Global Dependence and Productivity Catching-up: A Conditional Nonparametric Frontier Analysis

Camilla Mastromarco
University of Salento, CESifo and RCEA
Léopold Simar
Institut de Statistique, Université Catholique de Louvain.

The 20th International Panel Data Conference
Hitotsubashi University, Tokio, Japan 9 - 10 July 2014

Outline

1. Motivation
2. Methodology & Our contribution
3. Estimation Strategy
4. Data
5. Preliminary results
6. Conclusions

Motivation

$\rightarrow$ Globalization and and interconnection among countries.

$\rightarrow$ **Effects of external common factors on productivity of countries:**

- history, geography and trade relations;
- technological interdependency generated by externalities $\rightarrow$ important in explaining conditional convergence process across countries.

$\rightarrow$ **Importance of time and cross section dependence**

Motivation contd.

- factor base SFA:
  - Ahn et al. (2007): factor-based production function for fixed T
  - Kneip et al. (2012): flexible factor-based production function specification in which time-varying individual effects are represented by linear combinations of a small number of unknown basis functions with heterogeneous coefficients.
    $\rightarrow$ similar to Bai (2009) for fixed T:
    $\rightarrow$ Bai approach: i) valid for the large N and T - under the more general case where regressors are correlated with both factors and loading -; ii) if T is large relative to N, the least squares estimator is consistent even in presence of cross-sectional correlation and heteroskedasticity of unknown form.

Motivation contd.

- Inclusion of time and cross-section dependence Pesaran (2006) (PCCE), Bai (2009) (PC), Kapetanios et al. (2014) in SFA: Mastromarco, Serlenga and Shin (JPA 2013)
- Strong assumptions on the specification of the production function and distributional assumptions on the error components.

$\rightarrow$ Parametric approach: suffers of missspecification problems when the data generating process is unknown, as usual in the applied studies.
$\rightarrow$ Nonparametric methods: often give the most reliable results.
$\rightarrow$ The purpose of this paper: robustness method which simultaneously addresses the problem of model specification uncertainty, potential endogeneity and spatial dependence in the analysis of productivity. It also accounts for heteroskedasticity.

Questions

- In the context of production function - **explicitly taking into account CSD** - we seek the identification of the role of Foreign Direct Investment (FDI) in spurring technological catching-up (efficiency) among countries.
- Corss-country framework:
  - Inefficiencies: sluggish adoption of new technologies $\rightarrow$ efficiency improvement: productivity catch-up via technology diffusion;
  - $\rightarrow$ how best to model dynamic adjustments as to how technology efficiency and globalization factors such as FDI relate to each other remains ambiguous
  - Openness to trade, foreign direct investment and imported capital represent important channels for the transmission of new technologies (Borensztein et al. 1998)
    - FDI through the diffusion of more advanced technology and management practices: Findlay (1978);
    - technology progress diffuses through the introduction of new varieties of capital goods at lower costs: Borensztein et al. (1998), Romer (1990), Grossman and Helpman (1991), Barro and Sala-i-Martin (1995).

  $\rightarrow$ importance of global factors on catching-up approach

$\rightarrow$ importance of CSD for the catch-up factor

Methodology

$\rightarrow$ **Two-stage approaches** $\rightarrow$ **Nonparametric Conditional Efficiency Frontier - Panel Data**

- Aim: analyzing the role of technology diffusion in spurring productivity growth via the analysis of the impact over time of environmental factors (FDI) on efficiency.
- FDI may influence the production process in two way:
  - May affect the attainable set: $\Psi^z = \{(x, y) \mid Z = z,\ x \textit{ can produce } y\}$,
  - May affect the efficiency distribution

$\rightarrow$ **Cazals et al. (2002), Daraio and Simar (2005), Bădin et al. (2012), Mastromarco and Simar (2014)**

Methodology contd.

$\rightarrow$ dynamic transmission mechanism through which environmental factors may affect production process (time delays, delivery lags and installation costs);

$\rightarrow$ important issue in time-series;

- Methodology: Nonparametric dynamic two-step approach to estimate SF with common factors $\rightarrow$ inefficiency model to study the evolution of efficiency and its determinants; $\rightarrow$ two-stage approach to capture heterogeneity caused by these factors taking into account partial adjustment process due to time delays.
  $\rightarrow$ influence of common and environmetal factors on efficiency over time.
- Literature of frontier model with environmental factor :
  Simar and Wilson (2007, 2011): problem with two stage approach
  $\rightarrow$ separability issue: unit facing different environmental conditions
  $\rightarrow$ Cazals et al. (2002), Daraio and Simar (2005), Bădin et al. (2012)

$\rightarrow$ **Our approach: nonparametric efficiency frontier panel data model with common and environmental factors**

Nonparametric Time Conditional Efficiency Frontier, contd.

$\rightarrow$ Conditional Frontier and Efficiencies (Cazals et al. 2002, Daraio and Simar 2005):

1. given generic input vector $X \in \mathbb{R}_+^p$, generic output vector $Y \in \mathbb{R}_+$, and $Z \in \mathbb{R}^d$ the generic vector of environmental variables (FDI in our study), at each time period t, define the attainable set $\Psi_t^z \subset \mathbb{R}_+^{p+1}$ as the support of the conditional probability (Cazals et al. 2002):

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

2. Conditional Farrell output measure of efficiency of a production plan $(x, y) \in \Psi_t^z$, at time t facing conditions z:

$$\lambda_t\left(x, y|z\right) = \sup\left\{\lambda | S_{Y|X,Z}^t\left(\lambda y|x, z\right) > 0\right\} = \sup\left\{\lambda | H_{Y,X|Z}^t\left(x, \lambda y|z\right) > 0\right\}$$

3. Suppose we have a panel of data $(x_{i,t}, y_{i,t}, z_{i,t})$ for the cross section unit $i = 1, \ldots, n$ and at time $t = 1, \ldots, s$, the unconditional and conditional attainable sets can be estimated and nonparametric estimator of the conditional survival function $S_{Y|X,Z}^t(y|x, z)$ and of $\lambda_t\left(x, y|z\right)$ could be obtained (as e.g. in Badin et al. , 2010).

Our approach

$\rightarrow$ **Nonparametric Conditional Efficiency Frontier: following Florens, Simar and van Keilegom (JoE2014):**

-
  - $(x_{i,t}, y_{i,t}, z_{i,t})$ for the cross section unit $i = 1, \ldots, n$ and at time $t = 1, \ldots, s$
  - $\mathbf{f}_t = (t, x_{.t}, y_{.t})$: time specific factor which is supposed to capture an exogenous technological change (e.g. Ahn and Sickles (2000)) and is expected to provide a good proxy for any remaining nonlinear and complex trending patterns associated with globalization and the business-cycles;
- our method aims to envelope the effect of time (CSD) ($\mathbf{f_t}$) and Z (FDI) on the production process.

Our approach, contd.

→ **Florens et al. (2014): clear I/O variables by flexible Location-Scale Models:**

1. estimation of the location functions $\mu_\ell(z_{it}, \mathbf{f}_t)$;
2. estimation of the variance functions $\sigma^2_\ell(z_{it}, \mathbf{f}_t)$ by regressing the resulting square residuals of the first step on $(z_{it}, \mathbf{f}_t)$. For the first step we use local linear and for the second step local constant to avoid negative values of the estimated variances.

Our approach, contd.

$\rightarrow$ Florens, Simar and van Keilegom (JoE2014):

- From this first analysis we obtain the residuals

$$\begin{aligned}
\widehat{\varepsilon}_{1,it}^{(j)} &= \frac{X_{it}^{(j)} - \mu_1^{(j)}(z_{it}, \mathbf{f}_t)}{\widehat{\sigma}_1^{(j)}(z_{it}, \mathbf{f}_t)}, \quad \text{for } j = 1, \ldots, p \\
\widehat{\varepsilon}_{2,it} &= \frac{Y_{it} - \mu_2(z_{it}, \mathbf{f}_t)}{\widehat{\sigma}_2(z_{it}, \mathbf{f}_t)}
\end{aligned}$$

- $\widehat{\varepsilon}_{1,it}^{(j)}$ and $\widehat{\varepsilon}_{2,it} \perp (z_{it}, \mathbf{f}_t) \rightarrow$ can be tested see Florens et al. (2014)
- These are the whitened inputs and output obtained by eliminating the influence of the external and other environmental variables as common factors.
- our approach implies that production technology is driven by exogenously given common factors (Pesaran 2006, Bai 2009) and international diffusion of technology through FDI

Our approach, contd.

→ **The first step 'cleaning' allows us:**

- to control for endogeneity due to reverse causation between production factors, output and external variables;
- to accommodate for CSD;
- to clean external factors dependence avoiding the problem of curse of dimensionality due to the dimension of the external variables.

Our approach: 2nd step

$\rightarrow$ **Efficiency Frontier in $\widehat{\varepsilon}_{1,it}^{(j)}$ and $\widehat{\varepsilon}_{2,it}$**

- attainable set of pure inputs and output $(\varepsilon_1, \varepsilon_2)$:
  - $\Psi_\varepsilon = \left\{(e_1, e_2) \in \mathbb{R}^{p+1} | H_{\varepsilon_1,\varepsilon_2}(e_1, e_2) = \mathrm{Prob}(\varepsilon_1 \leq e_1, \varepsilon_2 \geq e_2) > 0\right\}$
- statistical properties (Florens et al. 2014):
  - consistency for the full-frontier FDH estimator and $\sqrt{n}$-consistency and asymptotic normality for the robust order-m frontiers
  - 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. $n^{1/(p+1)}$

Our approach: 2nd step,contd.

$\rightarrow$ **'"pure" measure of efficiency:**

- direction distance, output orientation (Simar and Vanhems 2012) $d_1 = 0$ and we can chose $d_2 = 1$ (Daraio and Simar, 2014) :

$$\widehat{\beta}(\hat{e}_1, \hat{e}_2; d_1, d_2) = \sup\{\beta | (\hat{e}_1, \hat{e}_2 + \hat{\beta} d_2) \in \widehat{\Psi}_{FDH,\varepsilon}\},$$

- The conditional output-oriented frontier in the original units of the inputs and output is directly obtained at any value of $(y_{it}, x_{it}, z_{it}, \mathbf{f}_t)$ as

$$\widehat{\tau}(x_{it}, z_{it}, \mathbf{f}_t) = \widehat{\mu}_2(z_{it}, \mathbf{f}_t) + \widehat{\varphi}(\widehat{\varepsilon}_{1,it})\widehat{\sigma}_2(z_{it}, \mathbf{f}_t).$$

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

Our approach: 2nd step,contd.

- The conditional output-oriented frontier in the original units can be equivalently written as

$$\widehat{\tau}(x_{it}, z_{it}, \mathbf{f}_t) = y_{it} + \widehat{\beta}(\widehat{\varepsilon}_{1,it}, \widehat{\varepsilon}_{2,it})\widehat{\sigma}_2(z_{it}, \mathbf{f}_t),$$

  which has a nice interpretation: we see that the directional distance from the observed input-output point (x_{it}, y_{it}) facing external conditions (z_{it}, t) to the efficient frontier is given by the "pure" efficiency measure evaluated at the pure input-outputs $(\widehat{\varepsilon}_{1,it}, \widehat{\varepsilon}_{2,it})$ rescaled by the local standard deviation $\widehat{\sigma}_2(z_{it}, \mathbf{f}_t)$.
- Farrell-type efficiency estimates are obtained as

$$\widehat{\lambda}_t\,(x_{it}, y_{it}|z_{it}) = \frac{\widehat{\tau}(x_{it}, z_{it}, \mathbf{f}_t)}{y_{it}} \geq 1, \tag{1}$$

  with equality to 1 for points on the estimated conditional frontier (having pure efficiency $\widehat{\beta}$ equal to zero).

Nonparametric Time Conditional Efficiency Frontier

- first stage: average behaviour of production process (x and y) as function of common and external factors;
- second stage: 'idiosyncratic' efficiency cleaned by the time and common effects dependences. This represents the unexplained part of the conditional efficiency measures and can be interpreted as the idiosyncratic part of the efficiency, i.e. an efficiency scores cleaned by the external effects (here common factor and Z = FDI).
  - $\mathbf{f}_t = (t, x_{.t}, y_{.t})$: enables us to eliminate the common time factor effect (which captures CSD), 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 affecting by common shocks (captured by common factors $\mathbf{f}_t$) and facing different environmental factors at different time periods (Z_{it}).

Robust Parametric Estimation of Frontier

$\rightarrow$ Florens and Simar (JoE 2005):

- Once we have the 'pure' inputs and output we can check is some popular parametric production function (Cobb-Douglas) can fit the obtained frontier.
- Project all the data points on the FDH frontier.
- Using OLS we fit our 'favourite' parametric model (Cobb-Douglas).

Robust Parametric Estimation of Frontier, contd.

→ Florens and Simar (JoE 2005):

- Cobb-Douglas parametric frontier model (linear model because the units are already in log)

$$\widehat{\varepsilon}^{\delta}_{y,it} = \alpha + \beta_1 \widehat{\varepsilon}_{1,it} + \beta_2 \widehat{\varepsilon}_{2,it} + \eta_{it}$$

  where η_{it} is the fitting noise;

- Estimated Cobb-Douglas frontier parametric model

$$\widehat{\varepsilon}^{\delta}_{y,it} = \widehat{\alpha} + \widehat{\beta}_1 \widehat{\varepsilon}_{1,it} + \widehat{\beta}_2 \widehat{\varepsilon}_{2,it}$$

Robust Parametric Estimation of Frontier, contd.

$\rightarrow$ Florens and Simar (JoE 2005):

- Estimated Cobb-Douglas frontier parametric model in the original units

$$\widehat{y}^{\delta}_{it}(x_{it} \mid w_{it}) = \mu_y(w_{it}) + \sigma_y(w_{it}) \left(\frac{x_{1,it} - \mu_1(w_{it})}{\sigma_1(w_{it})}, \frac{x_{2,it} - \mu_2(w_{it})}{\sigma_1(w_{it})} \right)$$

  thus

$$\widehat{y}^{\delta}_{it}(x_{it} \mid w_{it}) = \mu_y(w_{it}) + \sigma_y(w_{it})[\widehat{\alpha} - \widehat{\beta}_1 \frac{\mu_1(w_{it})}{\sigma_1(w_{it})} - \widehat{\beta}_2 \frac{\mu_2(w_{it})}{\sigma_2(w_{it})}]$$

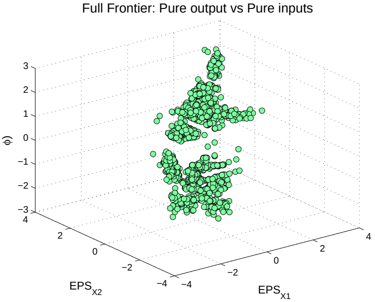
$$+ \widehat{\beta}_1 \frac{\sigma_y(w_{it})}{\sigma_1(w_{it})} x_{1,it} + \widehat{\beta}_2 \frac{\sigma_y(w_{it})}{\sigma_2(w_{it})} x_{2,it}$$

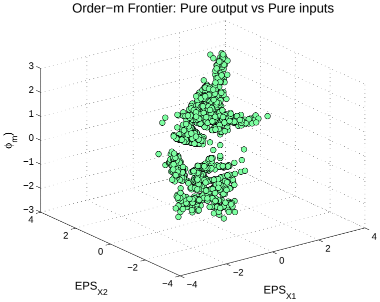
- We loot at the exponential of:
  - $\tilde{\alpha}_{it} = \mu_y(w_{it}) + \sigma_y(w_{it})[\widehat{\alpha} - \widehat{\beta}_1 \frac{\mu_1(w_{it})}{\sigma_1(w_{it})} - \widehat{\beta}_2 \frac{\mu_2(w_{it})}{\sigma_2(w_{it})}]$ in function of time
  - $\tilde{\beta}_{1,it} = \widehat{\beta}_1 \frac{\sigma_y(w_{it})}{\sigma_1(w_{it})}$
  - $\tilde{\beta}_{2,it} = \widehat{\beta}_2 \frac{\sigma_y(w_{it})}{\sigma_2(w_{it})}$

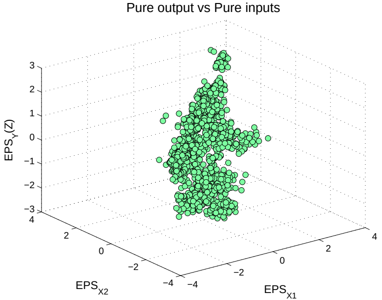
Data

- 23 OECD members: Australia, Austria, Belgium, Canada, Hong Kong, Denmark,, Finland, France, Germany, Greece, Ireland, Israel, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden,Turkey, the United Kingdom and the United States;
- 21 developing countries: Argentina, Bolivia, Chile, Côte d'Ivoire, Dominican Republic, Ecuador, Honduras, Jamaica, Kenya, Madagascar, Malawi, Mexico, Morocco, Nigeria, Panama, Philippines, Thailand, Venezuela, Zambia, Zimbabwe.
- Observation period: 1970-2007.
  - output: GDP in millions of US dollars at 2005;
  - effective inputs:
    - labour defined as average annual hours actually worked;
    - capital stock in millions of US dollars at 2005 (constructed using the perpetual inventory method - depreciation rate 6% (*e.g.*, Hall and Jones, 1999; Iyer *et al.*, 2008).
  - FDI: net inflow foreign direct investment
- Data sources: Capital 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.

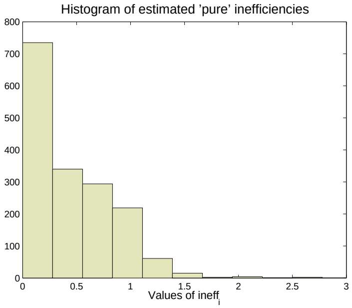
Empirical Evidence: 1

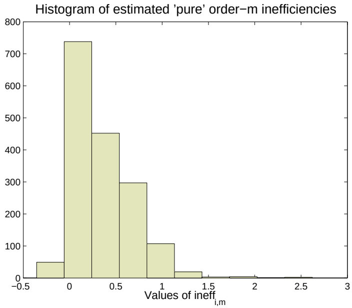




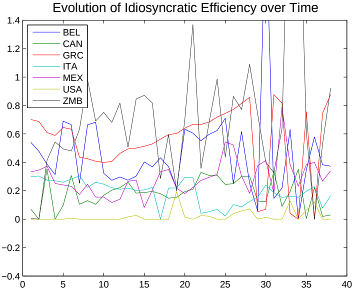


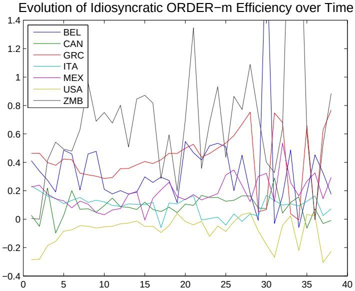
Empirical Evidence: 2



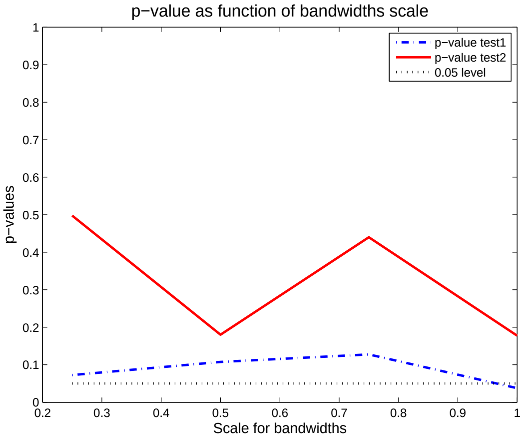


Empirical Evidence: 3

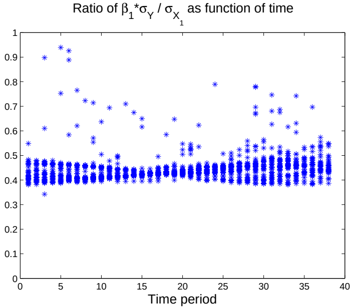


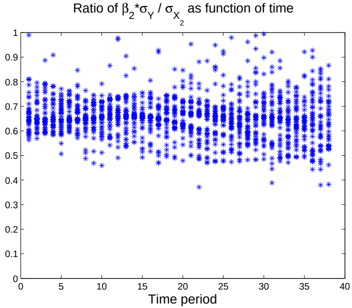


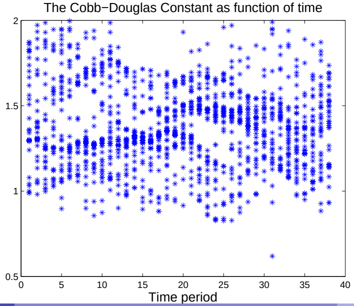
Test of independence between Z and *epsilon*



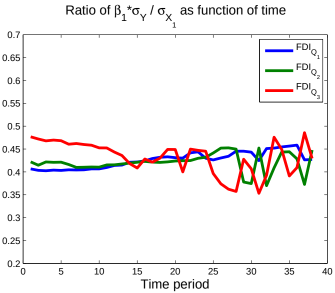
Empirical Evidence: 4

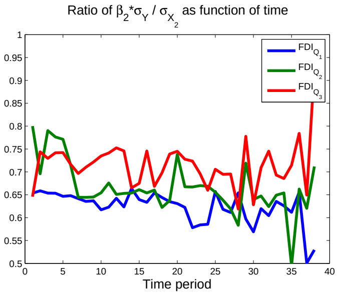


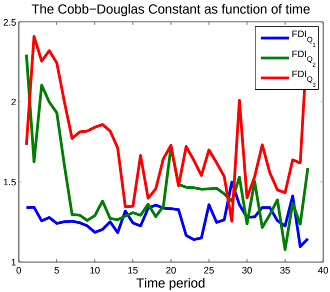




Empirical Evidence: 5







Conclusion

- unified non parametric framework for accommodating simultaneously the problem of model specification uncertainty, potential endogeneity and cross-section dependence in modelling technical efficiency in frontier models
- two-step procedure $\rightarrow$ **Florens et al. (2014);** effects on the production technology (shift of the frontier) and on efficiency distribution (movement away/towards the frontier) in a set-up dealing with both endogeneity and cross section dependence;
- importance of interdependencies of efficiency and global factors attributable to:
  - economy-wide shocks that affect all units in the cross section but with different intensities, by assuming a multifactor error process characterized by a finite number of unobserved common factors;
  - as global shocks and business cycles;
  - cross-section dependence (CSD) introduced by unobserved (heterogeneous) time-specific factors the conventional estimators would be seriously biased (Pesaran 2006, Bai 2009).

Conclusion, contd.

- robust parametric estimation of Cobb-Douglas frontier $\rightarrow$ **Florens and Simar (2005)**
  - Cross country framework (44 countries over 1970-2007) $\rightarrow$ production inefficiencies $=$ distance of the individual country's production from the frontier $\rightarrow$ sluggish adoption of new technologies (Ahn and Sickles 2000) $\rightarrow$ efficiency improvement will represent productivity catch-up via technology diffusion;
  - the effect seems to change over time;
  - FDI affect technological changes (shifts of the frontier) $\rightarrow$ increase competition between home and foreign firms;
  - the frontier itself shifts over time.

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

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.

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

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

Findlay, R.: 1978, Relative backwardness, direct foreign investment, and the transfer of technology: a simple dynamic model, *Quarterly Journal of Economics* **92**, 1–16.

Florens, J. and Simar, L.: 2005, Parametric approximations of nonparametric frontiers, *Journal of Econometrics* **124**, 91–116.

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

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

Kapetanios, G., Mitchell, J. and Shin, Y.: 2014, A nonlinear panel data model of cross-sectional dependence, *Journal of Econometrics* **179**, 134–157.

Mastromarco, C. and Simar, L.: 2014, Effect of fdi and time on catching-up: New insights from a conditional nonparametric frontier analysis, *Journal of Applied Econometrics* **forthcoming**.

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

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

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

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.