

HOW IMPORTANT IS INNOVATION? A BAYESIAN FACTOR-AUGMENTED PRODUCTIVITY MODEL BASED ON PANEL DATA

GEORGES BRESSON

Université Paris II

and

Sorbonne Universités

JEAN-MICHEL ETIENNE

Université Paris-Sud 11

PIERRE MOHNEN

UNU-MERIT, Maastricht University

This paper proposes a Bayesian approach to estimating a factor-augmented GDP per capita equation. We exploit the panel dimension of our data and distinguish between individual-specific and time-specific factors. On the basis of 21 technology, infrastructure, and institutional indicators from 82 countries over a 19-year period (1990 to 2008), we construct summary indicators of each of these three components in the cross-sectional dimension and an overall indicator of all 21 indicators in the time-series dimension and estimate their effects on growth and international differences in GDP per capita. For most countries, more than 50% of GDP per capita is explained by the four common factors we have introduced. Infrastructure is the greatest contributor to total factor productivity, followed by technology and institutions.

Keywords: Bayesian Factor-Augmented Model, Innovation, Markov Chain Monte Carlo, Panel Data, Productivity

1. INTRODUCTION

Why growth rates differ, what explains differences in productivity, and what are the conditions for economic development are fundamental questions that continue

We would like to thank participants at the International Forum of Statistics, Renmin University, Beijing, China (July 14–15, 2012), the International Conference on Panel Data, Banque de France, Paris (July 5–6, 2012), the Latin American Meeting of the Econometric Society, Santiago, Chile (November 10–13, 2011), and the Conference on MicroEvidence on Innovation in Developing Countries (MEIDE), San José, Costa Rica (June 27–28, 2011). We would also like to thank Jacques Mairesse, the Guest Editors Théophile Azomahou, Raouf Boucekkine, and Bart Verspagen, and an anonymous referee for helpful comments and suggestions. The usual disclaimer applies. Address correspondence to: Pierre Mohnen, UNU-MERIT Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail : mohnen@merit.unu.edu.

to preoccupy economists. As noted by Easterly and Levine (2001), starting with Solow (1957), a growing body of research has suggested that after accounting for physical and human capital accumulation, “something else” accounts for the majority of cross-country differences in the level and the growth rate of gross domestic product (GDP) per capita. The term “total factor productivity (TFP)” is used to refer to that “something else.” Whereas TFP was viewed by Abramowitz (1956) as a “measure of our ignorance,” various explanations have been proposed: technology, externalities, the sector composition of production, scale economies, capacity utilization, institutions, and adoption of information and communication technologies, to name only a few. Empirical evidence on the relative importance of each of these influences is far from unanimous.

As emphasized by Moral-Benito (2012), the main area of effort after a decade in the empirical growth literature has been to select appropriate variables to include in linear growth regressions. Including all potential explanatory variables at once leads to the curse of dimensionality. Using Bayesian model averaging (BMA) techniques, some authors have selected variables proposed as growth determinants among a total of more than 140 variables [see also Fernandez et al. (2001) and Durlauf et al. (2005)]. If K is the number of potential explanatory variables, there are 2^K possible models and BMA, using Markov Chain Monte Carlo (MCMC) model composition, selects the “best” model. But one major disadvantage with BMA is that some important variables, such as R&D or education, may be eliminated in favor of other variables such as life expectancy, culture, or religion. BMA, which can be described as a generalized stepwise method, may leave the reader doubtful about the interpretation of the resulting model in the absence of the usual explanatory variables.

We propose to use factor-augmented regressions (FAR) to avoid these problems. Besides the traditional inputs (labor and capital), some other variables such as technology, infrastructure, and institutions are often advanced to explain the evolution and the country gaps in productivity. How exactly those explanatory variables affect the TFP residual requires some structural multiequation modeling. This is not the purpose of this paper. Our intention here is to try to ascertain the importance of these “explanations” of the residual on the basis of a certain number of indicators that supposedly capture those broad explanations. To take the example of innovation, instead of including separately measures such as R&D, the number of new products, the number of patent applications or grants, the number of trademark applications, and the number of publications, we shall try to create some kind of index of innovation on the basis of which we could compare the performance and the contribution to productivity in different countries.

To construct these indices we use panel data on 21 indicators for 82 countries and 19 periods to identify common factors that pertain to technology, infrastructure, and institutions. We generalize in the panel dimension the approach of common factors, which has been proposed a long time ago [see Bartholomew et al. (2009)] and has received new attention in recent years in the literature on common features [Anderson et al. (2006)]. We then apply these common factors to the explanation

of international differences in productivity. The idea behind factor analysis is to summarize various indicators into a limited number of common factors that explain most of the correlations between the individual indicators in order to reduce the dimension of the analysis.

Compared with the previous studies, we innovate in two respects. On one hand, we allow two kinds of common factors, those in the cross-sectional dimension and those in the time-series dimension. This is like replacing the cross-sectional-specific and the time-series-specific errors in the two-way error components model with common factors in the two respective dimensions. On the other hand, instead of first estimating the common factors and afterward inserting them into a FAR on productivity, we propose a more robust Bayesian approach based on uninformed priors and MCMC simulation, where all equations are estimated simultaneously.

The paper is organized as follows. In Section 2 we present the background on FAR. In Section 3 we propose two ways to estimate, based on panel data, a factor-augmented productivity model with multiple indicators: the frequentist approach and the Bayesian approach. After presenting the data in Section 4, we analyze in Section 5 the results, the individual and time factor scores, and the estimated factor-augmented productivity equations. Last, we conclude on the effects of these indicators on total factor productivity using FAR on panel data (Section 6).

2. THE BACKGROUND ON FACTOR-AUGMENTED REGRESSIONS

In the literature on FAR, the usual specification for a time series is

$$\begin{cases} y_t = X_t\beta + \zeta' f_t + u_t, \\ z_{tm} = \sum_{j=1}^{pr} \gamma_{mj} f_{tj} + e_{tm}, \quad m = 1, \dots, M, t = 1, \dots, T, \end{cases} \tag{1}$$

where y_t is the dependent variable at time t , X_t is a $(1 \times K_X)$ vector of the primary inputs at time t , and $z_t = (z_{t1}, z_{t2}, \dots, z_{tM})'$ is an $(M \times 1)$ vector of predictors at time t . M may be large. The unobserved regressors $f_{tj} : f_t = (f_{t1}, f_{t2}, \dots, f_{tpr})'$ are the common factors and the γ_{mj} are the factor loadings. e_{tm} is an idiosyncratic error term [see Stock and Watson (2002), Bai and Ng (2002, 2006, 2008), Hecq et al. (2006), Ludvigson and Ng (2009), Gonçalves and Perron (2012), and Gospodinov and Ng (2013), to mention only a few].

We can write

$$z_t \underset{(M \times 1)}{=} \underset{(M \times pr)}{\Gamma} \underset{(pr \times 1)}{f_t} + \underset{(M \times 1)}{e_t}. \tag{2}$$

z_t is the vector of predictors (or control variables), Γ contains the factor loadings, and f_t is the vector of factor scores. None of the $\{\gamma_{mj}\}$, $\{f_{tj}\}$, and $\{e_{tm}\}$ are observed, but they can be estimated from the observed data set $\{z_{tm}\}$. In particular, with factor analysis, we can obtain estimated factors $\tilde{f}_t = E[f_t|z_t]$ and run a standard regression of y_t on X_t and on \tilde{f}_t to estimate β and ζ . Bai and Ng (2006) have shown that the standard errors of the second-step parameter estimates

(β', ζ') must account for the estimation error from the first step. Fortunately, no such adjustment is necessary if $\sqrt{T}/M \rightarrow 0$.

In a panel data context, we may be interested in the estimation of a static linear panel model, defined as

$$y_{it} = X_{it}\beta + Z_{it}\omega + u_{it}, i = 1, \dots, N, t = 1, \dots, T, \tag{3}$$

where y_{it} is the dependent variable for cross section i at time t , X_{it} is a $(1 \times K_X)$ vector of the main inputs, and Z_{it} is a $(1 \times K_Z)$ vector of predictors, where K_Z is large and some Z_{it} may be time-invariant or individual-invariant. u_{it} is the sum of individual (α_i)-and time (ξ_t)-specific effects and a remainder term (ε_{it}). The specific effects may be either fixed or random. In the latter case, the disturbance u_{it} is a two-way error component:

$$u_{it} = \alpha_i + \xi_t + \varepsilon_{it} \text{ with } \alpha_i \sim N(0, \sigma_\alpha^2), \xi_t \sim N(0, \sigma_\xi^2), \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \tag{4}$$

One way to estimate this model is to use FE (or LSDV with individual and time dummies) or RE specifications. But taking into account a large number of predictors (i.e, the Z_{it} explanatory variables) leads to the curse of dimensionality if we try to include them all simultaneously. Controlling for fixed effects, by estimating them directly, is not without difficulty and is known as the ‘‘incidental parameter problem,’’ which manifests itself in biases and inconsistency, at least under fixed T [see Chamberlain (1980), and Nickell (1981), among others]. In other words, LSDV estimations of α_i are not consistent for short panels.

More radically, in the cross-dependence and common factors literature [see Pesaran (2006), Bai and Ng (2008), Eberhardt and Bond (2009), and Eberhardt and Teal (2011)], the usual specification avoids gathering information on a large number of predictors Z_{it} and reduces the model to

$$\begin{cases} y_{it} = X_{it}\beta + u_{it}, \\ u_{it} = A_{i,0} + \lambda'_i f_t + \varepsilon_{it}. \end{cases} \tag{5}$$

In this literature, TFP (here u_{it}) is viewed as a combination of a country-specific level $A_{i,0}$ and a set of common factors f_t with country-specific factor loadings λ_i . So, TFP is unobserved and, generally, specifications of u_{it} omit the term $A_{i,0}(= A_0 + \alpha_i)$, which expresses random differences in TFP levels across countries. The preceding model with a factor error structure encompasses the two-way fixed effect model. If $u_{it} = \lambda'_i f_t + \varepsilon_{it}$ and if we suppose that there are two common factors ($r = 2$), with $f'_t = (1, \xi_t)$, then $\lambda'_i = (\alpha_i, 1)$ for all i and t . So the individual effects α_i and the time effects ξ_t enter the model additively instead of interactively [see Holtz-Eakin et al. (1988)].

The literature on common factors for panel data is now huge [see Anderson et al. (2006), Moench et al. (2009), Gonçalves and Perron (2012), Komunjer and Ng (2010) and Kneip et al. (2012)] but many applications are variations of equation (5). A few authors have used FAR on panel data with time-invariant common factors [Lanjouw and Schankerman (2004); Fagerberg and Shrolec (2008)]. Nobody, to

our knowledge, has estimated a FAR on panel data with multiple-indicator factor models.

3. THE FACTOR-AUGMENTED REGRESSION MODEL: FREQUENTIST VS. BAYESIAN APPROACHES

Let us suppose that, in the FAR, Z_{it} is a $(1 \times K_Z)$ vector of inputs where K_Z is large. We need to rewrite equation (1):

$$\begin{cases} y_{it} &= X_{it}\beta + \zeta' f_t + u_{it}, u_{it} = \alpha_i + \xi_t + \varepsilon_{it}, \\ \begin{pmatrix} Z_{1,it} \\ Z_{2,it} \\ \vdots \\ Z_{K_Z,it} \end{pmatrix} &= \begin{pmatrix} \lambda_i^{(1)} & & & \\ & \lambda_i^{(2)} & & \\ & & \ddots & \\ & & & \lambda_i^{(K_Z)} \end{pmatrix} (e_{K_Z} \otimes f_t) + \begin{pmatrix} e_{1,it} \\ e_{2,it} \\ \vdots \\ e_{K_Z,it} \end{pmatrix}, \end{cases} \quad (6)$$

where e_{K_Z} is a $(K_Z \times 1)$ vector of ones. The ideal would be to estimate the preceding system of equations simultaneously with the matrix of factor loadings $\lambda_i^{(k)}$ and the slope parameters β and ζ . The main problem is that to estimate an $(r \times 1)$ vector of common factors f_t we would need a $(T \times NK_Z)$ matrix Z , whereas generally we have only a $(T \times N)$ matrix Z . There is thus a serious identification problem.

But if we still want to use the two dimensions of the panel data indicators, we can introduce additively both p_N individual common factors $q_i = (q_{i1}, \dots, q_{ip_N})'$ and p_T time common factors $f_t = (f_{t1}, \dots, f_{tp_T})'$. These individual common factors may be interpreted as capabilities. In our case, the FAR specification becomes

$$\begin{cases} y_{it} &= X_{it}\beta + \delta' q_i + \mu' f_t + \alpha_i + \xi_t + \varepsilon_{it} \\ &= X_{it}\beta + v_{it}, \\ \bar{Z}_i &= \Lambda q_i + e_i, \\ \bar{Z}_t &= \Gamma f_t + e_t, \end{cases} \quad (7)$$

where $\bar{Z}_i = (\bar{Z}_{i,1}, \dots, \bar{Z}_{i,L_N})'$, $\bar{Z}_t = (\bar{Z}_{t,1}, \dots, \bar{Z}_{t,L_T})'$. $\bar{Z}_{i,l}$ are the individual means ($= \sum_{t=1}^T Z_{it,l}/T$) with $L_N \leq K_Z$ and $\bar{Z}_{t,m}$ are the time means ($= \sum_{i=1}^N Z_{it,m}/N$) with $L_T \leq K_Z$. Λ (resp. Γ) is a matrix of constants called the individual factor loading matrix (resp. the time factor loading matrix) such that

$$\Lambda q_i = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1p_N} \\ & \ddots & \\ \lambda_{L_N 1} & \cdots & \lambda_{L_N p_N} \end{pmatrix} \begin{pmatrix} q_{i1} \\ \vdots \\ q_{ip_N} \end{pmatrix} \text{ and } \Gamma f_t = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1p_T} \\ & \ddots & \\ \gamma_{L_T 1} & \cdots & \gamma_{L_T p_T} \end{pmatrix} \begin{pmatrix} f_{t1} \\ \vdots \\ f_{tp_T} \end{pmatrix}. \quad (8)$$

The e_i (resp. the e_t) are assumed to be mutually uncorrelated and multivariate normally distributed. This specification can be viewed as a general factor-augmented two-way error component model.

This specification allows for the possibility that TFP is in part common to all countries ($\mu' f_t + \xi_t$), (i.e., representing the global dissemination of scientific knowledge) and in part country-specific ($\delta' q_i + \alpha_i$). So it leads to a more interesting economic interpretation than the standard common factor model of equation (5): $y_{it} = X_{it}\beta + \lambda_i' f_t + \varepsilon_{it}$.

3.1. The Frequentist Approach

We can use a multistep estimation method with factor analysis for the first two steps, which give the estimated factor scores for individual means \tilde{q}_{ij} ($= E[q_{ij}|\bar{Z}_{i,l}]$) (resp. for time means \tilde{f}_t ($= E[f_t|\bar{Z}_{t,m}]$)). In these two steps, generated regressors ($\tilde{q}_{i_1}, \dots, \tilde{q}_{i_{p_N}}$) and ($\tilde{f}_{t_1}, \dots, \tilde{f}_{t_{p_T}}$) can be obtained as the fitted values from regressions of multiple indicators ($\bar{Z}_{i,l}$) (resp. ($\bar{Z}_{t,m}$)) related to the individual latent common factors ($q_i, \dots, q_{i_{p_N}}$) [resp. the time latent common factors ($f_t, \dots, f_{t_{p_T}}$)]. Regression scores will appear as expected values of the factors, given the indicators [see Bartholomew et al. (2009)]. The third step uses the ML estimator of the two-way error component model.

Multistep estimation may lead to biased and inefficient estimators in the third step. We have not derived the asymptotic properties of this multistep approach and hence we do not know if the condition ($\sqrt{T}/N \rightarrow 0$) has two equivalents for $\bar{Z}_{i,l}$ and $\bar{Z}_{t,m}$. If we transpose the Bai and Ng (2006) condition for not having to adjust the standard errors at the third step, we obtain the conditions ($\sqrt{N}/L_N \rightarrow 0$) and ($\sqrt{T}/L_T \rightarrow 0$), which are unlikely to hold for short panels.

3.2. The Bayesian Approach

One way to avoid these problems in the frequentist multistep method is to estimate jointly, in one step, the system (7). This can be done using the Bayesian approach [see Press and Shigemasu (1997)]. The e_i (resp. the e_t) are assumed to be mutually uncorrelated and multivariate normally distributed as $MN(0, \Psi)$ [resp. as $MN(0, \Phi)$]. Ψ and Φ are not assumed to be diagonal. In other words, the probability laws $\ell(\cdot)$ for y_{it} , \bar{Z}_i , and \bar{Z}_t are

$$\begin{cases} \ell(y_{it}|X_{it}, \beta, \delta, \mu, \sigma_\varepsilon^2) = N(X_{it}\beta, \sigma_\varepsilon^2), \\ \ell(\bar{Z}_i|\Lambda, q_i, \Psi) = MN(\Lambda q_i, \Psi), \\ \ell(\bar{Z}_t|\Gamma, f_t, \Phi) = MN(\Gamma f_t, \Phi), \end{cases} \tag{9}$$

with $\sigma_y^2 = \delta'\delta + \mu'\mu + \sigma_\alpha^2 + \sigma_\xi^2 + \sigma_\varepsilon^2$, or equivalently

$$\begin{pmatrix} y_{it} \\ \bar{Z}_i \\ \bar{Z}_t \end{pmatrix} \sim MN \left[\begin{pmatrix} X_{it}\beta \\ \Lambda q_i \\ \Gamma f_t \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \delta'\Lambda' & \mu'\Gamma' \\ \Lambda\delta & \Psi & 0 \\ \Gamma\mu & 0 & \Phi \end{pmatrix} \right]. \tag{10}$$

Following Lindley and Smith (1972) [see Bresson and Hsiao (2011), Bresson et al. (2011)], we express model (7) in three stages of hierarchy.

1. The first stage of the hierarchy postulates a joint density function of the data $(y_{it}, \bar{Z}_i, \bar{Z}_t)$ conditional on $(X_{it}, q_i, f_i, \pi, \Lambda, \Gamma)$ such that

$$p(y_{it}, \bar{Z}_i, \bar{Z}_t | X_{it}, q_i, f_i, \pi, \Lambda, \Gamma) \propto p(y_{it} | X_{it}, q_i, f_i, \pi) \cdot p(\bar{Z}_i | q_i, \Lambda) p(\bar{Z}_t | f_i, \Gamma), \tag{11}$$

where $\pi = (\beta', \delta', \mu)'$.

2. The second stage of the hierarchy postulates prior distributions of $(\pi, \Lambda, \Gamma, \sigma_\alpha^2, \sigma_\xi^2, \sigma_\varepsilon^2, \Psi, \Phi)$:

$$p(\pi | X_{it}) \sim MN(\bar{\pi}, \Omega_\pi), \tag{12}$$

$$p(\Lambda | \bar{Z}_i) \sim MN(\bar{\Lambda}, \Omega_\Lambda), \tag{13}$$

$$p(\Gamma | \bar{Z}_t) \sim MN(\bar{\Gamma}, \Omega_\Gamma), \tag{14}$$

$$p(\Psi^{-1}) \sim W_{LN}((\rho_\Psi R_\Psi)^{-1}, \rho_\Psi), \tag{15}$$

$$p(\Phi^{-1}) \sim W_{LT}((\rho_\Phi R_\Phi)^{-1}, \rho_\Phi), \tag{16}$$

$$p(\sigma_\alpha^2) \sim IG\left(\frac{\tau_\alpha}{2}, \frac{\eta_\alpha}{2}\right), \tag{17}$$

$$p(\sigma_\xi^2) \sim IG\left(\frac{\tau_\xi}{2}, \frac{\eta_\xi}{2}\right), \tag{18}$$

$$p(\sigma_\varepsilon^2) \sim IG\left(\frac{\tau_\varepsilon}{2}, \frac{\eta_\varepsilon}{2}\right). \tag{19}$$

3. The third stage of the hierarchy postulates prior distributions of $(\bar{\pi}, \bar{\Lambda}, \bar{\Gamma})$:

$$\bar{\pi} \sim MN(\bar{\bar{\pi}}, \Omega_{\bar{\pi}}), \bar{\Lambda} \sim MN(\bar{\bar{\Lambda}}, \Omega_{\bar{\Lambda}}), \text{ and } \bar{\Gamma} \sim MN(\bar{\bar{\Gamma}}, \Omega_{\bar{\Gamma}}). \tag{20}$$

We have assumed that Ψ^{-1} follows a Wishart distribution (a multivariate generalization of the gamma distribution) with scale matrix $(\rho_\Psi R_\Psi)$ and degrees of freedom ρ_Ψ . We have also assumed that $\sigma_\alpha^2, \sigma_\xi^2$, and σ_ε^2 are independent and follow inverse gamma distributions. The scale hyperparameter $(\frac{\eta_\alpha}{2}, \frac{\eta_\xi}{2}$, or $\frac{\eta_\varepsilon}{2}$) controls the precision of the priors. Small values of $(\frac{\eta_\alpha}{2}, \frac{\eta_\xi}{2}$, or $\frac{\eta_\varepsilon}{2}$) correspond to precise priors and the view that $\sigma_\alpha^2, \sigma_\xi^2$, or σ_ε^2 is probably constant over individuals, implying nearly homoskedastic disturbances. Large values of $(\frac{\eta_\alpha}{2}, \frac{\eta_\xi}{2}$, or $\frac{\eta_\varepsilon}{2}$) convey the view that disturbances may be quite variable or heteroskedastic. Generally, to implement the Gibbs sampler, we fix the values of the shape $(\frac{\tau_\alpha}{2}, \frac{\tau_\xi}{2}$, or $\frac{\tau_\varepsilon}{2})$ and scale $(\frac{\eta_\alpha}{2}, \frac{\eta_\xi}{2}$, or $\frac{\eta_\varepsilon}{2})$ hyperparameters.

Let $Z^{(i)'} = (\bar{Z}_1, \dots, \bar{Z}_N)$, $Z^{(t)'} = (\bar{Z}_1, \dots, \bar{Z}_T)$, $Q^{(i)'} = (q_1, \dots, q_N)$, and $F^{(t)'} = (f_1, \dots, f_T)$. Combining the priors (12)–(20) with the data $(y_{it}, \bar{Z}_i, \bar{Z}_t)$, we can obtain the posteriors of $\theta = (\pi, \Lambda, \Gamma, \Psi, \Phi, \sigma_\alpha^2, \sigma_\xi^2, \sigma_\varepsilon^2)$:

$$\begin{aligned}
 p(\theta | y_{it}, \bar{Z}_i, \bar{Z}_t) &\propto \prod_{i=1}^N \prod_{t=1}^T (2\pi\sigma_y^2)^{-\frac{1}{2}} \left\{ \exp \left[-\frac{1}{2\sigma_y^2} (y_{it} - X_{it}\beta)' (y_{it} - X_{it}\beta) \right] \right\} \\
 &\times |\Psi|^{-\frac{N}{2}} \exp \left[-\frac{1}{2} \text{tr} \Psi^{-1} (Z^{(i)} - Q^{(i)}\Lambda)' (Z^{(i)} - Q^{(i)}\Lambda) \right] \\
 &\times |\Phi|^{-\frac{T}{2}} \exp \left[-\frac{1}{2} \text{tr} \Phi^{-1} (Z^{(t)} - F^{(t)}\Gamma)' (aZ^{(t)} - F^{(t)}\Gamma) \right] \\
 &\times |\Omega_\pi|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\pi - \bar{\pi})' \Omega_\pi^{-1} (\pi - \bar{\pi}) \right] \\
 &\times |\Omega_\Lambda|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\Lambda - \bar{\Lambda})' \Omega_\Lambda^{-1} (\Lambda - \bar{\Lambda}) \right] \\
 &\times |\Omega_\Gamma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\Gamma - \bar{\Gamma})' \Omega_\Gamma^{-1} (\Gamma - \bar{\Gamma}) \right] \\
 &\times (\sigma_\alpha^{-2})^{\frac{\tau_\alpha}{2} + 1} \left(\frac{\eta_\alpha}{2} \right)^{\frac{\tau_\alpha}{2}} \exp \left[-\sigma_\alpha^{-2} \frac{\eta_\alpha}{2} \right] \\
 &\times (\sigma_\xi^{-2})^{\frac{\tau_\xi}{2} + 1} \left(\frac{\eta_\xi}{2} \right)^{\frac{\tau_\xi}{2}} \exp \left[-\sigma_\xi^{-2} \frac{\eta_\xi}{2} \right] \\
 &\times (\sigma_\varepsilon^{-2})^{\frac{\tau_\varepsilon}{2} + 1} \left(\frac{\eta_\varepsilon}{2} \right)^{\frac{\tau_\varepsilon}{2}} \exp \left[-\sigma_\varepsilon^{-2} \frac{\eta_\varepsilon}{2} \right] \\
 &\times |\Psi|^{-\frac{1}{2}(\rho_\Psi - L_N - 1)} \exp \left[-\frac{1}{2} \text{tr} [(\rho_\Psi R_\Psi) \Psi^{-1}] \right] \\
 &\times |\Phi|^{-\frac{1}{2}(\rho_\Phi - L_T - 1)} \exp \left[-\frac{1}{2} \text{tr} [(\rho_\Phi R_\Phi) \Phi^{-1}] \right]. \tag{21}
 \end{aligned}$$

Unfortunately, there is no closed form for the posteriors. The posterior distributions of $\theta = (\pi, \Lambda, \Gamma, \Psi, \Phi, \sigma_\alpha^2, \sigma_\xi^2, \sigma_\varepsilon^2)$, given the observed data, are very complicated and are not amenable to analytical calculation or to direct Monte Carlo sampling. Hence MCMC is used to approximate the desired posterior distributions and we use the statistical package OpenBUGS [the open source variant of WinBUGS; see Spiegelhalter et al. (2000)].

In principle, all prior distributions are specified to be as uninformative as possible. A multivariate normal distribution, $N(0_K, 10^2 I_K)$, is chosen for the $(K \times 1)$ vector of hyperparameters $\bar{\pi}$. To increase the speed of convergence, we use the

one-way FE estimates for hyperparameters of the intercept, capital, and labor variables. An inverse-gamma prior (0.1, 0.1, 0.1) is chosen for the variance parameters σ_α^2 , σ_ξ^2 , and σ_ε^2 . Selecting a prior for the covariance matrices Ω_Λ and Ω_Γ turned out to be a more interesting and challenging problem. The conjugate prior, inverse Wishart with scale matrix R_Ψ (resp. R_Φ) and degrees of freedom ρ_Ψ (resp. ρ_Φ), is commonly used in practice. The degrees of freedom must satisfy $\rho_\Psi \geq L_N$ (resp. $\rho_\Phi \geq L_T$) to yield a proper prior distribution. The prior scale matrices R_Ψ and R_Φ are set to $10^{-2}I_{L_N}$ and $10^{-2}I_{L_T}$. $\bar{\pi}$ is a $(K \times 1)$ zero vector and $\Omega_{\bar{\pi}} = 10^2 I_K$. $\bar{\Lambda}$ and $\bar{\Gamma}$ are initialized as estimated factor loadings coming from the factor analysis. $\Omega_{\bar{\Lambda}} = 10^{-1}I_{L_N}$ and $\Omega_{\bar{\Gamma}} = 10^{-1}I_{L_T}$. Results from convergence diagnostics indicated that it was sufficient to burn in the first 5, 000 samples and take the subsequent 10, 000 samples.

4. THE DATA SET

The data for the indicators that we use to infer the common factors in the cross-sectional and time series dimensions are taken from the World Development Indicators compiled by the World Bank from officially recognized international sources¹ and the CANA data set [Castellacci and Natera (2011)]. For the productivity equation, we use the GDP (in 1997 constant US\$), the capital stock² (in 1997 constant US\$ and base year PPPs), the labor force, and the population from 82 countries over a 19-year period (from 1990 to 2008).

In our search for useful indicators we are driven by two criteria: first, we expect the common factors to pertain to innovation, infrastructure, and institutions indicators, and second, we retain variables that are available for a great number of countries and for which there are not too many missing observations³ over time. In the end we choose 21 indicators from 82 countries over a 19-year period (from 1990 to 2008).

The five technological indicators are R&D expenditures (in 1997 US\$), U.S. patents granted per country of origin, royalties and license fees (in 1997 US\$), scientific publications, and trademark applications.

The seven infrastructure indicators are telecommunication revenue (in 1997 US\$), Internet users, secure Internet servers, mobile and fixed-line subscribers, electricity consumption (in kwh), paved roads as a percentage of the whole roads length of the country, and registered carrier departures worldwide.

The nine institutional indicators are index of democracy and autocracy, from +20 (democratic) to 0 (autocratic), electoral self-determination, from 0 (no freedom) to 3 (high freedom), political rights, from 1 (low freedom) to 7 (total freedom), civil liberties, from 1 (low freedom) to 7 (total freedom), women's rights, from 0 (low women rights) to 9 (high women rights), physical integrity human rights, from 0 (no government respect) to 8 (full government respect), freedom of the press, degree of print, broadcast, and Internet freedom, from 0 (no freedom) to 100 (high freedom), freedom of speech, from 0 (government censorship) to 2 (no

government censorship) and freedom of association, from 0 (total restriction) to 2 (no restrictions).^{4,5}

5. THE RESULTS

5.1. The Benchmark

We first give results of one-way fixed effects (FE) regressions for productivity equations. We regress the log of the output per capita on three variables (log of capital per capita, log of labor per capita, and a trend) (see Table 1). The FE model leads to constant returns of scale with elasticities of capital and labor of 0.41 and 0.45. The exogenous technical progress, proxied by a time trend, is estimated to be 2.3%. The constant-returns-to-scale assumption cannot be rejected at a 10% level. The estimated coefficient of capital is close to those found by Easterly and Levine (2001). The absence of a correlation between the three inputs and the individual effects is rejected by the Hausman test, and hence also the validity of the random effects model. The variance of these individual effects represents more than 98% of the total residual variance. When we explicitly assume constant returns to scale [column (2)], the capital elasticity of output is estimated to be 0.408, leading to similar results as without imposing constant returns to scale. One may be surprised to get such a high estimated elasticity of capital (more than 0.4), but this result is quite usual and reminiscent of many problems in estimating production functions on macro data, especially for developing countries. As emphasized by Eberhardt and Teal (2011), a majority of empirical studies produce capital coefficients far in excess of 0.3.

Adding the 21 indicators Z_{it} [column (3)] strongly reduces the elasticity of capital from 0.4 to 0.2, which is compensated for by impacts of electricity consumption (0.16), telecom revenue (0.10), and to a lesser extent carrier departures (0.03), patents (0.014), and R&D (0.012). But, as expected, many indicators do not have significant effects or surprising signs. We hit the curse of dimensionality mentioned earlier. We now turn to FAR estimates that reduce the dimension of the explanatory variables of TFP.

5.2. The Individual and Time Factor Scores

All variables are log-transformed. For the 82 countries over the period 1990–2009, $L_N = 21$ individual means ($\bar{z}_{i,l}$) of indicators are constructed. For the construction of the time means ($\bar{z}_{t,m}$), we only keep the 14 indicators with the largest time variabilities.

As usual in factor analysis, the variables are standardized. We use the mean and standard deviation of ($\bar{z}_{i,l}$) for $l = 1, \dots, L_N$ and of ($\bar{z}_{t,m}$) for $m = 1, \dots, L_T$. This implies that a change of a composite variable over individuals (resp. over time) will reflect changes in each country's position relative to the other countries (resp. changes in the importance of the underlying indicators over time relative to other indicators).

TABLE 1. Fixed effects productivity equations

log(GDP/pop)	One-way FE			One-way FE			One-way FE with Z_{it}			One-way RE with Z_{it}		
	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat
log(capital/pop)	0.4105	0.0178	23.1241				0.1940	0.0195	9.9556			
log(labor/pop)	0.4567	0.0848	5.3861				0.2942	0.0673	4.3729			
Trend	0.0230	0.0008	29.5332	0.0223	0.0007	34.0014	0.0148	0.0015	9.5924	0.0125	0.0015	8.0913
log(capital/labor)				0.4088	0.0177	23.0559				0.2108	0.0197	10.6779
log(R&D)							0.0123	0.0031	3.9188	0.0114	0.0032	3.5714
log(patents)							0.0147	0.0047	3.1514	0.0174	0.0048	3.6477
log(royalties)							0.0033	0.0019	1.7543	0.0035	0.0019	1.8176
log(publications)							-0.0294	0.0038	-7.7994	-0.0291	0.0038	-7.5627
log(trademark appl.)							-0.0099	0.0036	-2.7300	-0.0040	0.0036	-1.1051
log(telecom rev.)							0.1051	0.0071	14.8482	0.1052	0.0072	14.5779
log(internet users.)							-0.0071	0.0021	-3.4266	-0.0081	0.0021	-3.8617
log(internet servers)							-0.0015	0.0009	-1.6363	-0.0012	0.0009	-1.3023
log(phones)							0.0120	0.0076	1.5896	0.0101	0.0077	1.3094
log(electricity)							0.1601	0.0180	8.9048	0.1363	0.0181	7.5464
log(carriers)							0.0321	0.0058	5.5749	0.0348	0.0059	5.9526
Paved roads							-0.0010	0.0003	-3.0585	-0.0010	0.0003	-2.9796
Democracy							-0.0009	0.0015	-0.6290	-0.0016	0.0015	-1.0634
Elect. self determ.							0.0015	0.0064	0.2384	-0.0023	0.0065	-0.3458

TABLE 1. Continued

log(GDP/pop)	One-way FE			One-way FE			One-way FE with Z_{it}			One-way RE with Z_{it}		
	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat	Coef.	S.E.	<i>T</i> -stat
Political rights							-0.0087	0.0049	-1.7767	-0.0063	0.0050	-1.2695
Civil liberties							0.0040	0.0055	0.7208	0.0048	0.0056	0.8491
Women rights							-0.0054	0.0024	-2.2858	-0.0065	0.0024	-2.7293
Physical rights							-0.0073	0.0022	-3.3796	-0.0068	0.0022	-3.0922
Freedom press							0.0013	0.0004	3.2040	0.0014	0.0004	3.4504
Freedom speech							-0.0033	0.0057	-0.5732	-0.0039	0.0059	-0.6721
Freedom assoc.							-0.0111	0.0064	-1.7298	-0.0084	0.0065	-1.2806
var (indiv)	0.7657			0.7619			0.7793			0.7310		
var (epsilon)	0.0132			0.0132			0.0075			0.0078		
Hausman test (<i>p</i> -value)	329.7900	0.0000		325.5800	0.0000		497.5300	0.0000		372.7400	0.0000	
c.r.s. (<i>p</i> -value)	2.6300	0.1051					11.8300	0.0006				

As we have three groups of indicators (technology, infrastructures, and institutions), we specify our FAR model as

$$\left\{ \begin{aligned} y_{it} &= X_{it}\beta + \delta'_{\text{techno}}q_{i,\text{techno}} + \delta'_{\text{infra}}q_{i,\text{infra}} + \delta'_{\text{institut}}q_{i,\text{institut}} \\ &\quad + \mu'f_t + \alpha_i + \xi_t + \varepsilon_{it}, \\ \bar{z}_{i,\text{techno}} &= \Lambda_{\text{techno}}q_{i,\text{techno}} + e_{i,\text{techno}}, \\ \bar{z}_{i,\text{infra}} &= \Lambda_{\text{infra}}q_{i,\text{infra}} + e_{i,\text{infra}}, \\ \bar{z}_{i,\text{institut}} &= \Lambda_{\text{institut}}q_{i,\text{institut}} + e_{i,\text{institut}}, \\ \bar{z}_t &= \Gamma f_t + e_t. \end{aligned} \right.$$

We could have merged all 21 indicators. But we preferred to split them into *three* clusters for two main reasons: first, it seems natural to suppose that indicators belonging to a specific cluster (for instance, the technology cluster) represent a coherent set and are not necessary correlated with other indicators belonging to another cluster (for instance, institutions).⁶ Second, we have tried to estimate the factor scores of the whole 21 indicators. We found only three common factors for the individual means, which seems consistent with our expectations, but the ranking of the countries on the basis of the scores led to a few unrealistic results.

The factor analysis is used to analyze the correlation matrix of the 21 indicators. The factor loadings are computed using the squared multiple correlations as estimates of the communality (the variance shared with other variables). Because all indicators are normalized, the sum of all eigenvalues of the correlation matrix is the total number of variables. The number of principal components that we should retain depends on how much information (i.e., unaccounted variance) we are willing to sacrifice, which, of course, is a question of judgment.

The Kaiser criterion, the scree plot, and the parallel analysis [see Allen and Hubbard (1986)] suggest retaining only one factor each for the technological indicators, the infrastructure indicators, and the institutional indicators (see the Supplementary Materials). The factor explains 76.94% (resp. 63.49%, 81.72%) of the total variance for the five technological indicators (resp. the seven infrastructure indicators and nine institutional indicators). All the variables have, in general, high positive weights: from 0.80 for publications to 0.95 for R&D among the technological capabilities, from 0.51 for electricity revenue to 0.95 for Internet servers among the infrastructure capabilities, and from 0.77 for the political rights to 0.97 for the physical rights among the institutional capabilities (see the Supplementary Materials).⁷ As all factor loadings are of the same sign and roughly of the same order of magnitude, the principal components for our three clusters of variables do indeed capture common factors.

The Bayesian approach gives similar results.⁸ The posterior mean, the standard error, and the Monte Carlo standard error of the mean [MC error; see Roberts (1995)] of the factor loadings for the individual means of indicators are close to the results of factor analysis (see the Supplementary Materials). One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error (MC error) for each parameter. This is an estimate of the difference between the mean

of the sampled values (which we are using as our estimate of the posterior mean for each parameter) and the true posterior mean. As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 10% of the sample standard error [see Brooks and Gelman (1998)].

The Kaiser criterion, the scree plot, and the analysis also suggest retaining only one time common factor, which explains 90% of the total variance of the time averages of the 14 individual indicators (see the Supplementary Materials). One could argue that this factor measures common shocks or *time trends*. All 14 variables have positive weights higher than 88%, indicating again the common factor interpretation of the first principal component. The Bayesian approach again gives similar results.

Figures 1–3 rank countries according to the estimated factors scores using the frequentist and Bayesian approaches. One can note a few differences between the two methods, but by and large the two approaches display the same distribution. We shall concentrate on the results from the Bayesian approach for reasons laid out in the next section. The interesting result concerns the relative positions of the countries for the technological, infrastructure, and institutional capabilities. In each figure, we have also drawn the $\pm 0.5\sigma$ and $\pm\sigma$ confidence intervals (short dash-dot and dashed lines). These confidence intervals include respectively 38% and 68% of the distribution around the mean. The $\pm\sigma$ confidence interval allows us to define 4 groups of country: $[-2; -\sigma]$: low capabilities, $]-\sigma; 0]$: medium low capabilities $]0; +\sigma]$: medium high capabilities and $] +\sigma; 2]$: high capabilities. Remember that the three factor scores have a standard normal distribution. The individual factor scores for our 82 countries will hence be distributed according to this standard normal distribution. Regarding the technological common factor, we notice the usual suspects at the upper end of the distribution (with high capabilities): the G-7 countries, but also some countries such as Spain, Belgium, Switzerland, Sweden, and Australia and two of the BRICS countries, China and Brazil. Most of the other EU countries, the remaining three BRICS countries, and the countries of the G-20 group belong to the medium-high group. It is somewhat astonishing that Israel and Finland do not belong to the high group. To a large extent the ranking of the countries in the four categories remains the same for the infrastructure and institutions common factors, with a few notable exceptions: the BRICS countries fall into the medium-low category regarding infrastructure and India even into the lowest category. Many of the latest new member states in the EU belong to the medium-high category, possibly thanks to structural adjustment funds from the EU. It is on the institutional front that there is the highest concentration of countries in the two tails of the distribution. China and Russia now belong to the low group, South Africa and India to the medium-low, and Brazil to the medium-high. Again, the most developed countries belong to the high end of the distribution, whereas many of the African countries are at the lower end of the distribution.

We have also estimated time scores using the frequentist and Bayesian approaches, showing a linear time trend over the period 1990–2008 (see the Supplementary Materials).

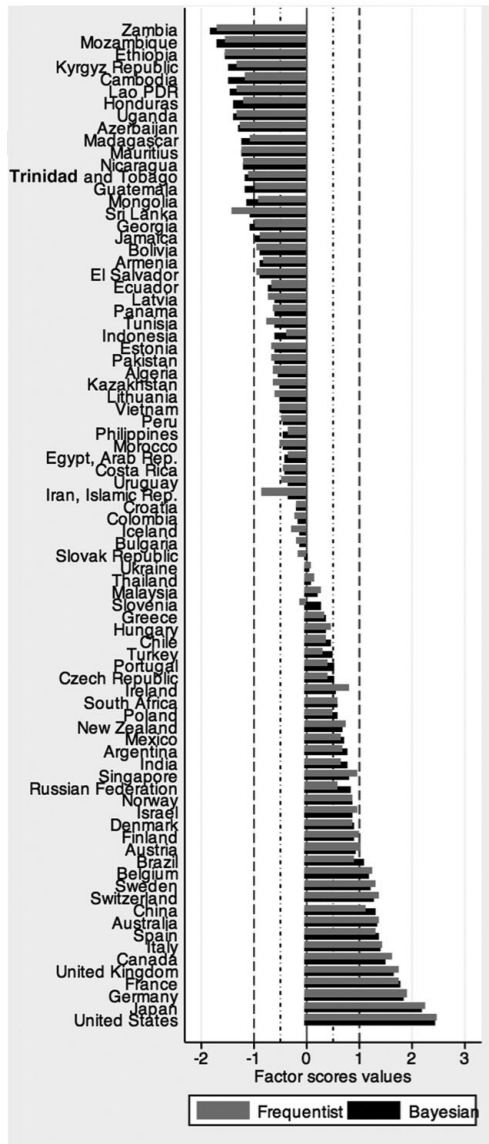


FIGURE 1. Individual factor scores on technological capabilities.

5.3. The Factor-Augmented Productivity Equation

These estimated factor scores, both for the individual means and for the time means, are used in the productivity equation as generated regressors. Table 2 gives the ML estimation of the general factor-augmented two-way error component model and the Bayesian posterior means, standard errors, and MC errors of this

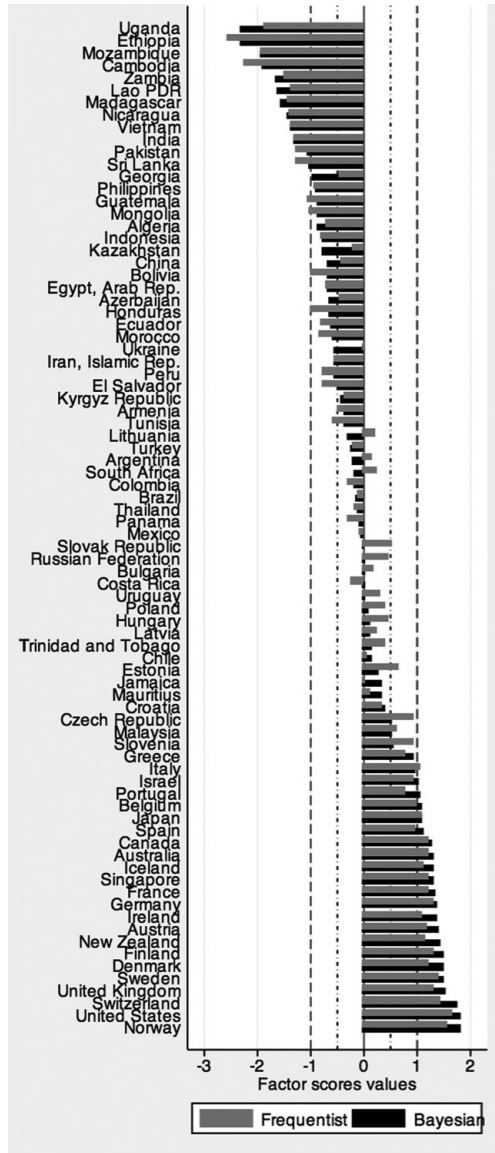


FIGURE 2. Individual factor scores on infrastructure capabilities.

productivity equation. We have assumed that the GDP per capita (in logs) depends on two inputs, the capital stock per capita (in logs) and the labor force per capita (in logs), augmented with the four estimated factor scores: three in the individual dimension (technology, infrastructure, and institutions) and one in the time dimension.

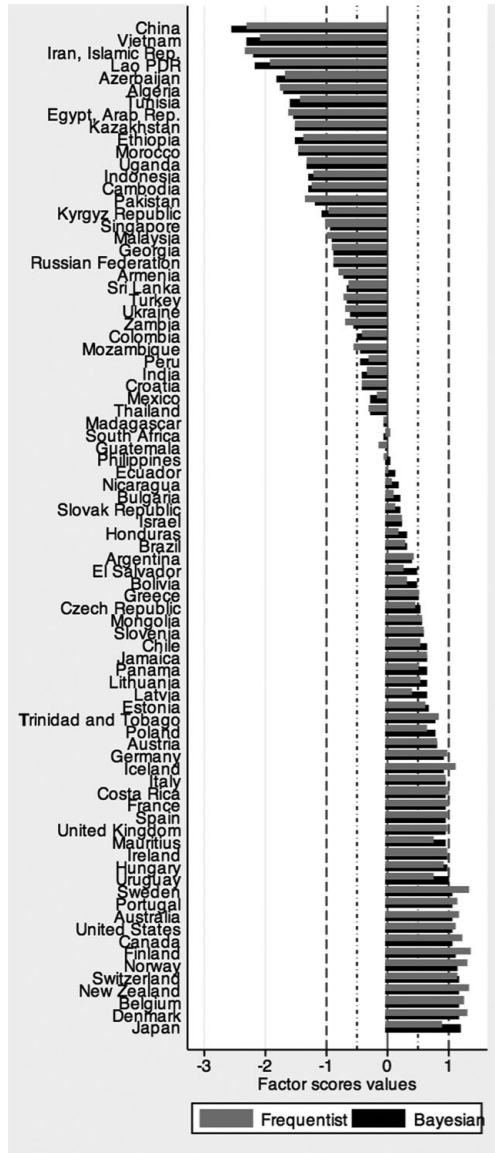


FIGURE 3. Individual factor scores on institutional capabilities.

If the estimated elasticity of production relative to capital (0.417) seems plausible, the elasticity of production relative to labor appears underestimated (0.256). Technological capabilities have a positive effect on productivity (0.23). But infrastructure seems to have the strongest impact (0.43) and, more surprisingly, institutions have a positive effect on productivity (0.20) quite similar to that of technological capabilities. Finally, the common trends (0.13) cannot be

TABLE 2. Factor-augmented productivity equations

log(GDP/pop)	Two-way FAR			Bayesian FAR		
	Coef.	S.E.	<i>T</i> -stat	Post. mean	S.E.	MC error
log(capital/pop)	0.4172	0.0167	24.9607	0.2879	0.0522	0.0052
log(labor/pop)	0.2568	0.0790	3.2512	0.5575	0.1292	0.0115
q_{1i} (technological)	0.2304	0.0678	3.3990	0.2553	0.0755	0.0072
q_{2i} (infrastructures)	0.4379	0.0828	5.2856	0.6050	0.0670	0.0061
q_{3i} (institutions)	0.2007	0.0569	3.5275	0.1989	0.0539	0.0050
f_t	0.1300	0.0097	13.4454	0.1407	0.0266	0.0023
Intercept	2.5684	0.2609	9.8460	2.5890	0.2191	0.0127
var(indiv)	0.1388			0.1407	0.0317	0.0021
var(time)	0.0015			0.0141	0.0054	0.0001
var(epsilon)	0.0120			0.0127	0.0006	0.0000
log-L (Y_{it})	990.9299			1213.9150		

associated with the time trend in the standard one-way FE Cobb–Douglas function (estimated at 0.023). This ML estimation of the general FAR two-way error component model leads to strong decreasing returns to scale: the 95% confidence interval is [0.524; 0.823], which seems unrealistic. Moreover, the mean relative error between observed and estimated output per capita is large: -8.58% .

These unrealistic results may come from biased and inconsistent estimates using the two-step frequentist approach. The posterior means obtained with the Bayesian method give more realistic results. The estimated elasticities for capital and labor are respectively (0.287) and (0.557), leading to quasi-constant returns to scale: [0.741; 0.949]. Factor scores on technology, infrastructure, and institutions all have a positive effect on productivity: the strongest is associated with infrastructure (0.605), followed by technology (0.255) and institutions (0.198). These estimated coefficients may be interpreted as semi-elasticities, i.e., percentage changes in GDP per capita due to a one-unit change (i.e., 1σ) in the capabilities. Finally, a one-standard-deviation change in the time common factor increases TFP by 0.14%. With the Bayesian approach, the variance of the individual effect σ_α^2 is similar to that of the FAR regression (0.14), but the variance of the time-specific effect is higher: 0.01 against 0.001. There is no significant difference in residual variance between the two approaches. The log-likelihood of the productivity equation shows the superiority of the Bayesian approach (1, 213.91) over the multistep ML estimation (990.92). This better fit is confirmed by the very low mean relative errors between observed and estimated output per capita: 0.61%.

Figure 4 shows the percentage contribution of TFP to output per capita. The part of output not due to the use of labor and capital is typically larger for the more developed countries. For only 10 countries of our sample is the contribution of TFP to output per capita lower than 50%. For the top third of our countries, this contribution is nearer to 60%, and this set comprises mostly the OECD countries.

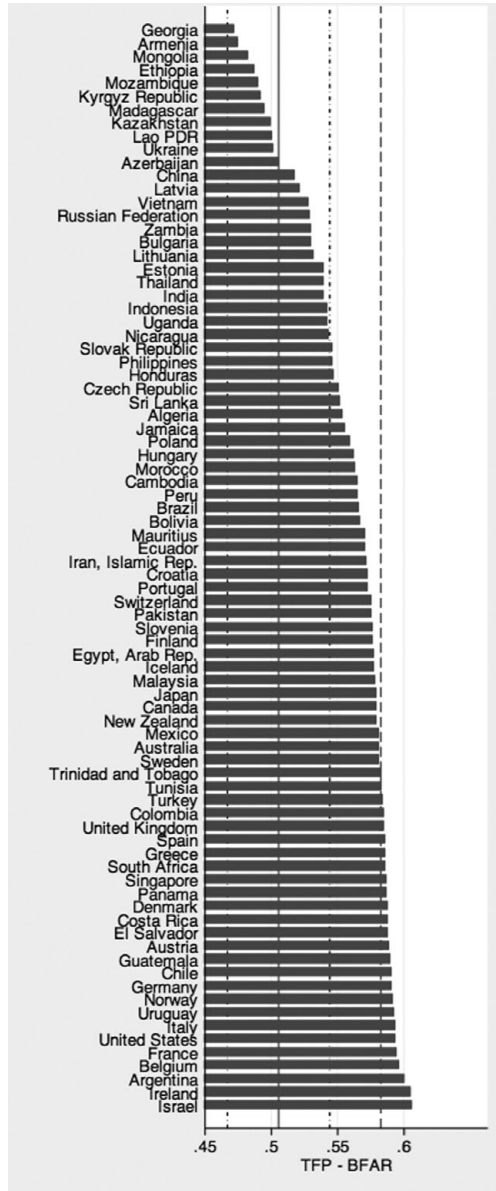


FIGURE 4. Share contributed by total factor productivity (individual means).

Figure 5 shows that the evolution of this share seems to be increasing over time. The contribution of TFP to output per capita has been increasing by about two percentage points over the past 15 years.

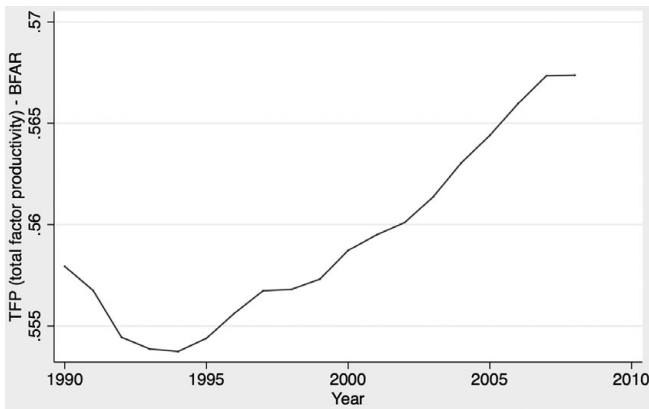


FIGURE 5. Share contributed by total factor productivity (time means).

We can compare our results with two earlier studies that used a somewhat different approach. Based on a sample of 115 countries and 25 indicators, Fagerberg and Srholec (2008) perform a factor analysis using two subsets of 3-year averages of these 25 indicators (1992–1994 and 2002–2004) and obtain four measures of capabilities. They do not assume separability between the four clusters of indicators, but they notice that some indicators are highly correlated with some factors and little with the others. The two factors that come out as significant in their GDP per capita regression are “innovation system” and “governance.” The other two, “political system” and “openness,” are insignificant. Their “innovation system” is essentially our factor “technology,” part of our factor “infrastructure,” and some variables regarding education that we have not included. Our factor “institutions” is closest to their factor “political system.” Their “governance” and “openness” factors are not really captured by our indicators.

A positive correlation between the global innovation index (which regroups seven clusters of variables: institutions, human capital and research, infrastructure, market sophistication, business sophistication, knowledge and technology outputs, and creative outputs) is also reported in the Global Innovation Index 2014 constructed by Cornell University, INSEAD, WIPO (2014). The Global Innovation Index contains more variables, but is not based on a factor analysis (although robustness experiments have been made using a factor analysis). In summary, our result of positive correlations between our three factors and GDP per capita is confirmed in other studies. The leading role of infrastructure over technology and institutions is something that does not come out so clearly from these other studies.

6. CONCLUSION

To evaluate the importance of technology, infrastructure, and institutions in explaining differences in total factor productivity among 82 countries between 1990 and 2008, we have estimated a factor-augmented GDP equation with 21 technology, infrastructure, and institutional indicators and unobserved country- and time-specific individual effects. The data are taken from the World Development Indicators database of the World Bank and the CANA data set. First, we have done a factor analysis, in which we have allowed for two kinds of common factors, those based on individual (country) means and those based on yearly means. Second, we have inserted the individual and time common factors into a factor-augmented productivity equation using the frequentist approach. Third, we have proposed a more robust Bayesian approach based on noninformative priors and a MCMC simulation, where all equations are estimated simultaneously. The Bayesian estimator leads to a better fit and somewhat more reasonable input elasticities. The explanation of TFP is, however, quite similar between the two approaches.

A sizeable portion, for most countries more than 50%, of total factor productivity is explained by the four common factors we have introduced. We show in addition that the TFP residual is driven by three clusters of variables: technology, infrastructure, and institutions. We have not tested the direction of causality, but we think that it is more likely to run from these factors, which take time to build up, than the other way around. Because there are many ways of measuring these influences, we have taken a series of measures related to them and estimated common factors for each. These common factors are then included in an extended Cobb–Douglas production function to explain TFP. Our results show that infrastructure is the greatest contributor to TFP, followed by technology and finally institutions. Infrastructure is at least twice as important as technology, whereas often it is assumed that TFP captures just technological change. Finally, our analysis reveals the weaknesses of certain countries regarding some of the determinants to TFP.

Our results might of course differ if we introduced another set or a larger set of indicators than the 21 that we have used, for which data were readily available. Some of our original data are obtained by interpolation. Increasing the quality of the construction of the data is another area of future research. For many countries, unfortunately, too many raw data were missing. Extending the analysis to more countries would be another worthy research endeavor.

NOTES

1. See <http://data.worldbank.org/data-catalog/world-development-indicators>.
2. The capital stock is computed from the gross capital formation data by the permanent inventory formula:

$$K_{it} = (1 - \delta) K_{it-1} + I_{it}, I_{it} = (\delta + g_y) K_{i0},$$

$$g_y = \left(\frac{\Delta Y}{Y} \right)_{90-92}, \delta = 5\%.$$

3. Missing values were linearly interpolated instead of using missing imputation methods.
4. The data for all 21 indicators, except secure Internet servers and trademark applications, come from the CANA data set.
5. Missing values were replaced by interpolation for each country. The numbers of interpolated observations are 7 for GDP, labor and population, R&D, patents, royalties, and telecom revenue, 11 for GDP growth, 26 for gross fixed capital formation, 12 for GDP deflator, 159 for trademark applications, and 994 for Internet servers. Negative values for some of the variables have been replaced by the lowest positive values in the sample: R&D (100,000, 7 cases), royalties (1,000, 37 cases), patents (1,328 cases), publications (0.5, 35 cases), trademark applications (10, 30 cases), Internet users (10, 220 cases), and Internet servers (1,000, 466 cases).
6. We would like to thank Jacques Mairesse for this suggestion.
7. Additional tables and figures are available as Supplemental Materials.
8. For the Bayesian approach to be comparable to the frequentist approach, we forced the number of common factors for each cluster of indicators to be equal to one, as we found with the frequentist approach.

REFERENCES

- Abramowitz, Moses (1956) Resource and output trends in the United States since 1870. *American Economic Review* 46(2), 5–23.
- Allen, Stuart J. and Raymond Hubbard (1986) Regression equations of the latent roots of random data correlation matrices unities on the diagonal. *Multivariate Behavioral Research* 21, 393–398.
- Anderson, Heather M., João Victor Issler, and Farshid Valid (2006) Common features. *Journal of Econometrics* 132, 1–15.
- Bai, Jushan and Serena Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bai, Jushan and Serena Ng (2006) Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions. *Econometrica* 74, 1133–1150.
- Bai, Jushan and Serena Ng (2008) Large dimensional factor analysis. *Foundations and Trends in Econometrics* 3 (2), 89–163.
- Bartholomew, David J., Ian J. Deary, and Martin Lawn (2009) The origin of factor scores: Spearman, Thomson and Bartlett. *British Journal of Mathematical and Statistical Psychology* 62, 569–582.
- Bresson, Georges and Cheng Hsiao (2011) A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris. *Advances in Statistical Analysis* 95(4), 501–529.
- Bresson, Georges, Cheng Hsiao, and Alain Pirotte (2011) Assessing the contribution of R&D to total factor productivity—A Bayesian approach to account for heterogeneity and heteroscedasticity. *Advances in Statistical Analysis* 95(4), 435–452.
- Brooks, Stephen P. and Andrew Gelman (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434–455.
- Castellacci, Fulvio and Jose Miguel Natera (2011) A New Panel Dataset for Cross-Country Analyses of National Systems, Growth and Development (CANA). Munich personal RePEc archive paper 28376.
- Chamberlain, Gary (1980) Analysis of covariance with qualitative data. *Review of Economic Studies* 47, 225–238.
- Cornell University, INSEAD, WIPO (2014) *The Global Innovation Index 2014: The Human Factor in Innovation*. Ithaca, NY.
- Durlauf, Steven N., Paul A. Johnson, and Jonathan R. Temple (2005) Growth econometrics. In Philippe Aghion and Steven Durlauf (eds.), *Handbook of Economic Growth*, Vol. 1, pp. 555–677. Amsterdam: Elsevier.

- Easterly, William and Ross Levine (2001) What have we learned from a decade of empirical research on growth? It's not factor accumulation: Stylized facts and growth models. *World Bank Economic Review* 15(2), 177–219.
- Eberhardt, Markus and Steve Bond (2009) Cross-Section Dependence in Nonstationary Panel Models: A Novel Estimator. University Library of Munich, Munich personal RePEc archive paper 01/2009.
- Eberhardt, Markus and Francis Teal (2011) Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys* 25(1), 109–155.
- Fagerberg, Jan and Martin Srholec (2008) National innovation systems, capabilities and economic development. *Research Policy* 37, 1417–1435.
- Fernandez, Carmen, Edurado Ley, and Mark Steel (2001) Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16, 563–576.
- Gonçalves, Silvia and Perron, Benoît (2012) Bootstrapping Factor-Augmented Regression Models. CIRANO working paper 2012s-12.
- Gospodinov, Nikolay and Serena Ng (2013) Commodity prices, convenience yields and inflation. *Review of Economics and Statistics* 95(1), 206–219.
- Hecq, Alain, Franz Palm, and Jean-Pierre Urbain (2006) Common cyclical features analysis in VAR models with cointegration. *Journal of Econometrics* 132(1), 117–141.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey Rosen (1988) Estimating vector autoregressions with panel data. *Econometrica* 56, 1371–1395.
- Kneip, Alois, Robin C. Sickles, and Wonho Song (2012) A new panel data treatment for heterogeneity in time trends. *Econometric Theory* 28, 590–628.
- Komunjer, Ivana and Serena Ng (2010) Indirect Estimation of Models with Latent Error Components. Mimeo, Columbia University.
- Lanjouw, Jean O. and Mark Schankermann (2004) Patent quality and research productivity: Measuring innovations with multiple indicators. *Economic Journal* 114, 441–465.
- Lindley, Dennis V. and Adrian F.M. Smith (1972) Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* 34, 1–41.
- Ludvigson, Sydney and Serena Ng (2009) Macro factors in bond risk premia. *Review of Financial Studies* 22, 5027–5067.
- Moench, Emanuel, Serena Ng, and Simon Potter (2009) Dynamic Hierarchical Factor Models. Federal Reserve Bank of New York staff report 412, December.
- Moral-Benito, Enrique (2012) Determinants of economic growth: A Bayesian panel data approach. *Review of Economics and Statistics* 94(2), 566–579.
- Nickell, Stephen (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1426.
- Pesaran, M. Hashem (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967–1012.
- Press, S. James and Kazuo Shigemasa (1997) Bayesian Inference in Factor Analysis. Technical report 243 (revised), Department of Statistics, University of California, Riverside.
- Roberts, Gareth O. (1995) Markov chain concepts related to sampling algorithms. In Wally R. Gilks, David Spiegelhalter, and Sylvia Richardson (eds.), *MCMC in Practice*, pp. 45–58. London: Chapman and Hall.
- Solow, Robert (1957) Technical change and the aggregate production function. *Review of Economics and Statistics* 39, 312–320.
- Spiegelhalter, David, Andrew Thomas, and Nicky Best (2000) WinBUGS, Bayesian Inference Using Gibbs Sampling, Version 1.3. User manual, MRC Biostatistics Unit, Cambridge, UK.
- Stock, James H. and Mark Watson (2002) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.