tsionas, E. G.: 2006, Inference in dynamic stochastic frontier models, *Journal of Applied Econometrics* **21**, 669–676.
- Tybout, J. R.: 1992, Linking trade and productivity: new research direction, *World Bank Economic Review* **6**, 189–211.
- Wijeweera, A., Villano, R. and Dollery, B.: 2010, Economic growth and fdi inflows: A stochastic frontier analysis, *Journal of Developing Areas* **43**, 143–158.
- World Bank: 2004, Morocco - Recent Economic Development in Infrastructure (REDI) : water supply and sanitation sector, *World Bank Policy Research Report*.
- Xu, B.: 2000, Multinational enterprises, technology diffusion, and host country productivity growth, *Journal of Development Economics* **62**, 477–493.