How important is innovation? A Bayesian factor-augmented productivity model on panel data

Georges Bresson, Jean-Michel Etienne, Pierre Mohnen

▶ **To cite this version:**

Georges Bresson, Jean-Michel Etienne, Pierre Mohnen. How important is innovation? A Bayesian factor-augmented productivity model on panel data. 2011. halshs-00812155

HAL Id: halshs-00812155
https://shs.hal.science/halshs-00812155

Preprint submitted on 11 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TEPP WORKING PAPER

N° 2011 - 6

How important is innovation? A Bayesian factor-augmented productivity model on panel data

GEORGES BRESSON, JEAN-MICHEL ETIENNE, PIERRE MOHNEN

www.tepp.eu

TEPP - Institute for Labor Studies and Public Policies
TEPP - Travail, Emploi et Politiques Publiques - FR CNRS 3126

ISSN 2110-5472

How important is innovation? A Bayesian factor-augmented productivity model on panel data.

Georges Bresson[a,*], Jean-Michel Etienne[b], Pierre Mohnen[c]

[a] *ERMES (CNRS), Université Panthéon-Assas Paris II / Sorbonne Universités*
[b] *Université Paris-Sud 11* [c] *Maastricht University, UNU-MERIT.*

May 2011

Abstract - This paper proposes a Bayesian approach to estimate a factor augmented productivity equation. We exploit the panel dimension of our data and distinguish individual-specific and time-specific factors. On the basis of 14 technology and infrastructure indicators from 37 countries over a 10-year period (1998 to 2007), we construct summary indicators of these two components and estimate their effect on the growth and the international differences in GDP per capita.
Keywords: Bayesian factor-augmented model, innovation, MCMC, panel data, productivity.

JEL classification: C23, C38, O47

*Corresponding author. ERMES (CNRS), Université Paris II / Sorbonne Universités, 12 place du Panthéon, 75231 Paris Cedex 05, France.
E-mail adresses: bresson-georges@orange.fr(G. Bresson), jean-michel.etienne@u-psud.fr (Jean-Michel Etienne), mohnen@merit.unu.edu (Pierre Mohnen).

1 Introduction

Why growth rates differ, what explains differences in productivity and what are the conditions for economic development, those are fundamental questions that continue to preoccupy economists. Besides the traditional inputs - labor, capital, energy and other intermediate inputs - technology, infrastructure and institutions are some of the variables that are often advanced to explain the evolution and the country gaps in productivity. How exactly those explanatory variables affect the total factor productivity residual requires some structural, multi-equation, modeling. This is not the purpose of this paper. Our goal in this paper is to try and 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, the number of publications, ... could we not 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 indexes we use panel data on 14 indicators, for 35 countries and 10 periods, to identify common factors, some of which, as we shall show, will pertain to innovation and infrastructure. We generalize in the panel dimension the approach of common factors that has been proposed a long time ago (see Bartholomew et al. (2009)) and received new attention in recent years in the literature on common features (Anderson, Issler and Vahid, 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 so as to reduce the dimension of the analysis. Here are a few examples of applications of this procedure. In macroeconomics, Hecq, Palm and Urbain (2006) identify two or three common cyclical features to explain comovements in annual GDP series of 5 Latin American countries over a 50-year period. In the analysing the quality of patents, Lanjouw and Schankerman (2004) estimate by factor analysis a patent quality index common to four quality indicators (claims, citations, family size and technology area). Fagerberg and Srholec (2008) use factor analysis on pooled data for 115 countries, 25 indicators and 2 periods (average values for 1992-1994 and for 2002-2004) to isolate four

types of indicators, which they interpret as capabilities: the development of the innovation system, the quality of governance, the political system and the openness of the economy.
Compared to the previous studies we innovate in two respects. On the one hand, we allow for two kinds of common factors, those in the cross-sectional dimension and those in the time series dimension. It is like replacing the cross-sectional specific and the time series specific errors in the two-way error components model by common factors in the two respective dimensions. On the other hand, instead of estimating first the common factors and inserting them afterwards in a factor-augmented productivity equation, we propose a more robust Bayesian approach based on uninformed priors and Markov Chain Monte Carlo simulation where all equations are estimated simultaneously.
The paper is organized as follows. We first present the factor augmented productivity model and then we explain two ways to estimate this model: the frequentist approach and the Bayesian approach. After presenting the data, we analyze the results, the individual and time factor scores followed by the estimation of the factor-augmented productivity equation.

2 The factor-augmented productivity model

As Bai and Ng (2008) pp.114 say "*the motivation for considering factor analysis is to deal with large number of variables. Use of a small number of factors as conditioning variables is a parsimonious way to capture information in a data rich environment, and there should be efficient gains over (carefully) selecting a handful of predictors*".
Generally, the factor-augmented regression (FAR) is used to add factors estimated by principal components analysis to a standard equation. Let N be the number of cross-section units (countries) and T be the number of time

series observations (years). A standard specification of a FAR[1] is defined as:

$$\begin{cases} y_{it} &= f\left(X_{it}, \widetilde{f}_t\right) + u_{it} = X_{it}\beta + \mu_i \widetilde{f}_t + u_{it} \\ z_{it,l} &= \gamma_l f_t + e_{it,l} \ , \ l = 1, ..., L, \ i = 1, ..., N, \ t = 1, ..., T \end{cases} \tag{1}$$

where y_{it} is the dependent variable (*i.e.*, the productivity) for cross-section i at time t, X_{it} is the $(1 \times K)$ vectors of inputs, $\widetilde{f}_t \left(= E\left[f_t | z_{it,l}\right]\right)$ are the estimated factor scores and ε_{it} is an error component:

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

$z_{it,l}$ is the l^{th} observed indicator and these indicators may include R&D, publications, trademark, royalties, ... γ_l is referred to as the factor loadings, f_t is the unobserved common factors (or common shocks). Since f_t is unobservable, we normalize by setting it variance to one: $f_t \sim N\left(0, 1\right)$. Each indicator contains an idiosyncratic error $e_{it,l} \sim N\left(0, \sigma_{e_l}^2\right)$. The regression model given by (1) can usually encompass many applications of interest. It is simply a regression with generated regressors $\widetilde{f}_t$ and $f\left(X_{it}, \widetilde{f}_t\right)$ is the estimated conditional mean of y_{it}. Generated regressor $\widetilde{f}_t$ can be obtained as the fitted values from a first step regression of a multiple-indicator ($z_{it,l}$) related to the latent common factor. As shown in Pagan (1984), sampling variability from the first step estimation is $O_p(1)$ in the second step. Consequently, the standard errors of the second step parameter estimates $(\beta', \mu_i)'$ in (1) must account for the estimation error from the first step. But, Bai and Ng (2008) have shown that no such adjustment is necessary if $\sqrt{T}/N \to 0$.

[1] The literature on common factors for panel data is now huge (see Anderson, Issler and Valid (2006)) but many applications concern the FAR (see (eq. 1)) or the random error model with common factors $f_t : y_{it} = X_{it}\beta + u_{it}$ with $u_{it} = \lambda_i' f_t + \eta_{it}$. For instance and recently, Gonçalves and Perron (2010) propose a bootstrapping FAR in a very specific model: $y_t = \beta' W_t + \alpha' f_t + e_t$ where the unobserved regressors f_t are the common factors in the following panel factor model: $X_{it} = \lambda_i' f_t + \eta_{it}$. Moench et a. (2009) propose dynamic hierarchical factors models to characterize within-block and between-block variations as well as idiosyncratic noise in large dynamic panels. Their model is estimated using a Markov Chain Monte-Carlo algorithm that takes into account the hierarchical structure of the factors. Komujer and Ng (2010) propose an indirect estimation of model with latent error components: $y_{it} = X_{it}\beta + \lambda_i' f_t + \eta_{it}$with $X_{it} = \lambda_i' g_t + x_{it}$, etc.

2.1 The frequentist approach: a multi-step model

In our case, we introduce both p_N individual common factors $(q_{i1}, ..., q_{ip_N})$ and a time common factor[2] (f_t) to express the quality of innovation, following Lanjouw and Schankerman (2004) and Fagerberg and Srholec (2008). So, the factor-augmented regression is now defined as:

$$\begin{cases} y_{it} &= f\left(X_{it}, \widetilde{q_{i1}}, ..., \widetilde{q_{ip_N}}, \widetilde{f_t}\right) + u_{it} = X_{it}\beta + \sum_{j=1}^{p_N} \delta_j \widetilde{q_{ij}} + \mu \widetilde{f_t} + u_{it} \\ \overline{z}_{i,l} &= \sum_{j=1}^{p_N} \lambda_{lj} q_{ij} + e_{i,l} \ , \ l = 1, ..., L_N \text{ with } p_N < L_N \\ \overline{z}_{t,m} &= \gamma_m f_t + e_{t,m} \ , \ m = 1, ..., L_T \end{cases} \tag{3}$$

where $\overline{z}_{i,l}$ are the individuals means $\left(= \sum_{t=1}^{T} z_{it,l}/T\right)$, $\overline{z}_{t,m}$ are the time means $\left(= \sum_{i=1}^{N} z_{it,m}/N\right)$, $\widetilde{q_{ij}}$ $(= E\left[q_{ij}|\overline{z}_{i,l}\right])$(resp. $\widetilde{f_t}\,(= E\left[f_t|\overline{z}_{t,m}\right])$ are the estimated factor scores for individuals means $\overline{z}_{i,l}$ (resp. for time means $\overline{z}_{t,m}$). λ_{lj} and γ_m are factor loadings and unobservable common factors $((q_{i1}, ..., q_{ip_N})$ and $f_t)$ are $iid\ N(0,1)$.
We can use a multi-step estimation method with principal component analysis for the first two steps. In these two steps, generated regressors $(\widetilde{q_{i1}}, ..., \widetilde{q_{ip_N}})$ and $\widetilde{f_t}$ can be obtained as the fitted values from regressions of multiple-indicators $(\overline{z}_{i,l})$ (resp. $(\overline{z}_{t,m})$) related to the individual latent common factors $(q_{i1}, ..., q_{ip_N})$ (resp. the time latent common factor f_t). Regression scores will appear as expected values of the factors, given the indicators (see Bartholomew et al. (2009)). The third step uses ML estimator of the one-way error component model.

2.2 The Bayesian approach

Since we have not derived the asymptotic properties of this multi-step approach, estimation may lead to biased and inefficient estimators in the third step and we don't know if the condition $\left(\sqrt{T}/N \to 0\right)$ on (eq. 1) has two equivalents for (eq. 3). One way to avoid these problems is to estimate jointly, in one step, the system (3). This can be done using Bayesian approach (see

[2]We can generalize the specification with $p_T > 1$ time common factors.

Press and Shigemasu (1997)). Let us re-write the factor-augmented regression in matrix form:

$$\left\{\begin{array}{lcl} y_{it} & = & X_{it}\beta + \underset{(1\times p_N)}{\delta'}\underset{(p_N\times 1)}{q_i} + \mu f_t + \alpha_i + \varepsilon_{it} \\ & = & X_{it}\beta + v_{it} \\ \underset{(L_N\times 1)}{\overline{z}_i} & = & \underset{(L_N\times p_N)}{\Lambda}\underset{(p_N\times 1)}{q_i} + \underset{(L_N\times 1)}{e_i} \\ \underset{(L_T\times 1)}{\overline{z}_t} & = & \underset{(L_T\times 1)}{\Gamma}\underset{(1\times 1)}{f_t} + \underset{(L_T\times 1)}{e_t} \end{array}\right. \tag{4}$$

where $\overline{z}_i = (\overline{z}_{i,1}, ..., \overline{z}_{i,L_N})'$, $\overline{z}_t = (\overline{z}_{t,1}, ..., \overline{z}_{t,L_T})'$, Λ (resp. Γ) is a matrix (resp. a vector) of constants called the individual factor loading matrix (resp. the time factor loading vector) and $q_i = (q_{i1}, ..., q_{ip_N})'$ is the individual factor scores vector for cross-section i. 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. This specification can be viewed as a general factor-augmented two-way error component model with[3]:

$$v_{it} = \kappa_i + \omega_t + u_{it} = \delta' q_i + \mu f_t + \alpha_i + \varepsilon_{it} \tag{5}$$

In other words, the probability laws $\ell(.)$ of y_{it}, $\overline{z}_i$ and $\overline{z}_t$ are:

$$\left\{\begin{array}{lcl} \ell(y_{it}|\beta, \delta, \mu, \sigma_\alpha^2, \sigma_\varepsilon^2) & = & N\left(X_{it}\beta, \sigma_y^2\right) \\ \ell(\overline{z}_i|\Lambda, q_i, \Psi) & = & MN(\Lambda q_i, \Psi) \\ \ell(\overline{z}_t|\Gamma, f_t, \Phi) & = & MN(\Gamma f_t, \Phi) \end{array}\right. \tag{6}$$

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

$$\begin{pmatrix} y_{it} \\ \overline{z}_i \\ \overline{z}_t \end{pmatrix} \sim MN\left(\begin{pmatrix} X_{it}\beta \\ \Lambda q_i \\ \Gamma f_t \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \delta'\Lambda' & \mu\Gamma' \\ \Lambda\delta & \Psi & 0 \\ \Gamma\mu & 0 & \Phi \end{pmatrix}\right) \tag{7}$$

Following Lindley and Smith (1972) (see Bresson and Hsiao (2011), Bresson et al. (2011)), we express model (4) in three stages of hierarchy.

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

$$p(y_{it}, \overline{z}_i, \overline{z}_t|X_{it}, q_i, f_t, \pi, \Lambda, \Gamma) \propto p(y_{it}|X_{it}, q_i, f_t, \pi).p(\overline{z}_i|q_i, \Lambda)\, p(\overline{z}_t|f_t, \Gamma) \tag{8}$$

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

[3] We could imagine that $\sigma_\alpha^2 = 0$.

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

$$p(\pi|X_{it}) \sim MN(\overline{\pi}, \Omega_\pi) \tag{9}$$

$$p(\Lambda|\overline{z}_i) \sim MN\left(\overline{\Lambda}, \Omega_\Lambda\right) \tag{10}$$

$$p(\Gamma|\overline{z}_t) \sim MN\left(\overline{\Gamma}, \Omega_\Gamma\right) \tag{11}$$

$$p\left(\Psi^{-1}\right) \sim W_{L_N}\left(\left(\rho_\Psi R_\Psi\right)^{-1}, \rho_\Psi\right) \tag{12}$$

$$p\left(\Phi^{-1}\right) \sim W_{L_T}\left(\left(\rho_\Phi R_\Phi\right)^{-1}, \rho_\Phi\right) \tag{13}$$

$$p\left(\sigma_\alpha^2\right) \sim IG\left(\frac{\tau_\alpha}{2}, \frac{\eta_\alpha}{2}\right) \tag{14}$$

$$p\left(\sigma_\varepsilon^2\right) \sim IG\left(\frac{\tau_\varepsilon}{2}, \frac{\eta_\varepsilon}{2}\right) \tag{15}$$

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

$$\overline{\pi} \sim MN\left(\overline{\overline{\pi}}, \Omega_{\overline{\pi}}\right) \text{ , } \overline{\Lambda} \sim MN\left(\overline{\overline{\Lambda}}, \Omega_{\overline{\Lambda}}\right) \text{ and } \overline{\Gamma} \sim MN\left(\overline{\overline{\Gamma}}, \Omega_{\overline{\Gamma}}\right) \tag{16}$$

We have supposed 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 supposed that σ_α^2 and σ_ε^2 are independent and follow inverse-gamma distributions. The scale hyperparameter $\left(\frac{\eta_\alpha}{2} \text{ or } \frac{\eta_\varepsilon}{2}\right)$ controls the precision of the prior. Small values of $\left(\frac{\eta_\alpha}{2} \text{ or } \frac{\eta_\varepsilon}{2}\right)$ correspond to precise priors and the view that σ_α^2 or σ_ε^2 is probably constant over individuals, implying nearly homoscedastic disturbances. Large values

of $\left(\frac{\eta_\alpha}{2} \text{ or } \frac{\eta_\varepsilon}{2}\right)$ convey the view that disturbances may be quite variable or heteroscedastic. Generally, in order to implement the Gibbs sampler, we fix values of shape $\left(\frac{\tau_\alpha}{2} \text{ or } \frac{\tau_\varepsilon}{2}\right)$ and scale $\left(\frac{\eta_\alpha}{2} \text{ or } \frac{\eta_\varepsilon}{2}\right)$ hyperparameters.
Let $Z^{(i)\prime} = (\overline{z}_1, ..., \overline{z}_N)$, $Z^{(t)\prime} = (\overline{z}_1, ..., \overline{z}_T)$, $Q^{(i)\prime} = (q_1, ..., q_N)$ and $F^{(t)\prime} = (f_1, ..., f_T)$. Combining the priors (9)-(16) with the data $(y_{it}, \overline{z}_i, \overline{z}_t)$, we can obtain the posteriors of $\theta = (\pi, \Lambda, \Gamma, \Psi, \Phi, \sigma_\alpha^2, \sigma_\varepsilon^2)$:

$$
\begin{aligned}
p\left(\theta | y_{it}, \overline{z}_i, \overline{z}_t\right) \quad \propto \quad & \prod_{i=1}^{N}\prod_{t=1}^{T}\left(2\pi\sigma_y^2\right)^{-\frac{1}{2}}\left\{\exp\left[-\frac{1}{2\sigma_y^2}\left(y_{it}-X_{it}\beta\right)'\left(y_{it}-X_{it}\beta\right)\right]\right\} \qquad (17)\\
& \times|\Psi|^{-\frac{N}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\Psi^{-1}\left(Z^{(i)}-Q^{(i)}\Lambda'\right)'\left(Z^{(i)}-Q^{(i)}\Lambda'\right)\right]\\
& \times|\Phi|^{-\frac{T}{2}}\exp\left[-\frac{1}{2}\mathrm{tr}\Phi^{-1}\left(Z^{(t)}-F^{(t)}\Gamma'\right)'\left(Z^{(t)}-F^{(t)}\Gamma'\right)\right]\\
& \times|\Omega_\pi|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\pi-\overline{\pi}\right)'\Omega_\pi^{-1}\left(\pi-\overline{\pi}\right)\right]\\
& \times|\Omega_\Lambda|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\Lambda-\overline{\Lambda}\right)'\Omega_\Lambda^{-1}\left(\Lambda-\overline{\Lambda}\right)\right] \qquad (18)\\
& \times|\Omega_\Gamma|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\Gamma-\overline{\Gamma}\right)'\Omega_\Gamma^{-1}\left(\Gamma-\overline{\Gamma}\right)\right]\\
& \times\left(\sigma_\alpha^{-2}\right)^{\frac{\tau_\alpha}{2}+1}\left(\frac{\eta_\alpha}{2}\right)^{\frac{\tau_\alpha}{2}}\exp\left[-\sigma_\alpha^{-2}\frac{\eta_\alpha}{2}\right]\\
& \times\left(\sigma_\varepsilon^{-2}\right)^{\frac{\tau_\varepsilon}{2}+1}\left(\frac{\eta_\varepsilon}{2}\right)^{\frac{\tau_\varepsilon}{2}}\exp\left[-\sigma_\varepsilon^{-2}\frac{\eta_\varepsilon}{2}\right]\\
& \times|\Psi|^{-\frac{1}{2}(\rho_\Psi-L_N-1)}\exp\left[-\frac{1}{2}\mathrm{tr}\left[\left(\rho_\Psi R_\Psi\right)\Psi^{-1}\right]\right]\\
& \times|\Phi|^{-\frac{1}{2}(\rho_\Phi-L_T-1)}\exp\left[-\frac{1}{2}\mathrm{tr}\left[\left(\rho_\Phi R_\Phi\right)\Phi^{-1}\right]\right]
\end{aligned}
$$

Unfortunately, there is no closed form for the posteriors. The posterior distributions of $\theta = (\pi, \Lambda, \Gamma, \Psi, \Phi, \sigma_\alpha^2, \sigma_\varepsilon^2)$, 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 WinBUGS (Spiegelhalter, Thomas and Best (2000)).
In principle, all prior distributions are specified to be as noninformative as possible. A multivariate normal distribution, $N(0_K, 10^2 I_K)$ is chosen for the the $(K \times 1)$ vector of hyperparameters $\overline{\pi}$. An inverse-gamma prior $(0.1, 0.1)$ is chosen for the variance parameters σ_α^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 $\rho_\Psi \geq N - L_N$ to yield a proper prior distribution[4]. Ω_Λ^{-1} is assigned Wishart prior with $\rho_\Psi = 23$ degrees of freedom with the prior scale matrix R_Ψ set to $10^{-2} I_{L_N}$. In the same way, Ω_Γ^{-1} is assigned Wishart prior with $\rho_\Phi = 23$ degrees of freedom with the prior scale matrix R_Φ set to $10^{-2} I_{L_T}$. $\overline{\overline{\pi}}$ is a $(K \times 1)$ zero vector and $\Omega_{\overline{\pi}} = 10^2 I_K$. $\overline{\overline{\Lambda}}$ and $\overline{\overline{\Gamma}}$ are initialized as estimated factor loadings coming from the principal component analysis. $\Omega_{\overline{\Lambda}} = 10^{-1} I_{L_N}$ and $\Omega_{\overline{\Gamma}} = 10^{-1} I_{L_T}$. Results from convergence diagnostics indicated that it was sufficient to burn-in the first $5,000$ samples and take the subsequent $20,000$ samples.

3 The dataset

The data for the indicators that we use to infer the common factors in the cross section and time series dimensions are taken from the World Development Indicators compiled by the World Bank from officially recognized international sources[5]. In our search for useful indicators we are driven by two criteria: first, we expect the common factors to pertain to innovation and infrastructure indicators, and second, we retain variables that are available for a great number of countries and for which there are not too many missing observations over time. In the end we choose 14 indicators. Since we work with averages over countries and over time, we need not have observations for all years for all the variables, but well for a good number of them (corresponding to the dimension L_T). For some variables, we need to reconstruct some of the data by interpolation and extrapolation. Likewise for the choice of the countries and the period of our analysis we are to a large extent conditioned by the availability of data.

[4]We have tried several initial values for the priors but results remain the same.
[5]See http://data.worldbank.org/data-catalog/world-development-indicators

We selected 14 indicators from 37 countries over a 10-year period (from 1998 to 2007). The countries are Argentina, Brazil, Canada, Chile, China, Colombia, Costa Rica, Croatia, the Czech Republic, Ecuador, Estonia, Finland, France, Germany, Honduras, Hungary, India, Ireland, Israel, Italy, Japan, Korea, Lithuania, Mexico, the Netherlands, Norway, Poland, Russia, Singapore, the Slovak Republic, Slovenia, Spain, Turkey, Thailand, Great Britain, Panama, Peru, the United States of America, and Uruguay. The indicators are R&D expenditure per capita, publications per capita, trademark per capita, royalties and license fees per capita, arable land per capita, total country area, education spending per capita, electric power consumption per capita, percentage of total roads paved, length of railway route, life expectancy at birth, secure internet servers par capita, internet users per capita and fixed broadband internet subscribers per capita. The first four are technological indicators and the other ones are infrastructure indicators in the broad sense, including measures of health, education, transportation and communication.
In appendix table A1, we list the variables, their measures and sources, the number of constructed observations, and the means of the indicators (in logs) at the individual level and at the yearly level.

4 The results

4.1 The individual and time factor scores

All variables are log-transformed. For the 37 countries over the $1998 - 2007$ period, $L_N = 14$ individual means $(\overline{z}_{i,l})$ of indicators are constructed. For the construction of the time means $(\overline{z}_{t,m})$ we only keep the indicators with the largest time variabilities and eliminate from the above list the following indicators: electricity consumption per capita, arable lands per capita, total country area, percentage of roads paved, and length of railway route life.
As usual in factor analysis, the variables are standardized. We use the mean and standard deviation of $(\overline{z}_{i,l})$ for $l = 1, ..., L_N$ and of $(\overline{z}_{t,m})$ for $m = 1, ..., L_T$. It implies that a change of a composite variable over individuals (resp. over time) will reflect changes in each country's position relative to other countries (resp. changes in the importance of the underlying indicators over time, relative to other indicators). The principal-component method is used to analyze the correlation matrix. The factor loadings are computed

using the squared multiple correlations as estimates of the communality (the variance shared with other variables).
The sum of all eigenvalues is the total number of variables. Proportion indicates the relative weight of each factor in the total variance. Remember that 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 judgemental question. In general and in the case of standardized data, three suggested rules are used:

- the Kaiser criterion suggests to retain those factors with eigenvalues equal or higher than 1.
- the scree plot (plot of eigenvalues against each principal component) and its typical form with an elbow. The number of principal components that need to be retain is given by the elbow.
- The parallel analysis (*i.e.*, a regression equation to estimate the eigenvalues for random data for standardized data inputs (see Allen and Hubbard (1986)). The observed eigenvalues should be higher than the estimated eigenvalues obtained from this regression.

In our case, the first criterion leads to retain three factors (see Figure 1). But the two first factors (resp. the third factor) explain 52.03%, 23.8% (resp. 8.96%) of the total variance. Hence cumulatively the two first factors account for 75.83% of the total variance. Furthermore, the parallel analysis shows that the difference between the third eigenvalue and the estimated one is small (0.06) as compared to the first two ones (5.92 and 2.07). So, we prefer to retain only the first two factors and it will allow us to get a more tractable interpretation of these factors.
As usual in factor analysis, loadings of the 14 indicators on the two retained factors are adjusted through the so-called "rotation" to maximize the differences between them. We use the "varimax normalized" rotation which assumes that the underlying factors are uncorrelated[6].
Factor loadings are the weights and correlations between each variable and

[6] We can use a more flexible "oblique" rotation if we suppose that there may be a good deal of correlation between the different factors. These are rotations of the axes that preserve the norms of the rows of the loadings but not the angles between the axes or the angles between the rows.

the factor. Remember that the higher the load the more relevant is the corresponding variable in defining the factor's dimensionality. A negative value indicates an inverse impact on the factor. The factor loadings of variables (R&D, publication, royalties, education, electricity, life expectation and paved roads) with factor 1 are much greater than those with factor 2 (see Table 1 and Figure 2). Indeed, 85.42% $((0.97 + 0.95 + 0.82 + 0.93 + 0.95 + 0.74 + 0.62)/7)$ of the total factor loadings of factor 1 are due to these variables. The Bayesian approach gives similar results.

Table 1 reports the posterior mean, the standard error and the Monte Carlo standard error of the mean (MC error, see Roberts (1996)) of the factor loadings. 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 factor loadings of variables trademark, internet servers and fixed broadbands with factor 2 are much greater than those with factor 1: their weights are respectively 0.80, 0.81 and 0.84. One can note that total area, rails and, in a lower proportion, arable lands have negative impacts on factor 2: their weights are respectively -0.90, -0.84 and -0.26. In other words, this second factor is defined by opposite factor loadings. One could argue that the common factor 1 mesures "*technological capabilities*" while the common factor 2 measures "*networks and geographical capabilities*".

Uniqueness is the variance that is 'unique' to the variable and not shared with other variables. It is equal to 1− communality. For example, 90% (resp. 60%, 34%) of the variance in arable lands (resp. in paved roads, and life expectations) is not shared with other variables in the overall factor model. On the contrary, R&D (resp. electricity, rails, internet users and internet servers) has a low variance not accounted for by other variables 5.2% (resp. 8.3%, 8.7%, 8.9% and 9.2%). Notice that the greater 'uniqueness', the lower the relevance of the variable in the factor model. The squared multiple correlations (SMC) between each variable and all other variables is a theoretical lower bound for communality, so it is an upper bound for uniqueness. These SMC are high for the major variables.

Table 1 also gives the Kaiser-Meyer-Olkin (KMO) measure of sampling ad-

equacy[7]. The overall KMO is 0.82. One can note 4 groups of variables according to this KMO measure. The first one, whose KMO is greater than 0.9 is composed of education, royalties and internet users. The second one, whose KMO is close to 0.8 is composed of R&D, publications, trademark, electricity, life expectation, internet servers and fixed broadbands. The third one, whose KMO is close to 0.7 is composed of roads, rails and area. The last one is only composed of arable lands whose KMO is very small (0.3). Last, Table 1 gives the scoring coefficients which are the weights used for estimating the regression factor scores. One can see that weights associated with R&D (0.16), publication (0.15), electricity (0.15), education (0.14) and royalties (0.12) are relatively important for the technological capabilities. For the networks and geographical capabilities, the scoring coefficients of fixed broadbands (0.22), internet servers (0.19) and trademark (0.19), on one side and those of rails (-0.26) and total area (-0.24), on the other side, are the most important for estimating the regression factor scores.

Just as for the individual means, the factor analysis has also been done for the time means. The Kaiser criterion, the scree plot and the parallel analysis suggest to retain only one factor (see Figure 3) which explains 81.7% of the total variance. One could argue that this factor mesures common shocks or "*time trends*". Table 2 shows that all the variables have high postive weights (> 0.94) except for the fixed broadbands (-0.47) and correlatively low uniqueness (< 0.05) except for R&D (0.11), trademark (0.53) and fixed broadbands (0.78). One can note that the fixed broadbands have a negative impact on this factor associated with time trends and, furthermore, we get high squared multiple correlations (SMC) between each variable and all other variables (> 0.98). On the contrary, the overall Kaiser-Meyer-Olkin (KMO) measure is low (0.53) leading to mediocre values except for electricity and broadbands. Figure 4 plots factors scores and defines 4 large (or 16 small) quadrants from the lowest to the highest scores leading to clusters of countries depending on their relative positions.

Once these scoring coefficients are obtained, factor scores are computed by regression method (see Bartholomew et al. (2009)). Regression factor scores

[7] This measure takes values between 0 and 1, with small values meaning that overall the variables have too little in common to warrant a factor analysis. Kaiser recommends accepting values greater than 0.5 as acceptable. Values below this should lead either to collect more data or rethinking which variables to include. Furthermore, values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are great and values above 0.9 are superb.

predict the location of each individual on the factor. These regression scores therefore appear as expected values of the factors, given the indicators. In this sense, they are reasonable predictors of the common factors (see Table 3, Figures 5 and 6 for individual means and Table 4 and Figure 7 for time means).
On Table 3, the estimated factor scores for technological (resp. networks and geographical) capabilities have, of course, zero mean and unit variance: -3.78×10^{-9} and 1.000 (resp. 2.01×10^{-9} and 1.000). The Bayesian approach does not give strictly similar means of scores for Factor 1 and Factor 2. The MC error confirms the accuracy of the posterior estimates. On Figure 5, for technological capabilities, the minimum value is for Honduras and the maximum for United States. We have also drawn the $\pm 0.5\sigma$ and $\pm\sigma$ confidence intervals (shortdash dot and dash lines). These confidence intervals include respectively the 38% and the 68% of the distribution around the mean. The $\pm\sigma$ confidence interval allows us to define 4 groups of country:

- $[-2; -\sigma]$: very small technological capabilities (Honduras, India, Ecuador, Peru, China)
- $]-\sigma; 0]$: small technological capabilities
- $]0; +\sigma]$: high technological capabilities
- $]+\sigma; 2]$: very high technological capabilities (USA, Canada, Norway, Finland, France, Japan, UK and Ireland).

Figure 6 gives the estimated factor score for the networks and geographical capabilities. Due to the negative area effect, the countries with very small geographical and networks capabilities are India, China, the Russian Federation, USA and Brazil. On the contrary, the countries with very high geographical and networks capabilities are Honduras, Panama, Costa Rica, Slovenia, Estonia and Israël. These capabilities express the positive effect of dense networks and the negative effect of countries with large sparsely populated areas. Again, the Bayesian approach does not give strictly similar means of scores for the common factor and the MC error also confirms the accuracy of the posterior estimates. Figure 7 gives the estimated factor scores for time trend. It is a quasi-linear trend and the $\pm\sigma$ confidence interval define 4 sub-periods: $[1998 - 1999]$, $[2000 - 2002]$, $[2003 - 2005]$ and $[2006 - 2007]$.

4.2 The factor-augmented productivity equation

These estimated factor scores, both for individual means and the time means, are used in the productivity equation as generated regressors. Figure 8 plots the factor scores on technological capabilities against GDP per capita (in logs). There is a very close correlation between these two variables. In the lower left quadrant, there are deviations from the regression line which come from a group of large countries like India, China and the Russian Federation and from small countries like Honduras and Ecuador. The correlation is less important between the factor scores on geographical and networks capabilities and the GDP per capita (in logs) (see Figure 9). This will be confirmed by the lower estimated effect in the productivity equation. Figure 10 plots the factor scores on time trends against GDP per capita (in logs) and shows a convex relationship, which reflects a kind of accelaration of time common factors.

Table 5 gives the ML estimation of the error component model and the Bayesian posterior means, standard errors and the MC errors of the productivity equation. We have supposed that the GDP per capita (in logs) depends on three inputs: the gross fixed capital formation (GFCF) per capita (in logs), the school enrollment tertiary (in logs) and the energy use per capita (in logs) augmented with the estimated factors scores. The elasticity of the productivity relatively to the investment (resp. the tertiary, the energy) is 0.745 (resp. 0.112 and -0.087) for the ML estimator and 0.747 (resp. 0.123 and -0.133) for the Bayesian posteriors means. The elasticity of investment is 5.8 times larger than those of the two other inputs. The negative effect of energy on productivity is quite surprising. The more productive the countries are, the less energy users they are. Actually, energy is not a primary input in the value-added measure of output. This needs further examination. Both technological capabilities and geographical and networks capabilities have positive effects on productivity. These are semi-elasticities, *i.e.*, the percentage changes in the GDP per capital from a one unit change in the capabilities. Technological capabilities effects are expected to be 0.314 (resp. 0.343) for the ML estimation (resp. for the Bayesian posterior means). The effects of geographical and networks capabilities are 0.143 (resp. 0.207) for the ML estimation (resp. for the Bayesian posterior means). Last, the time effects are 0.024 (resp. 0.028) for the ML estimation (resp. for the Bayesian posterior means). The variance of the individual effects σ_α^2 (resp. the reminder term σ_ε^2) is lower for the Bayesian approach (0.0385 against

0.1722) (resp. 0.007 against 0.084). The log-likelihood of the productivity equation shows the superiority of the Bayesian approach (385.53) over the multi-step ML estimation (321.140).
Figure 11 shows the good ajustment quality of the Bayesian factor-augmented productivity model as compared to the ML estimation procedure. For the frequentist approach (*i.e.*, the multi-step model), the differences between individual means of observed and predicted output per capita are relatively small except for some countries as Norway, UK, Spain or Korea and some relative errors are larger than $\pm 20\%$ (see Figure 12). On the contrary, with the Bayesian approach, we get very good estimates of individual means of GDP per capita since the relative errors are less than $\pm 1\%$ (see Figures 12 and 13). This good adjustment also holds for time means of estimated GPD per capita (see Figures 14 and 15).

5 Conclusion

To evaluate the importance of technology and infrastructure in explaning differences in GDP per capita, we have estimated a factor augmenting GDP per capita equation with 14 technology and infrastructure indicators taken from the World Development Indicators of the World Bank. First, we have done a factor analysis on 37 countries and a 10-year period in which we allow for two kinds of common factors, those based on individual means and those based on yearly means. In a third step, the estimation of the two types of factor loadings and of the productivity equation has been done using ML estimation. But, we have shown that a one-step more robust Bayesian approach, where all equations are estimated simultaneously, leads to better results. The estimates show a sizeable effect of the technological capabilities in explaining GDP per capita.

References

Allen, S.J. and R. Hubbard, 1986, Regresssion equations of the latent roots of random data correlation matrices unities on the diagonal, *Multivariate Behavioral Research*, 21, 393-398.

Anderson, H.M, Issler, J.V. and F. Valid, 2006, Common features, *Journal of Econometrics*, 132, 1-15.

Bai, J. and S. Ng, 2008, Large Dimensional Factor Analysis, in *Foundations and Trends in Econometrics* (W.H. Greene, Ed.), Vol. 3, No. 2, 89-163.

Bartholomew, D.J, Deary, I.J. and M. Lawn, 2009, The origin of factor scores: Spearman, Thomson and Bartlett, *British Journal of Mathematical and Statistical Psychology*, 62, 569-582.

Bresson, G. and C. 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* (forthcomming).

Bresson, G., Hsiao, C. and A. 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* (forthcomming).

Brooks, S.P. and A. Gelman, 1998, Alternative methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics*, 7, 434-455.

Fagerberg, J. and M. Srholec, 2008, National innovation systems, capabilities and economic development, *Research Policy*, 37, 1417-1435.

Fagerberg, J., Srholec, M. and B.Verspagen, 2010, The role of innovation in development, *Review of Economics and Institutions*, 1, 2, 1-29.

Gonçalves, S. and Perron, B., 2010, Bootstrapping factor-augmented regression models, Département de sciences économiques, CIREQ and CIRANO, *Université de Montréal*, November.

Hecq, A. , F. Palm and J.P. Urbain, 2006, Common cyclical features analysis in VAR Models with cointegration, *Journal of Econometrics*, 132(1), 117-141.

Komujer, I. and S. Ng , 2010, Indirect Estimation of Models with Latent Error Components, *University of California, San Diego*, WP, April.

Lanjouw, J.O. and M. Schankermann, 2004, Patent quality and research productivity: measuring innovations with multiple indicators, *The Economic Journal*, 114, 441-465.

Lindley, D.V. and A.F.M. Smith, 1972, Bayes estimates for the linear model, *Journal of The Royal Statistical Society,* B 34, 1-41.

Moench, E., Ng, E. and S. Potter, 2009, Dynamic hierarchical factor models, Federal Reserve Bank of New York, Staff Report no. 412, December.

Pagan, A. 1984, Econometric issues in the analysis of regressions with generated regressors, *International Economic Review*, 25, 221-247.

Press, S.J. and K. Shigemasu, 1997, Bayesian Inference in Factor Analysis (revised), Technical report n°243, *Department of Statistics, University of California,Riverside.*

Roberts, G.O., 1996, Markov chain concepts related to sampling algorithms, in *Markov Chain Monte Carlo in Practice*, Gilks, W.R., Richardson, S., and D. Spiegelhalter, (eds), Chapman and Hall, London, UK.

Spiegelhalter, D., Thomas, A. and N. Best, 2000, WinBUGS, Bayesian inference using Gibbs sampling, Version 1.3, User manual, *MRC Biostatistics Unit, Cambridge*, UK.

Appendix

Table A1 Data sources and constructions

Variable	Original source	Units and constructions	# of inter- and extrapolated missing values*
GDP per capita	World Bank and OECD National Accounts data	1997 US dollars using GDP deflator	
Gross fixed capital formation per capita	World Bank and OECD National Accounts data	1997 US dollars using gross fixed capital formation/GDP, GDP deflator, GDP, and GDP per capita	
Total area		Square Kilometers	
Arable land per capita	Food and Agriculture Organization	Hectares per person	
School enrolment, tertiary	UNESCO Institute for Statistics	Total enrolment in tertiary education as a percentage of eligible population of that age group	47
Public spending on education per capita	UNESCO Institute for Statistics	In 1997 US dollars, using public spending on education/GDP, GDP deflator, GDP and GDP per capita	59
R&D per capita	UNESCO Institute for Statistics	In 1997 US dollars, using R&D expenditure/GDP, GDP deflator, GDP and GDP per capita	39
Royalties and license fees, payments	International Monetary Fund, Balance of Payments Statistics Yearbook	In 1997 US dollars per 1000 workers, using GDP deflator and labor force	1
Labor force, total	International Labour Organization	persons	
Scientific and technical journal articles	National Science Foundation, Science and Engineering Indicators	# of articles per thousand workers	
Patent applications by residents	World Intellectual Property Organization (WIPO), WIPO Patent Report: Statistics on Worldwide Patent Activity.	# per thousand workers	
Total trademark applications	World Intellectual Property Organization (WIPO), WIPO Patent Report: Statistics on	# per thousand workers	1

	Worldwide Patent Activity.		
Electric power consumption per capita	International Energy Agency	In KwH per capita	
Fixed broadband internet subscribers	International Telecommunication Union, World Telecommunication/ICT Development Report and database, and World Bank estimates	# per 100 people	41 (83 negative nb converted to 0.01)
Internet users	International Telecommunication Union, World Telecommunication/ICT Development Report and database, and World Bank estimates	# per 100 people	
Secure internet servers	Netcraft (http://www.netcraft.com/) and World Bank population estimates	# per million people	
Rail lines	World Bank, Transportation, Water, and Information and Communications Technologies Department, Transport Division	Total route-Km	65
Roads paved	International Road Federation, World Road Statistics and electronic files	% of total roads	127
Primary energy use	International Energy Agency	Kg of oil equivalent per capita	(18 numbers < 1 have been converted to 1)
Life expectancy at birth	United Nations Population Division, Eurostat, national statistical offices, Secretariat of the Pacific Community	# of years	1

All the data come from the World Development Indicators, World Bank (http://data.worldbank.org/data-catalog/world-development-indicators).

*Total number of observations: 370.

Table A2 Individual means of indicators

Country	R&D	Public ations	Trade- marks	Royal- ties	Elec- tricity	arable land	Educa- tion	total area	Life- time	roads	rails	Internet servers	Internet users	Broad -band
Argentina	7.71	-1.76	1.33	10.35	7.70	-0.26	9.97	14.83	4.31	3.40	10.48	-10.64	2.20	-12.27
Brazil	8.19	-2.40	0.08	9.57	7.56	-1.12	9.69	15.96	4.26	1.71	10.29	-12.28	2.12	-15.96
Canada	10.89	0.34	0.88	12.55	9.74	0.37	11.86	16.12	4.38	3.69	10.45	-8.70	3.96	-10.55
Chile	8.10	-1.59	1.82	10.71	7.92	-2.30	9.92	13.54	4.35	3.14	8.80	-9.21	2.76	-11.25
China	7.21	-3.30	-0.71	8.12	7.20	-2.26	7.64	16.08	4.28	3.44	11.01	-15.16	1.21	-16.57
Colombia	6.07	-3.85	0.02	8.54	6.79	-2.90	9.26	13.95	4.27	2.69	7.67	-11.26	1.67	-13.72
Costa Rica	7.23	-3.03	1.74	10.16	7.38	-2.99	9.86	10.84	4.36	3.15	6.55	-8.11	2.52	-8.38
Croatia	8.87	-0.92	1.34	10.78	8.04	-1.52	9.77	10.94	4.31	4.46	7.91	-7.97	2.74	-7.49
Czech Rep.	9.29	-0.59	1.21	10.50	8.70	-1.21	10.48	11.28	4.32	4.61	9.15	-8.58	2.96	-10.32
Ecuador	5.21	-4.81	0.80	9.17	6.55	-2.21	7.06	12.51	4.30	2.80	6.87	-10.79	0.96	-11.30
Estonia	8.55	-0.58	2.09	9.92	8.55	-0.73	10.47	10.72	4.27	3.13	6.85	-5.94	3.58	-6.62
Finland	11.52	0.60	1.17	12.47	9.69	-0.86	12.14	12.73	4.36	4.16	8.67	-6.89	3.97	-8.64
France	11.00	0.13	1.10	11.47	8.94	-1.18	11.96	13.42	4.38	4.61	10.31	-10.09	3.25	-12.64
Honduras	3.81	-5.76	0.88	8.73	6.33	-1.75	8.38	11.63	4.26	3.02	4.32	-10.17	0.92	-4.61
Hungary	8.70	-0.56	1.08	11.70	8.18	-0.78	10.47	11.44	4.28	3.74	8.96	-8.68	2.84	-10.79
India	6.05	-3.47	-1.58	6.99	6.09	-1.88	7.64	15.01	4.13	3.88	11.05	-16.08	0.09	-15.07
Ireland	10.64	-0.05	1.05	15.71	8.69	-1.25	11.95	11.16	4.35	4.61	7.56	-7.34	3.25	-7.47
Israel	11.30	0.87	1.21	12.01	8.77	-3.00	11.73	9.94	4.38	4.61	6.62	-8.11	2.99	-7.70
Italy	10.19	-0.04	0.85	11.02	8.61	-1.98	11.63	12.62	4.38	4.61	9.70	-10.10	3.17	-11.65
Japan	11.60	-0.19	0.65	12.11	8.99	-3.36	11.71	12.84	4.40	4.35	9.91	-10.33	3.73	-12.91
Korea Rep.	10.45	-0.63	1.52	11.91	8.79	-3.35	10.86	11.50	4.35	4.35	8.07	-9.26	3.81	-11.04
Lithuania	8.12	-1.58	1.43	9.06	7.99	-0.54	10.12	11.09	4.27	3.34	7.47	-7.82	2.63	-8.11
Mexico	7.93	-2.49	0.34	9.28	7.52	-1.39	10.31	14.49	4.31	3.54	10.19	-11.68	2.14	-12.61
Norway	11.23	0.31	1.74	11.99	10.11	-1.65	12.71	12.87	4.37	4.35	8.34	-6.81	3.92	-8.64
Panama	7.15	-3.29	1.70	10.29	7.23	-1.74	9.71	11.27	4.31	3.52	4.34	-8.19	2.15	-9.57
Peru	5.72	-4.77	0.30	8.74	6.63	-2.00	8.83	14.07	4.27	2.61	7.69	-10.49	2.01	-13.50
Poland	8.15	-1.05	0.31	10.53	8.12	-1.08	10.32	12.65	4.31	4.21	9.94	-10.02	2.84	-10.30
Portugal	9.36	-0.82	1.09	10.97	8.36	-1.96	11.26	11.43	4.35	4.45	7.95	-8.43	3.12	-10.36

Russia	7.94	-1.56	-0.56	7.94	8.61	-0.16	8.72	16.65	4.19	4.33	11.36	-12.54	1.65	-10.85
Slovak Rep	8.40	-0.98	1.31	10.24	8.52	-1.32	10.35	10.80	4.30	4.47	8.20	-7.80	3.09	-8.17
Slovenia	9.86	-0.09	2.00	11.28	8.76	-2.44	11.14	9.92	4.34	4.61	7.11	-6.57	3.33	-7.28
Spain	9.88	-0.16	1.35	11.63	8.63	-1.16	11.32	13.13	4.38	4.60	9.55	-9.84	3.10	-11.23
Thailand	6.37	-3.65	-0.20	10.32	7.42	-1.42	9.25	13.15	4.22	4.62	8.31	-11.46	1.92	-12.39
Turkey	7.53	-1.51	0.48	8.95	7.45	-1.07	7.81	13.57	4.26	3.95	9.07	-11.44	2.00	-11.31
UK	10.92	0.44	0.24	12.44	8.72	-2.31	11.96	12.48	4.36	4.61	9.72	-9.54	3.76	-10.47
USA	11.51	0.27	0.52	11.70	9.50	-0.50	12.23	16.05	4.35	4.22	12.09	-10.88	3.99	-13.20
Uruguay	7.33	-2.20	2.09	8.44	7.59	-0.91	9.48	12.08	4.32	4.50	8.01	-7.73	2.67	-6.50

Table A3. Annual means of indicators

year	R&D	Publicat.	Tradem.	royalties	Electric.	Educat.	lifetime	roads	rails	Int. ser.	Int. user	broadband
1999	8.34	-1.61	0.83	10.14	8.02	10.05	4.30	3.86	8.65	-10.62	1.68	-9.84
2000	8.37	-1.56	0.97	10.24	8.06	10.10	4.30	3.87	8.66	-10.16	2.15	-10.37
2001	8.41	-1.54	0.89	10.30	8.07	10.15	4.31	3.87	8.64	-9.86	2.45	-12.16
2002	8.45	-1.51	0.85	10.37	8.10	10.17	4.31	3.89	8.65	-9.50	2.82	-12.31
2003	8.61	-1.47	0.87	10.55	8.13	10.28	4.31	3.90	8.65	-9.30	3.02	-11.83
2004	8.78	-1.42	0.87	10.68	8.16	10.40	4.32	3.91	8.65	-9.17	3.16	-11.34
2005	8.93	-1.38	0.91	10.76	8.18	10.53	4.32	3.91	8.70	-9.03	3.31	-10.83
2006	9.06	-1.35	0.93	10.89	8.22	10.64	4.32	3.92	8.69	-8.90	3.45	-10.36
2007	9.24	-1.31	0.94	11.04	8.25	10.76	4.33	3.93	8.69	-8.75	3.61	-10.12
1998	8.29	-1.64	0.77	9.97	8.00	9.60	4.30	3.86	8.65	-11.18	1.11	-6.63

Table 1 - Factor loadings and factor scores. Frequentist and Bayesian approaches. Individual means.

	Principal component analysis							Bayesian approach . Factor loadings					
Variable	Factor loadings		Uniqueness	kmo	smc	Factor scores		Factor 1			Factor 2		
	Factor 1	Factor 2				Factor 1	Factor 2	mean	s.e	MC error	mean	se	MC error
zi(R&D)	0.9730	-0.0381	0.0519	0.8204	0.9785	0.1615	-0.0688	0.9437	0.0483	0.00168	-0.0563	0.0462	0.00134
zi(publications)	0.9503	0.0131	0.0968	0.8106	0.9685	0.1547	-0.0534	0.9250	0.0559	0.00159	0.0292	0.0543	0.00139
zi(Trademark)	0.3679	0.8012	0.2227	0.8226	0.8806	0.0121	0.1972	0.3443	0.0647	0.00145	0.7797	0.0673	0.00174
zi(Royalties)	0.8209	0.2538	0.2617	0.9104	0.7907	0.1191	0.0202	0.7907	0.0740	0.00165	0.2206	0.0757	0.00192
zi(Education)	0.9289	0.1620	0.1108	0.9450	0.8960	0.1423	-0.0114	0.8962	0.0505	0.00115	0.1414	0.0509	0.00121
zi(Electricity)	0.9569	0.0298	0.0834	0.8787	0.9521	0.1548	-0.0492	0.9166	0.0455	9.518E-4	0.0507	0.0474	0.00126
zi(Land)	0.1569	-0.2640	0.9057	0.2990	0.7457	0.0415	-0.0816	0.1736	0.1366	0.00405	-0.1840	0.1499	0.00587
zi(Area)	-0.1062	-0.9005	0.1778	0.6723	0.9228	0.0366	-0.2401	-0.0769	0.0730	0.00171	-0.8388	0.0824	0.00315
zi(Life)	0.7407	0.3279	0.3439	0.8294	0.8435	0.1015	0.0453	0.7074	0.0844	0.00218	0.2988	0.0895	0.00289
zi(Roads)	0.6259	0.0752	0.6027	0.7518	0.7083	0.0979	-0.0170	0.5647	0.1172	0.00332	0.0494	0.1233	0.00434
zi(Rails)	0.3364	-0.8942	0.0873	0.6906	0.9198	0.1087	-0.2649	0.3232	0.0561	0.00126	-0.8492	0.0594	0.00178
zi(Internet servers)	0.4898	0.8175	0.0918	0.7866	0.9681	0.0311	0.1943	0.4801	0.0424	0.00106	0.7900	0.0445	0.00132
zi(Internet users)	0.9230	0.2424	0.0892	0.9047	0.9444	0.1365	0.0109	0.8916	0.0445	9.439E-4	0.2267	0.0456	0.00108
zi(Broadbands)	0.1553	0.8470	0.2584	0.8262	0.8522	-0.0254	0.2225	0.1341	0.0733	0.00192	0.8091	0.0783	0.00257
Overall				0.8192									

kmo : Kaiser-Meyer-Olkin measure of sampling adequacy

smc : Squared multiple correlations of variables with all other variables

mean : posterior mean

s.e : posterior standard error

MC error: Monte Carlo standard error

Table 2 - Factor loadings and factor scores. Frequentist and Bayesian approaches. Time means.

	Principal component analysis					Bayesian approach		
Variable	Factor loadings	Uniqueness	kmo	smc	Factor scores	Factor loadings		
						mean	s.e	MC error
zt(R&D)	0.9271	0.1405	0.4931	1.0000	0.1255	1.0440	0.0677	0.00254
zt(Publications)	0.9796	0.0403	0.5313	0.9999	0.1327	1.0570	0.0547	0.00214
zt(Trademark)	0.6851	0.5306	0.4555	0.8987	0.0928	0.9081	0.1280	0.00252
zt(Royalties)	0.9832	0.0333	0.5815	0.9994	0.1331	1.0580	0.0531	0.00212
zt(Education)	0.9839	0.0320	0.6276	0.9980	0.1332	1.0500	0.0579	0.00197
zt(Life)	0.9879	0.0241	0.5479	0.9997	0.1338	1.0560	0.0535	0.00197
zt(Internet servers)	0.9830	0.0338	0.5171	1.0000	0.1331	1.0410	0.0634	0.00163
zt(Internet users)	0.9840	0.0317	0.5179	1.0000	0.1333	1.0410	0.0628	0.00165
zt(Broadbands)	-0.4986	0.7514	0.5287	0.9912	-0.0675	0.4227	0.2203	0.00462
Overall			0.5363					

kmo : Kaiser-Meyer-Olkin measure of sampling adequacy

smc : Squared multiple correlations of variables with all other variables

mean : posterior mean

s.e : posterior standard error

MC error: Monte Carlo standard error

Table 3 - Factor scores. Frequentist and Bayesian approaches. Individual means.

	Principal component analysis		Bayesian approach					
country	Factor 1	Factor 2	Factor 1			Factor 2		
			mean	se	MC error	mean	se	MC error
Argentina	-0.1287	-0.7127	-0.0483	0.1723	0.01010	-0.5809	0.2423	0.01472
Brazil	-0.6129	-1.5035	-0.4560	0.2228	0.01373	-1.3800	0.3329	0.02135
Canada	1.4722	-0.9298	1.6580	0.2378	0.01533	-0.6898	0.3323	0.02173
Chile	-0.1136	0.2510	-0.0825	0.1580	0.00910	0.2675	0.2182	0.01295
China	-1.0703	-1.9820	-1.1660	0.1464	0.00830	-2.0930	0.1979	0.01123
Colombia	-1.3701	-0.2873	-1.3440	0.1614	0.00961	-0.3512	0.2209	0.01292
Costa Rica	-0.7078	1.4275	-0.6945	0.1436	0.00862	1.3710	0.2018	0.01222
Croatia	0.0267	0.8439	-0.0115	0.1034	0.00565	0.8143	0.146	0.00814
Czech Republic	0.4390	0.2062	0.3794	0.1083	0.00552	0.2083	0.1453	0.00739
Ecuador	-1.8247	0.5043	-1.8060	0.1326	0.00668	0.4629	0.1872	0.00980
Estonia	-0.0118	1.3626	0.0926	0.2017	0.01243	1.5200	0.2983	0.01911
Finland	1.3714	0.2312	1.4610	0.1145	0.00661	0.3414	0.1572	0.00913
France	1.1497	-0.6558	1.1280	0.0976	0.00425	-0.6346	0.1263	0.00533
Honduras	-2.2193	1.6295	-2.1580	0.1692	0.00852	1.6210	0.2019	0.00889
Hungary	0.1934	0.1381	0.2155	0.1358	0.00697	0.1798	0.1908	0.01032
India	-1.8461	-2.1038	-2.0000	0.2285	0.01408	-2.2810	0.3207	0.02039
Ireland	1.0935	0.7873	1.1310	0.2111	0.01106	0.7487	0.2778	0.01471
Israel	0.7923	1.1199	0.7196	0.1820	0.01071	1.0030	0.2553	0.01567
Italy	0.8439	-0.3123	0.7836	0.1131	0.00616	-0.3606	0.1630	0.00953
Japan	1.1326	-0.4455	1.0320	0.2015	0.01324	-0.6039	0.2898	0.01912
Korea, Rep.	0.6490	0.4958	0.5193	0.1926	0.01154	0.3283	0.2601	0.01593
Lithuania	-0.3962	0.8064	-0.3176	0.1576	0.00952	0.9335	0.2305	0.01441
Mexico	-0.3478	-0.8777	-0.3285	0.0813	0.00341	-0.8631	0.1101	0.00497
Norway	1.4085	0.4754	1.4480	0.1250	0.00596	0.5581	0.1660	0.00787
Panama	-0.9105	1.4488	-0.8517	0.1216	0.00545	1.461	0.1691	0.00738
Peru	-1.4432	-0.1723	-1.3670	0.1460	0.00810	-0.1757	0.1989	0.01136
Poland	0.1471	-0.3695	0.1254	0.1013	0.00493	-0.3702	0.1347	0.00643
Portugal	0.4164	0.4622	0.3801	0.0868	0.00445	0.4196	0.1219	0.00670
Russian Federation	-0.3812	-1.8095	-0.4054	0.2407	0.01503	-1.7270	0.3304	0.02060
Slovak Republic	0.1432	0.7303	0.09783	0.1050	0.00589	0.7341	0.1520	0.00868
Slovenia	0.5404	1.4112	0.4729	0.1025	0.00547	1.3510	0.1427	0.00811
Spain	0.8518	-0.2356	0.8287	0.1034	0.00489	-0.2185	0.1367	0.00628
Thailand	-0.8136	-0.4822	-0.9077	0.1849	0.01012	-0.5971	0.2627	0.01526
Turkey	-0.7059	-0.0498	-0.7494	0.1411	0.00730	-0.5031	0.1847	0.00917
United Kingdom	1.1183	-0.3123	1.0680	0.1317	0.00754	-0.4130	0.1930	0.01152
United States	1.5269	-1.6476	1.5710	0.1133	0.00635	-1.5520	0.1549	0.00892
Uruguay	-0.4124	1.0065	-0.4166	0.1848	0.01035	1.0700	0.2505	0.01362
Total	-3.78e-09	2.01e-09	1,23E-03			1,20E-03		

mean : posterior mean

s.e : posterior standard error

MC error: Monte Carlo standard error

Table 4 - Factor scores. Frequentist and Bayesian approaches. Time means.

year	P.C.A	Bayesian approach		
		mean	s.e	MC error
1998	-1.9408	-1.3890	0.2521	0.01413
1999	-1.1576	-0.9498	0.1034	0.00525
2000	-0.5411	-0.4411	0.1624	0.00732
2001	-0.3572	-0.4282	0.0685	0.00289
2002	-0.1509	-0.2609	0.0816	0.00361
2003	0.1987	0.0765	0.0739	0.00348
2004	0.5259	0.3962	0.0794	0.00389
2005	0.8273	0.6973	0.0686	0.00330
2006	1.1348	0.9968	0.0724	0.00332
2007	1.4609	1.3020	0.0927	0.00405
Total	-2.98e-09	-2.00e-04		

mean : posterior mean

s.e : posterior standard error

MC error: Monte Carlo standard error

Table 5 - Productivity equation

	Multi-step FAR. Frequentist approach		Bayesian approach		
log(gdp)	coef.	s.e	mean	s.e	MC error
intercept	-0.1098	0.2546	0.1120	0.5688	0.03809
log(gfcf)	0.7484	0.0164	0.7468	0.0193	0.00044
log(tertiary)	0.1120	0.0415	0.1230	0.0446	0.00119
log(energy)	-0.0878	0.0369	-0.1338	0.0562	0.00327
Qi1	0.3141	0.0450	0.3437	0.0588	0.00308
Qi2	0.1429	0.0603	0.2073	0.0902	0.00530
Ft	0.0246	0.0076	0.0287	0.0106	0.00023
σ^2_α	0.1722	0.0209	0.0385	0.0109	0.00022
σ^2_ε	0.0841	0.0032	0.0078	6.166E-4	0.00000

Log-likelihood (zi)		-141,1125
Log-likelihood (zt)		24,5855
Log-likelihood (y)	321.14077	385,5285
Log-likelihood (total)		269,0021

mean : posterior mean

s.e : posterior standard error

MC error: Monte Carlor standard error

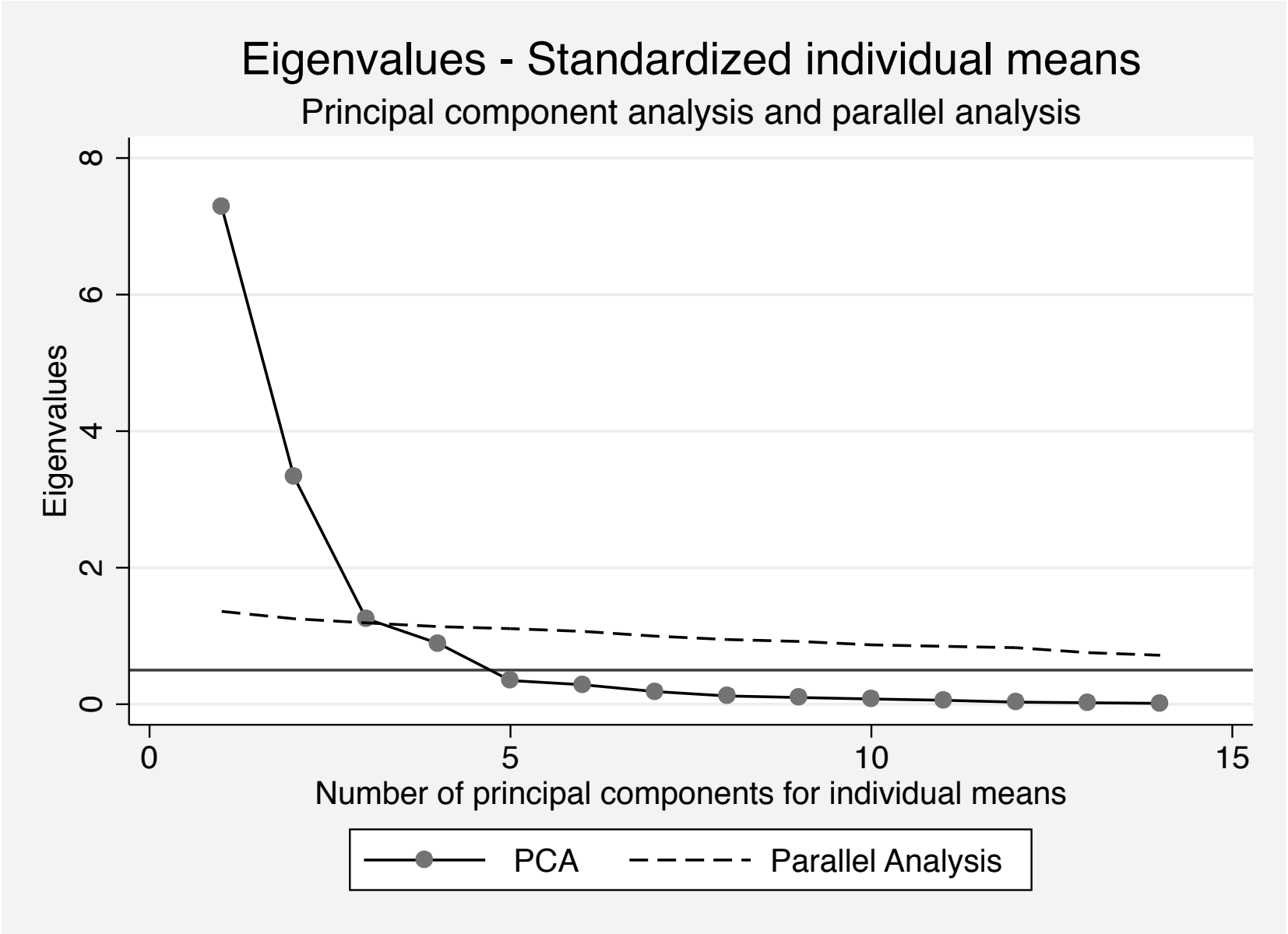

Eigenvalues - Standardized individual means
Principal component analysis and parallel analysis
Eigenvalues
0
2
4
6
8
0
5
10
15
Number of principal components for individual means
PCA
Parallel Analysis

Figure 1

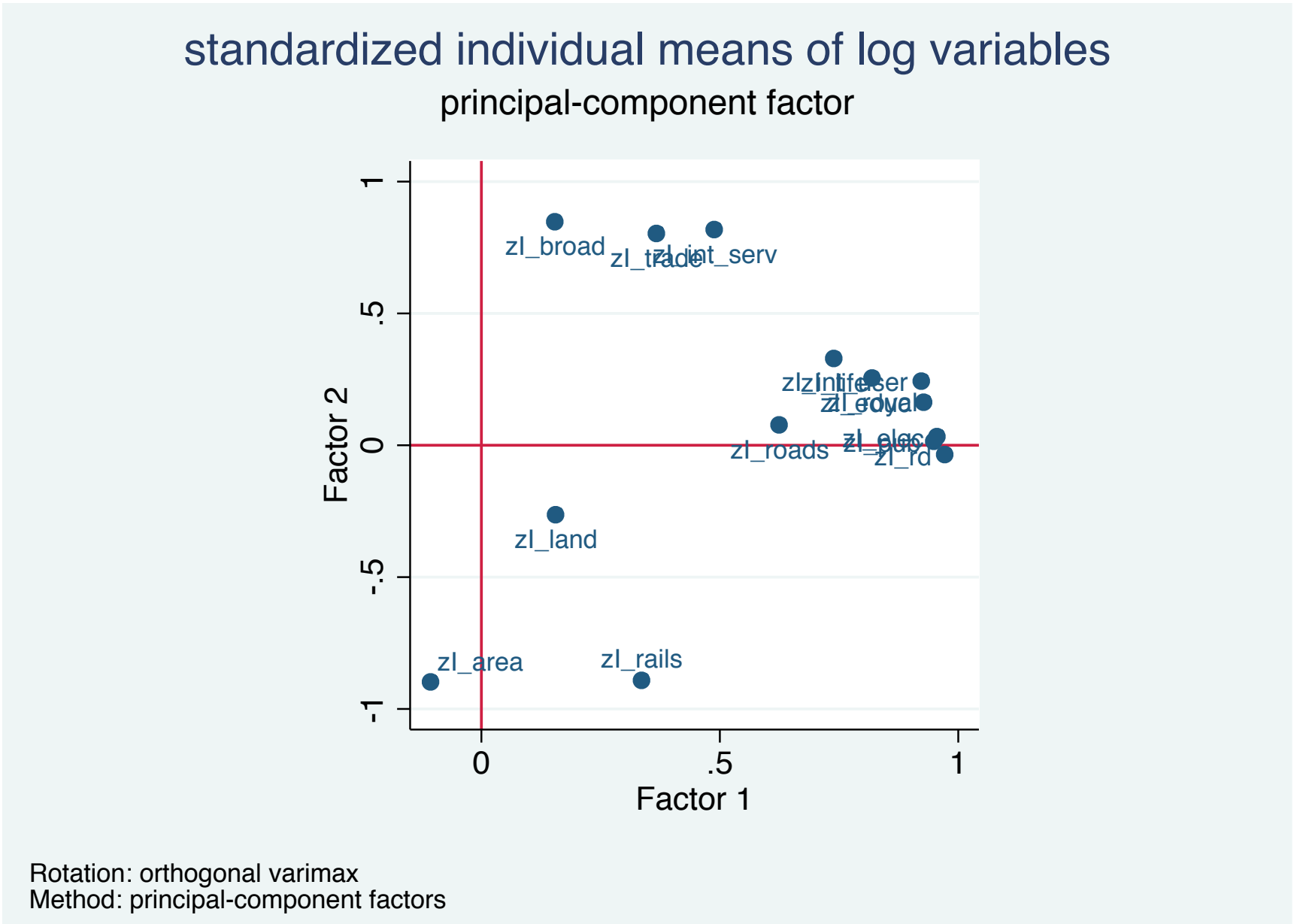

standardized individual means of log variables
principal-component factor
zl_broad
zl_trade
zl_int_serv
zl_land
zl_area
zl_rails
zl_roads
zl_rd
Factor 2
Factor 1
1
.5
0
-.5
-1
0
.5
1
Rotation: orthogonal varimax
Method: principal-component factors

Figure 2

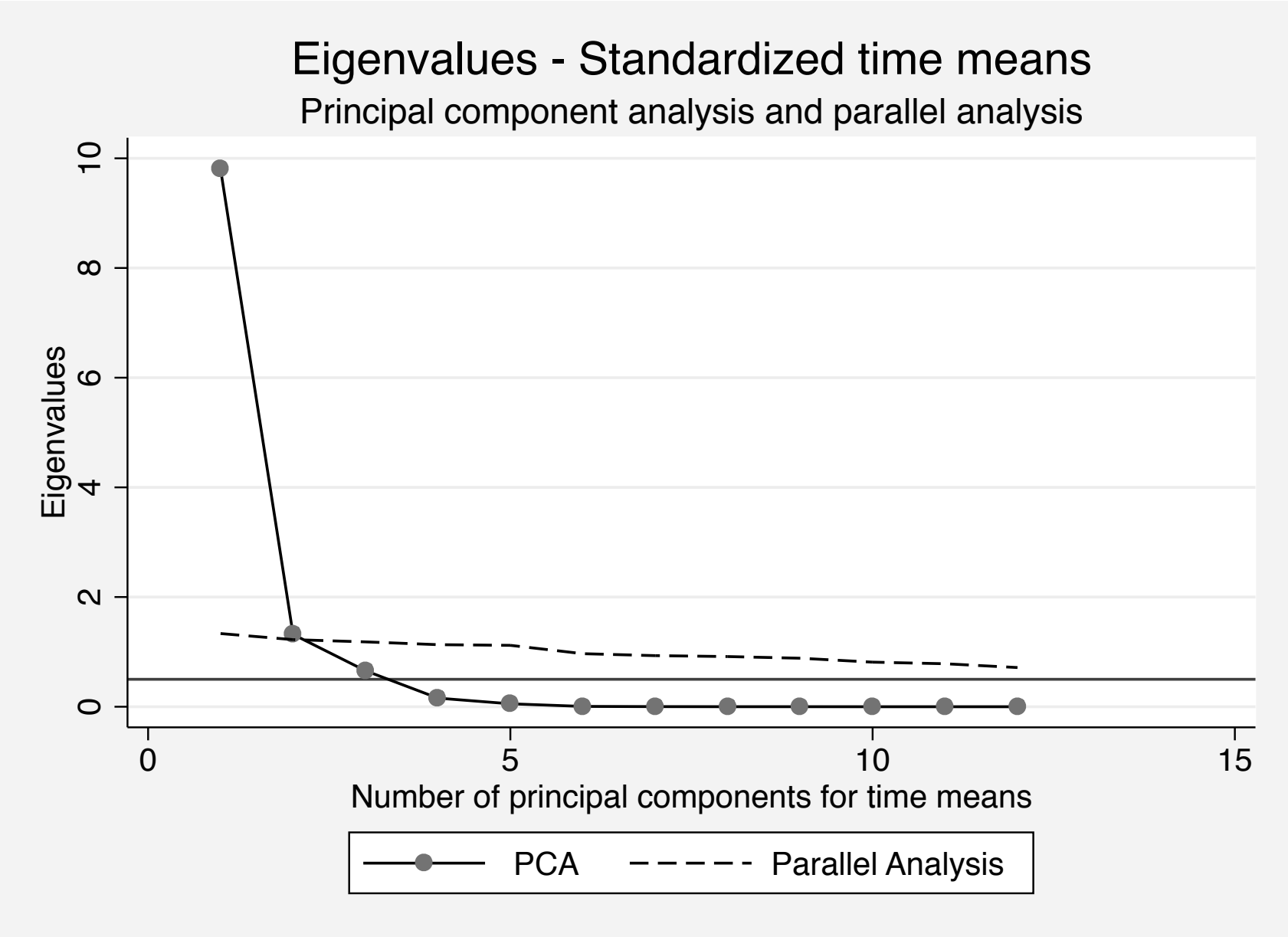

Eigenvalues - Standardized time means
Principal component analysis and parallel analysis
Eigenvalues
0
2
4
6
8
10
0
5
10
15
Number of principal components for time means
PCA
Parallel Analysis

Figure 3

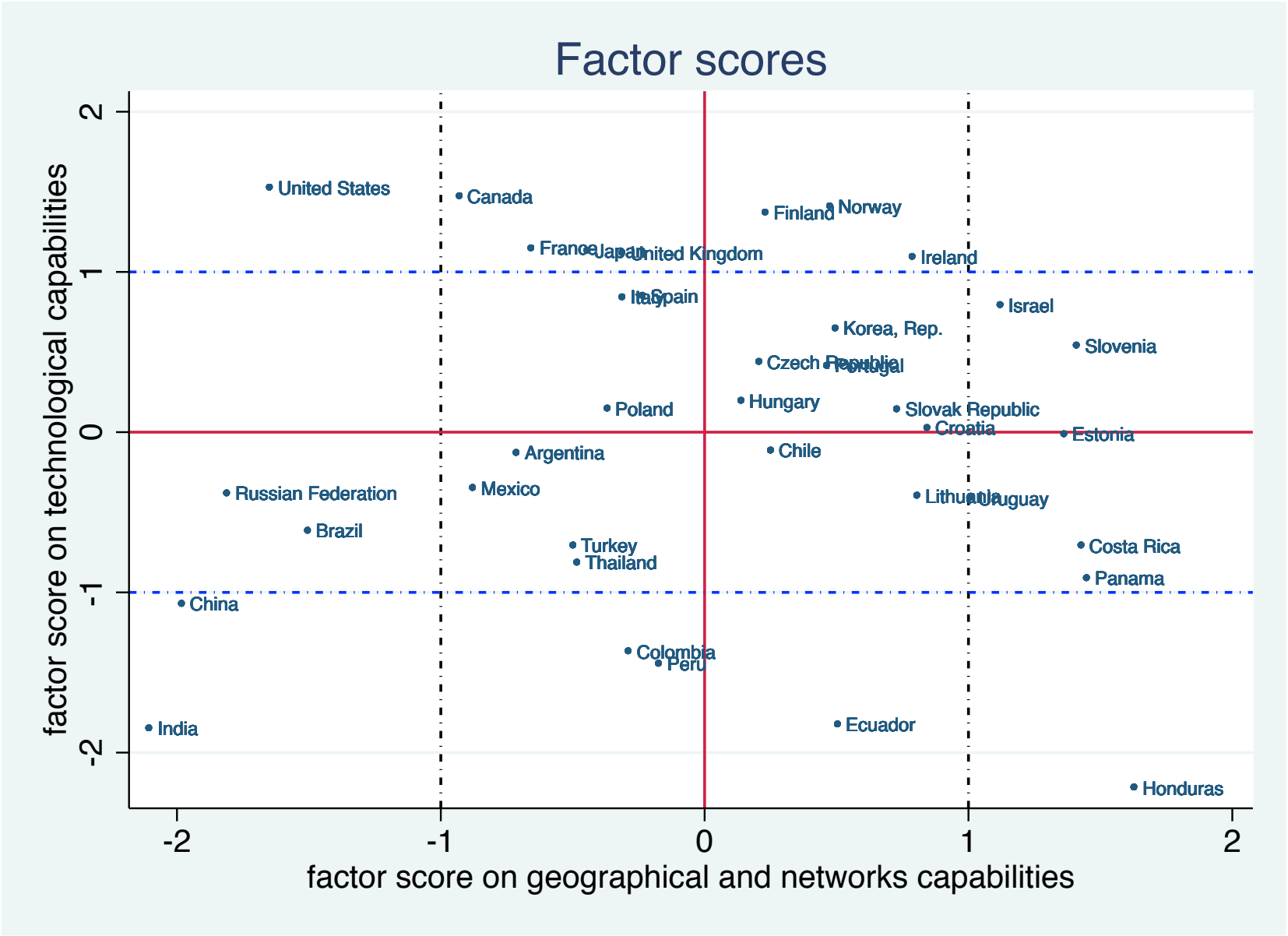

Factor scores
factor score on technological capabilities
factor score on geographical and networks capabilities
2
1
0
-1
-2
United States
Canada
Finland
Norway
France
Japan
United Kingdom
Ireland
Italy
Spain
Israel
Korea, Rep.
Slovenia
Czech Republic
Portugal
Poland
Hungary
Slovak Republic
Croatia
Estonia
Argentina
Chile
Russian Federation
Mexico
Lithuania
Uruguay
Brazil
Turkey
Thailand
Costa Rica
Panama
China
Colombia
Peru
India
Ecuador
Honduras

Figure 4

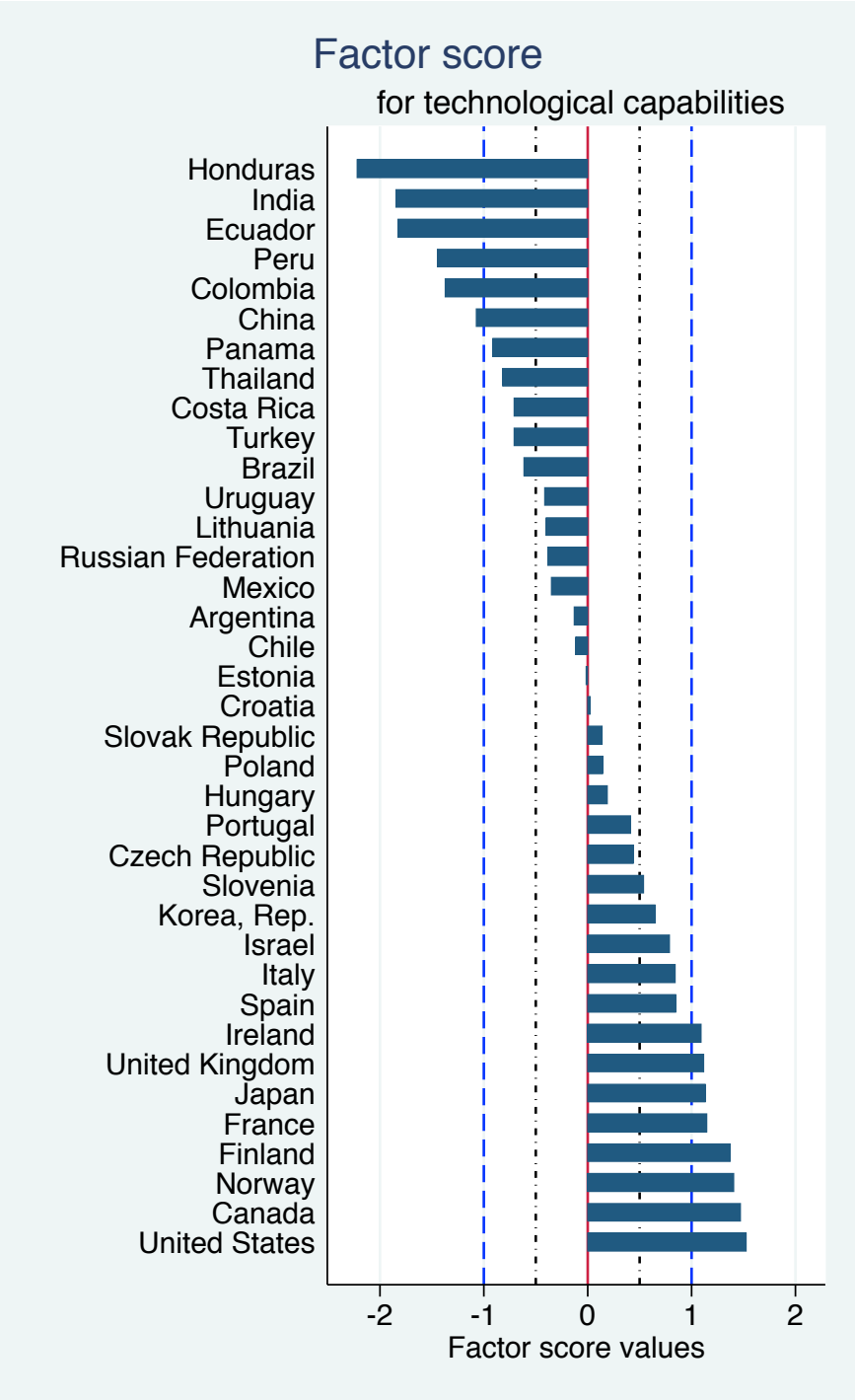
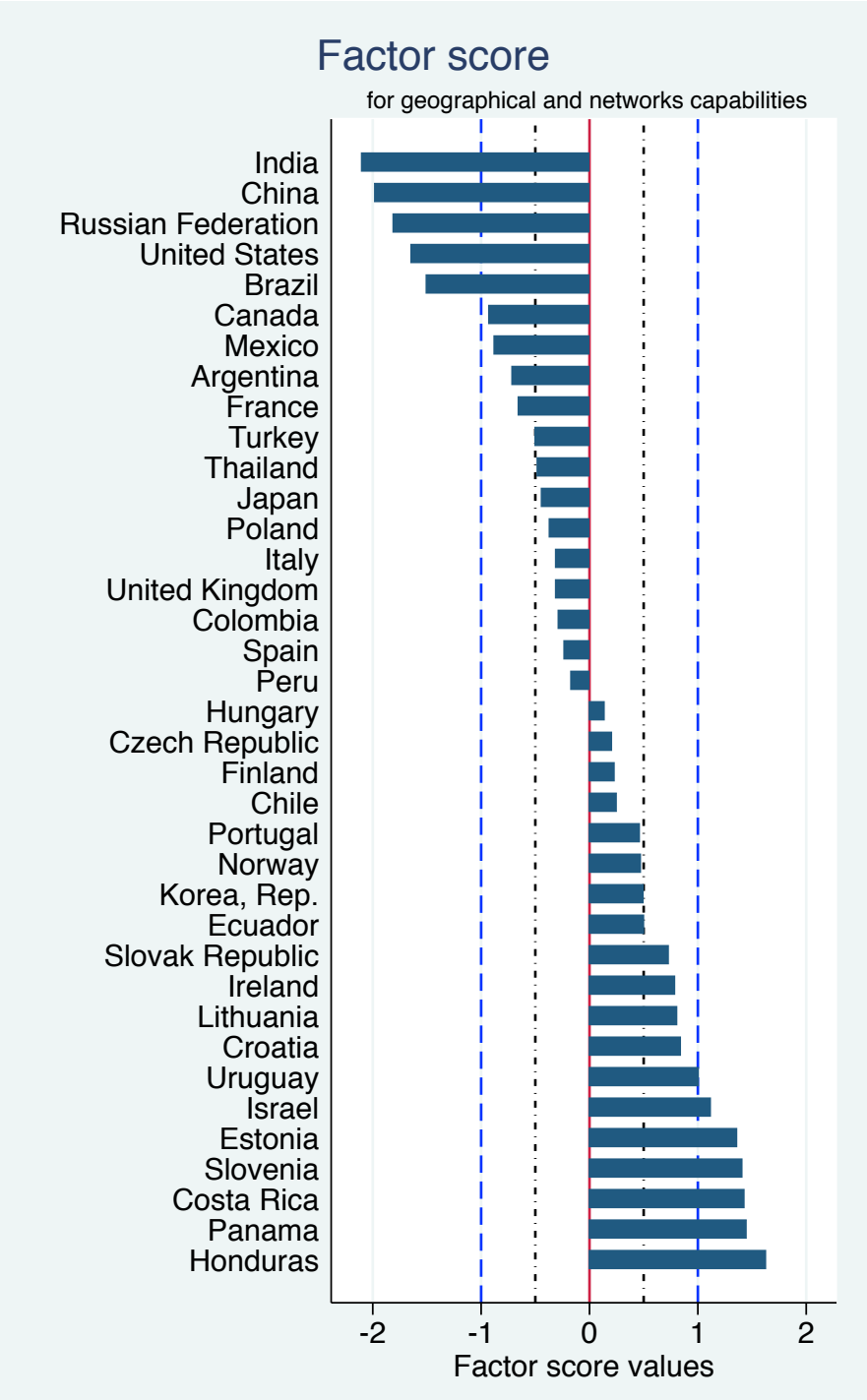

Factor score
for technological capabilities
Honduras
India
Ecuador
Peru
Colombia
China
Panama
Thailand
Costa Rica
Turkey
Brazil
Uruguay
Lithuania
Russian Federation
Mexico
Argentina
Chile
Estonia
Croatia
Slovak Republic
Poland
Hungary
Portugal
Czech Republic
Slovenia
Korea, Rep.
Israel
Italy
Spain
Ireland
United Kingdom
Japan
France
Finland
Norway
Canada
United States
-2
-1
0
1
2
Factor score values

Figure 5


Factor score
for geographical and networks capabilities
India
China
Russian Federation
United States
Brazil
Canada
Mexico
Argentina
France
Turkey
Thailand
Japan
Poland
Italy
United Kingdom
Colombia
Spain
Peru
Hungary
Czech Republic
Finland
Chile
Portugal
Norway
Korea, Rep.
Ecuador
Slovak Republic
Ireland
Lithuania
Croatia
Uruguay
Israel
Estonia
Slovenia
Costa Rica
Panama
Honduras
-2
-1
0
1
2
Factor score values

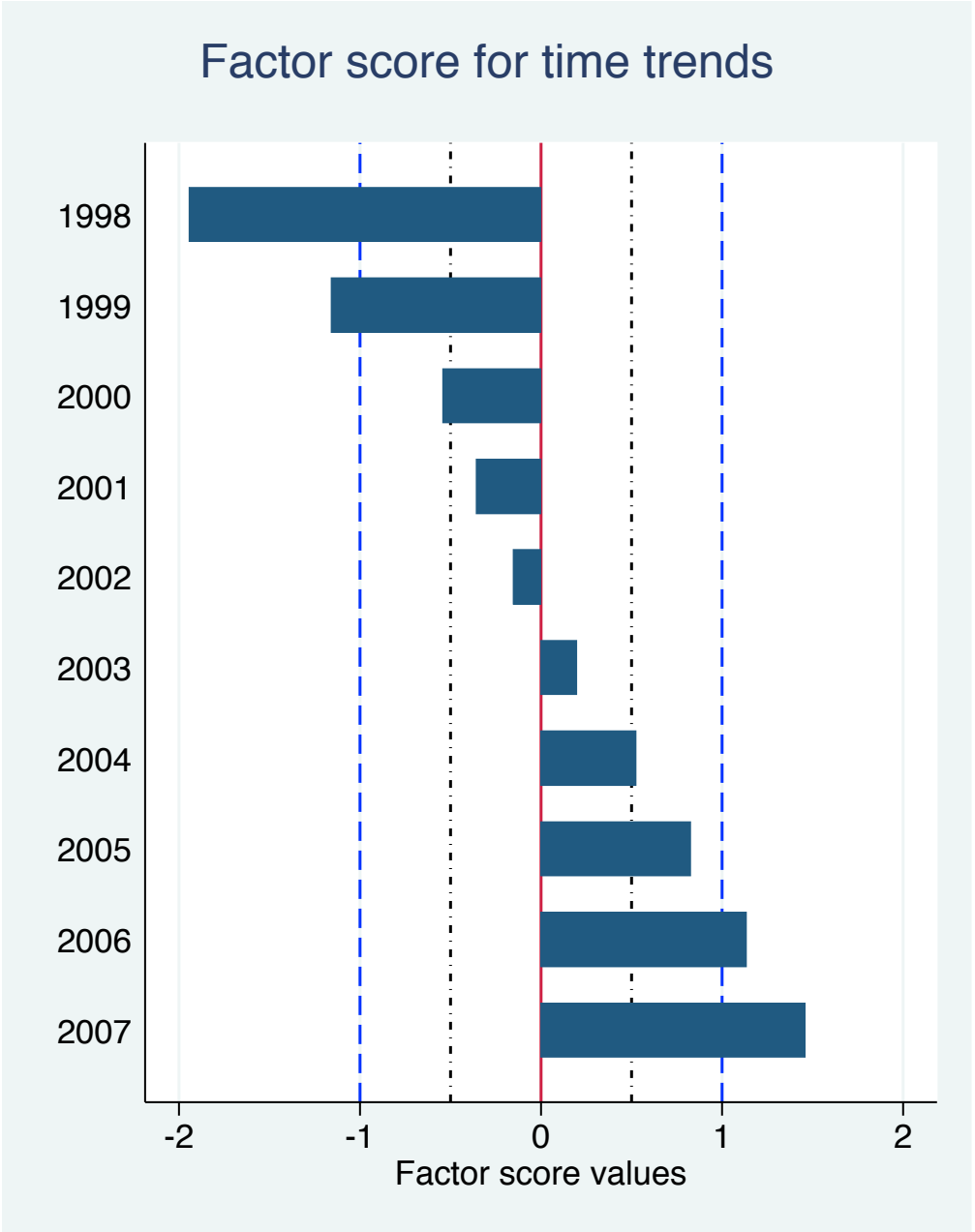

Factor score for time trends
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
-2
-1
0
1
2
Factor score values

Figure 7

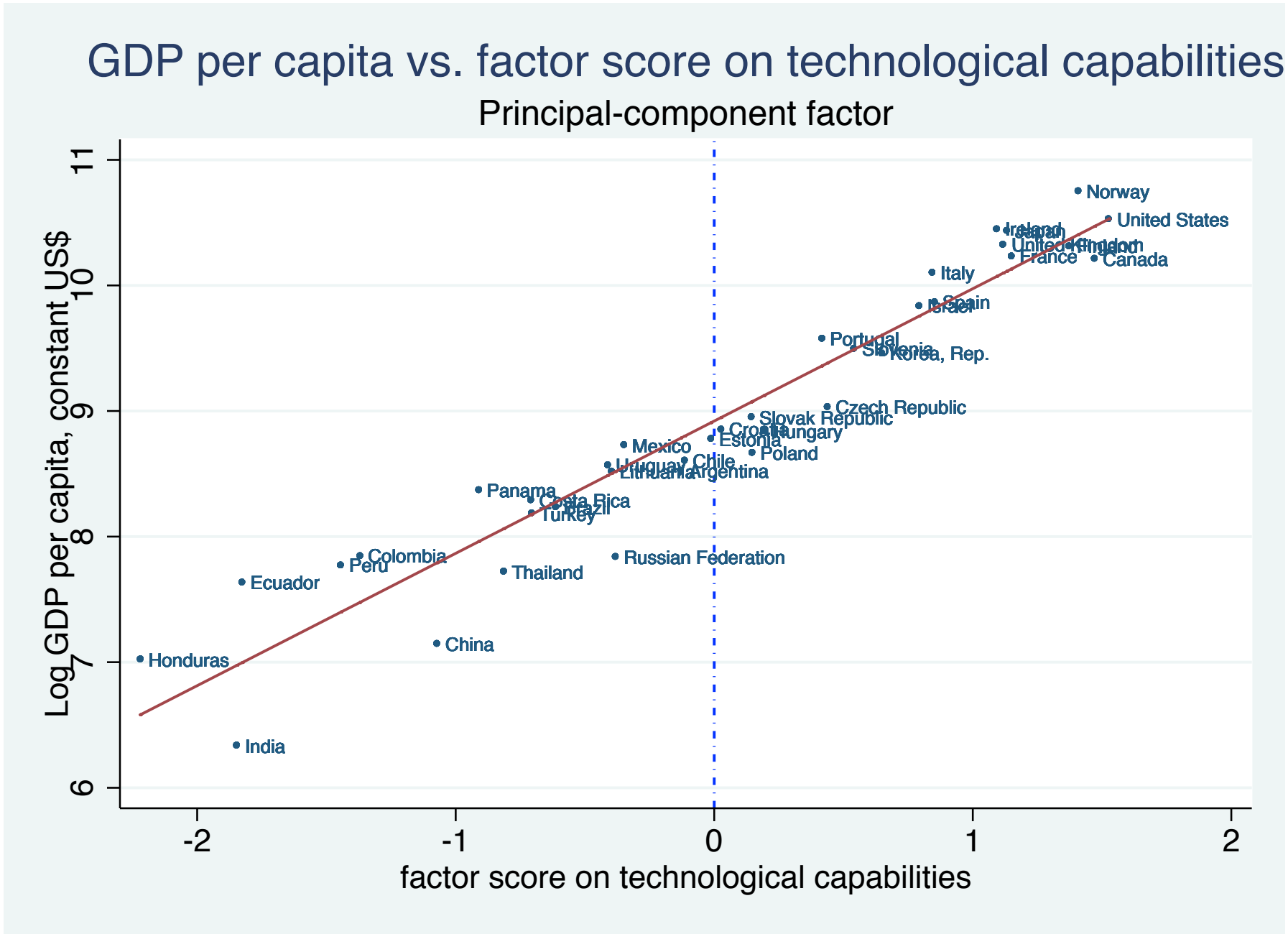

GDP per capita vs. factor score on technological capabilities
Principal-component factor
Log GDP per capita, constant US$
11
10
9
8
7
6
-2
-1
0
1
2
factor score on technological capabilities
Norway
United States
Canada
Italy
Spain
Portugal
Korea, Rep.
Czech Republic
Slovak Republic
Hungary
Estonia
Poland
Chile
Argentina
Mexico
Panama
Costa Rica
Turkey
Brazil
Colombia
Peru
Ecuador
Thailand
Russian Federation
China
Honduras
India

Figure 8

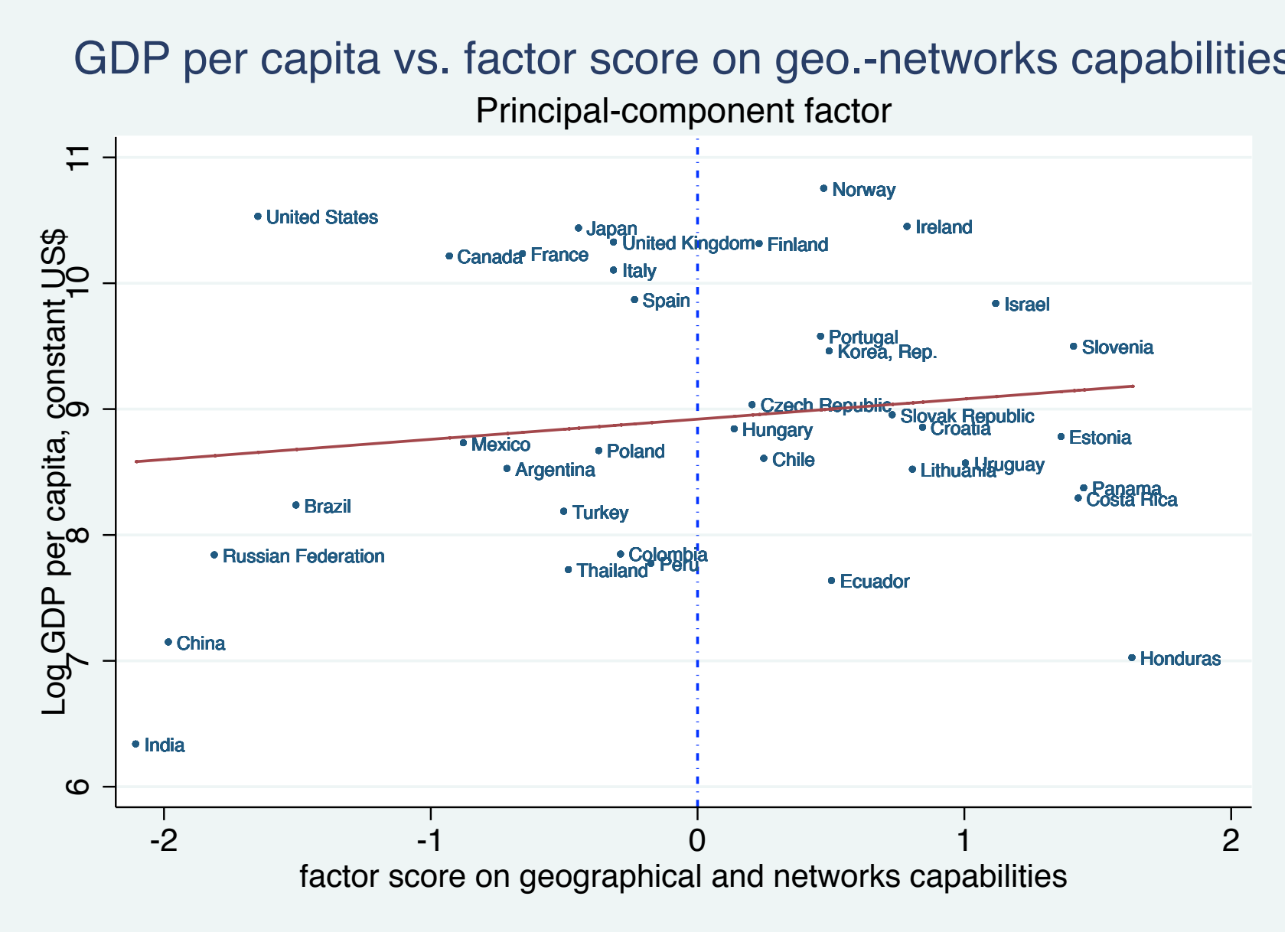

GDP per capita vs. factor score on geo.-networks capabilities
Principal-component factor
Log GDP per capita, constant US$
11
10
9
8
7
6
-2
-1
0
1
2
factor score on geographical and networks capabilities
Norway
United States
Ireland
Japan
United Kingdom
Finland
Canada
France
Italy
Spain
Israel
Portugal
Slovenia
Korea, Rep.
Czech Republic
Slovak Republic
Hungary
Croatia
Estonia
Mexico
Poland
Chile
Argentina
Uruguay
Lithuania
Panama
Costa Rica
Brazil
Turkey
Russian Federation
Colombia
Peru
Thailand
Ecuador
China
Honduras
India

Figure 9

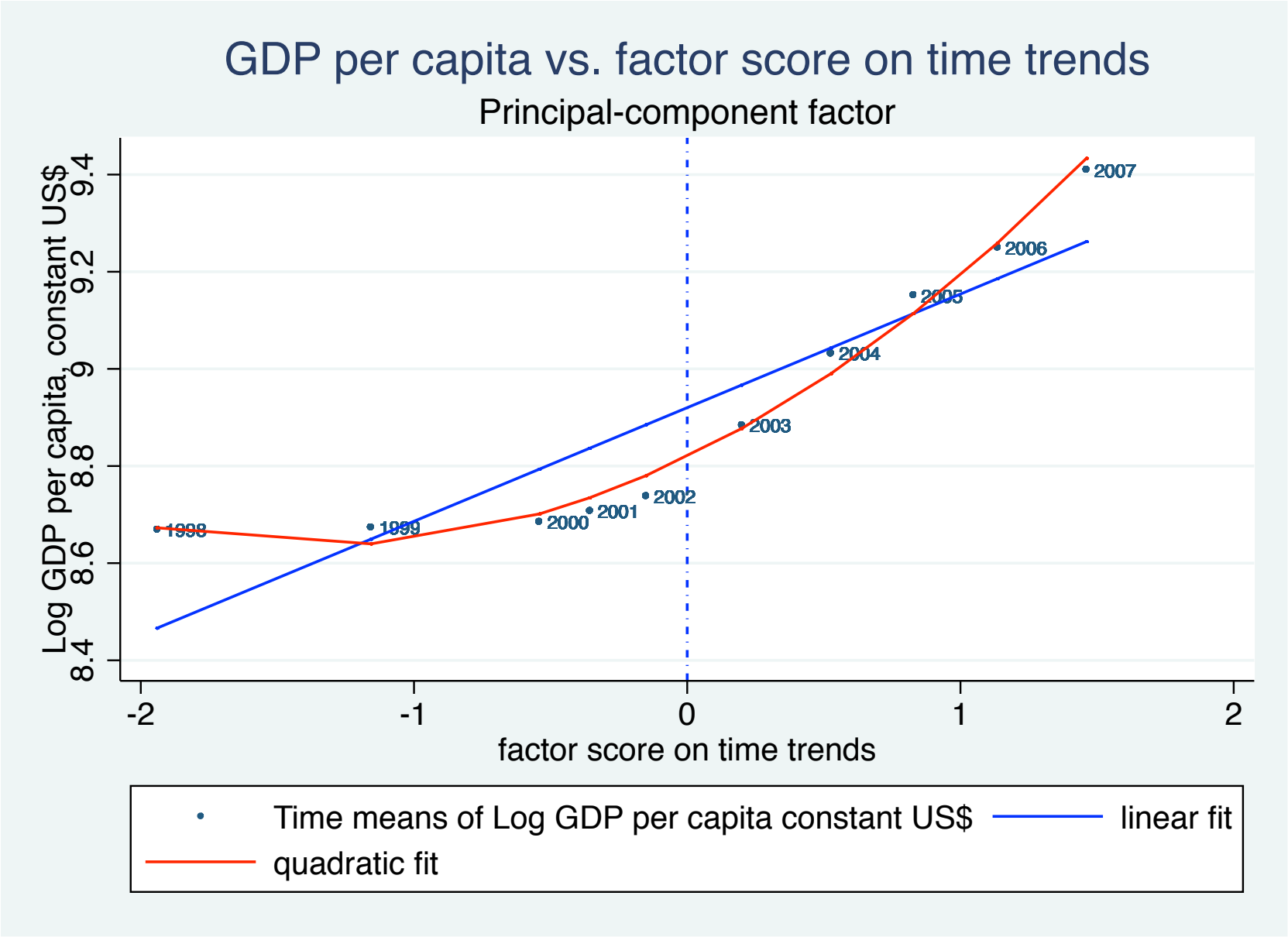

GDP per capita vs. factor score on time trends
Principal-component factor
Log GDP per capita, constant US$
8.4
8.6
8.8
9
9.2
9.4
-2
-1
0
1
2
factor score on time trends
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Time means of Log GDP per capita constant US$
linear fit
quadratic fit

Figure 10

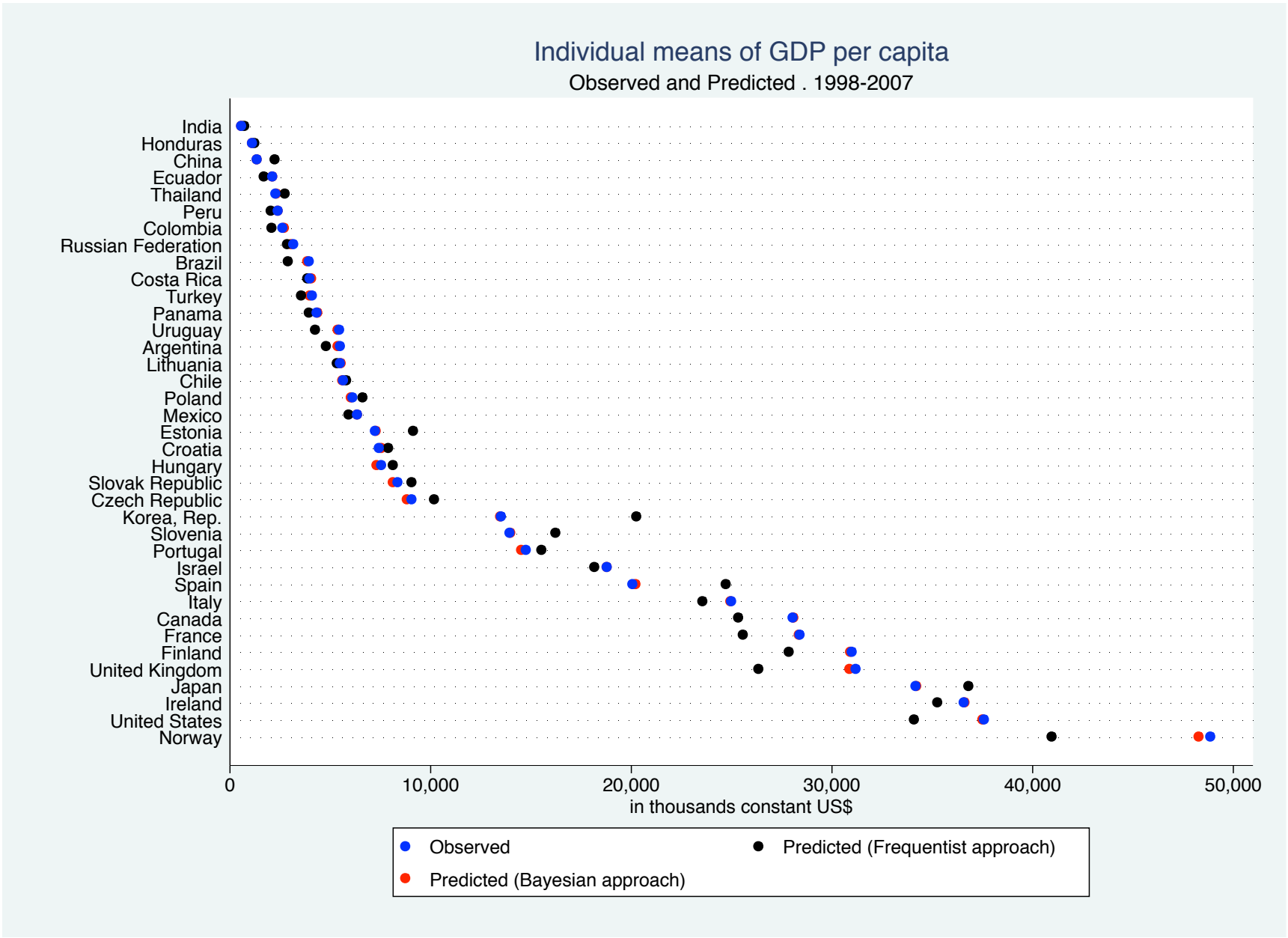

Individual means of GDP per capita
Observed and Predicted . 1998-2007
India
Honduras
China
Ecuador
Thailand
Peru
Colombia
Russian Federation
Brazil
Costa Rica
Turkey
Panama
Uruguay
Argentina
Lithuania
Chile
Poland
Mexico
Estonia
Croatia
Hungary
Slovak Republic
Czech Republic
Korea, Rep.
Slovenia
Portugal
Israel
Spain
Italy
Canada
France
Finland
United Kingdom
Japan
Ireland
United States
Norway
0
10,000
20,000
30,000
40,000
50,000
in thousands constant US$
Observed
Predicted (Frequentist approach)
Predicted (Bayesian approach)

Figure 11

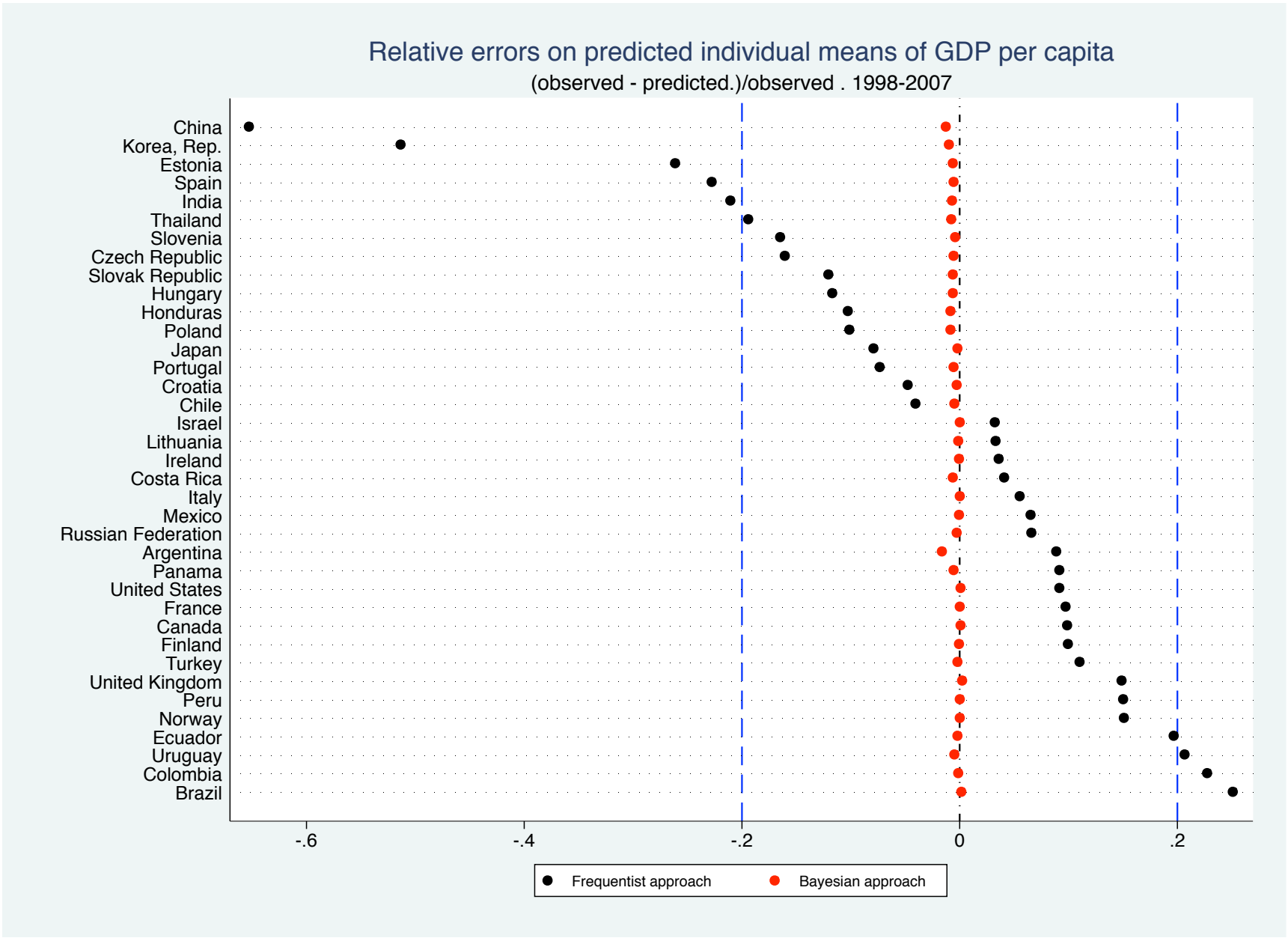

Relative errors on predicted individual means of GDP per capita
(observed - predicted.)/observed . 1998-2007
China
Korea, Rep.
Estonia
Spain
India
Thailand
Slovenia
Czech Republic
Slovak Republic
Hungary
Honduras
Poland
Japan
Portugal
Croatia
Chile
Israel
Lithuania
Ireland
Costa Rica
Italy
Mexico
Russian Federation
Argentina
Panama
United States
France
Canada
Finland
Turkey
United Kingdom
Peru
Norway
Ecuador
Uruguay
Colombia
Brazil
-.6
-.4
-.2
0
.2
Frequentist approach
Bayesian approach

Figure 12

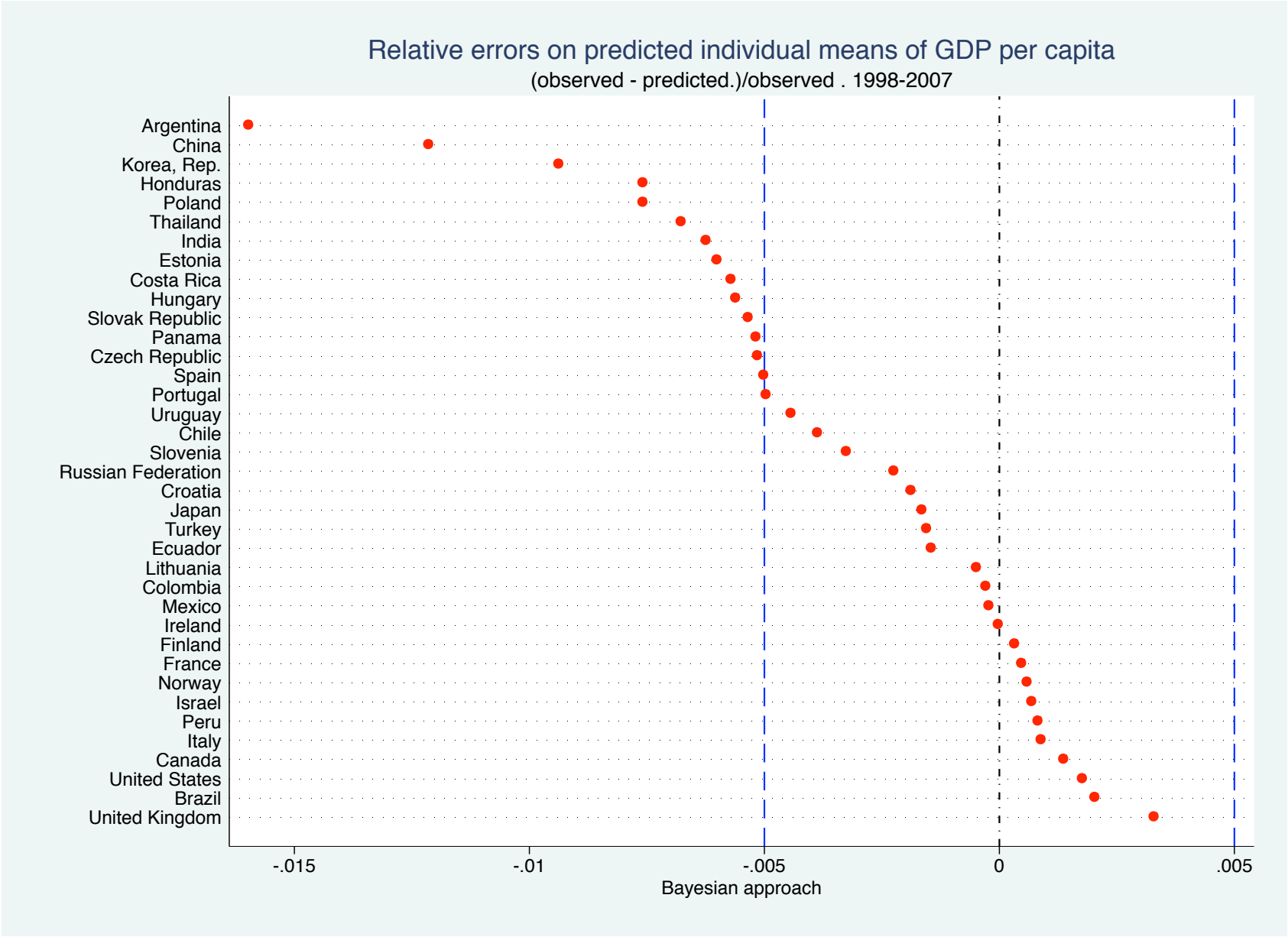

Relative errors on predicted individual means of GDP per capita
(observed - predicted.)/observed . 1998-2007
Argentina
China
Korea, Rep.
Honduras
Poland
Thailand
India
Estonia
Costa Rica
Hungary
Slovak Republic
Panama
Czech Republic
Spain
Portugal
Uruguay
Chile
Slovenia
Russian Federation
Croatia
Japan
Turkey
Ecuador
Lithuania
Colombia
Mexico
Ireland
Finland
France
Norway
Israel
Peru
Italy
Canada
United States
Brazil
United Kingdom
-.015
-.01
-.005
0
.005
Bayesian approach

Figure 13

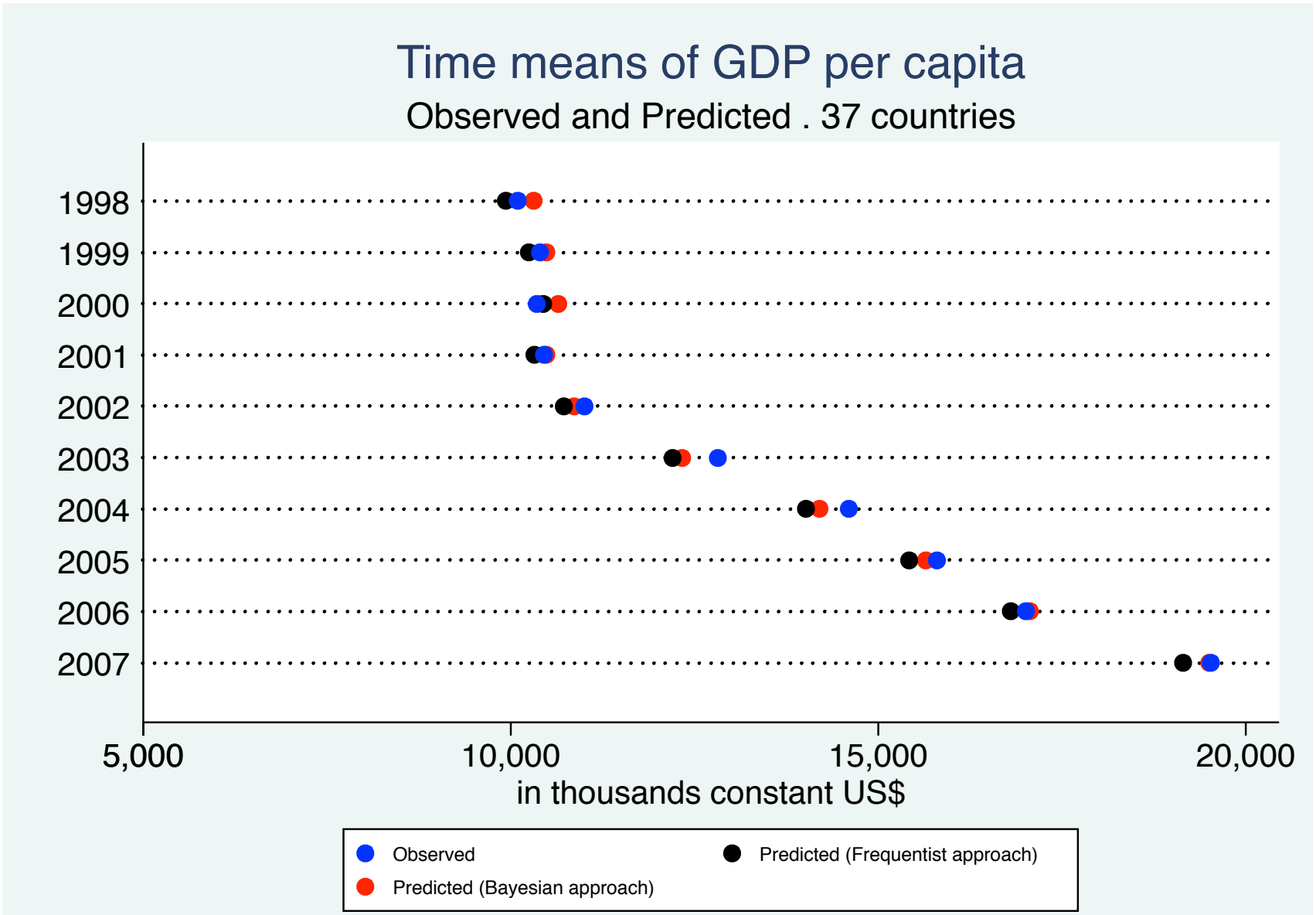

Time means of GDP per capita
Observed and Predicted . 37 countries
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
5,000
10,000
15,000
20,000
in thousands constant US$
Observed
Predicted (Bayesian approach)
Predicted (Frequentist approach)

Figure 14

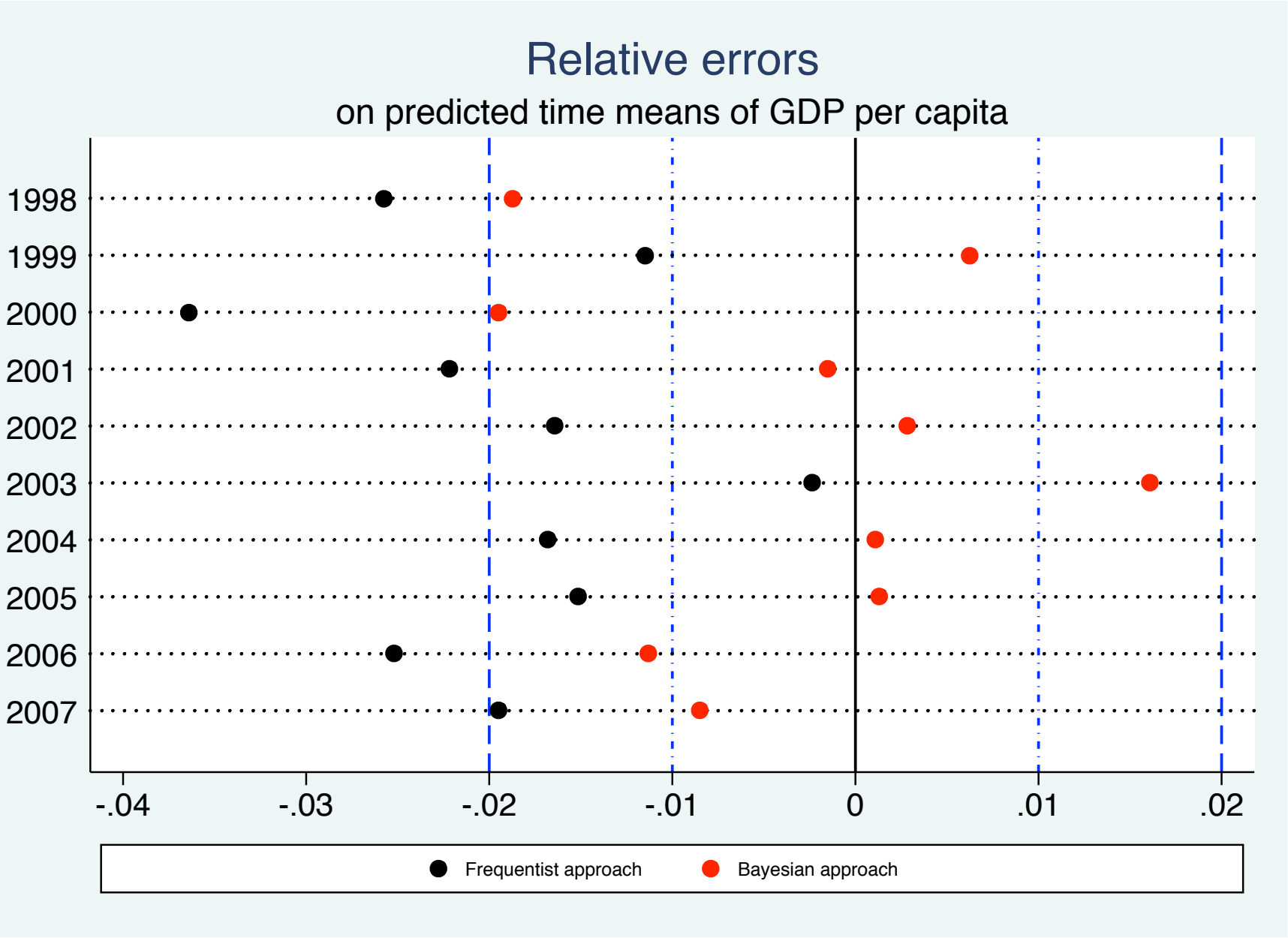

Relative errors
on predicted time means of GDP per capita
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
-.04
-.03
-.02
-.01
0
.01
.02
Frequentist approach
Bayesian approach

Figure 15

TEPP Working Papers 2010

10-1. Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area
Emmanuel Duguet, Noam Leandri, Yannick L'Horty, Pascale Petit

10-2. Couple's Work Hours, Satisfaction and reconciling Work and family Life
Nathalie Georges, Dominique Méda, Danièle Trancart

10-3. Housing ownership, social housing and unemployment: an econometric analysis of the Paris area
Emmanuel Duguet, Yannick L'Horty, Florent Sari

10-4. Do Public Subsidies Have an Impact on New Firm Survival? An Empirical Study with French Data
Lionel Désiage, Richard Duhautois, Dominique Redor

10-5. The effect of social security payroll tax reductions on employment and wages: an evaluation of the 2003 French reform
Matthieu Bunel, Fabrice Gilles, Yannick L'Horty

10-6. What are Entrepreneurs' Objectives When Starting a New Business?
Lionel Désiage

10-7. Health and Early Retirement: Evidence from French Data for individuals
Thomas Barnay, Karine Briard

10-8. Ageing, chronic conditions and the evolution of future drugs expenditures
Thomas Barnay, Sophie Thiébaut, Bruno Ventelou

10-9. Entrepreneurial motives and performance: Why might better educated entrepreneurs be less successful?
Arnab Bhattacharjee, Jean Bonnet, Nicolas Le Pape, Régis Renault

10-10. Returns to firm-provided training in France: Evidence on mobility and wages
Arnaud Chéron, Bénédicte Rouland, François-Charles Wolff

10-11. Taxation of early retirement windows and delaying retirement: the French experience
Pierre-Jean Messe

10-12. Pre Entry Motives into Entrepreneurship and Post Entry Entrepreneurial Orientation
Jean Bonnet, Nicolas Le Pape

10-13. Hiring Practices, Employment Protection and Temporary Jobs
Anne Bucher

10-14. Young-in Old-out: a new evaluation
Michela Bia, Pierre-Jean Messe, Roberto Leombruni

10-15. On the impact of the TFP growth on the employment rate: does training on-the-job matter?
Eva Moreno-Galbis

10-16. The dynamics of youth labor market integration
Anne Bucher

10-17. Fostering the potential endogenous development of European regions: a spatial dynamic panel data analysis of the Cohesion Policy on regional convergence over the period 1980-2005
Salima Bouayad-Agha, Nadine Turpin, Lionel Védrine

10-18. Cost-saving or Cost-enhancing Mergers: the Impact of the Distribution of Roles in Oligopoly
Nicolas Le Pape, Kai Zhao

10-19. Bankruptcy Risk, Product Market Competition and Horizontal Mergers
Bernard Franck, Nicolas Le Pape

10-20. Endogenous Job Destructions and the Distribution of Wages
Arnaud Chéron, Bénédicte Rouland

10-21. Employment Protection Legislation and Adverse Selection at the Labor Market Entry
Anne Bucher, Sébastien Ménard

TEPP Working Papers 2011

11-1. The French "Earned Income Supplement" (RSA) and back-to-work incentives
Denis Anne, Yannick L'Horty

11-2. The effect of place of residence on access to employment: a field experiment on qualified young job applicants in Ile-de-France
Yannick L'Horty, Emmanuel Duguet, Loïc du Parquet, Pascale Petit, Florent Sari

11-3. Why is there a faster return to work near the border?
Jonathan Bougard

11-4. Residential Discrimination and Ethnic Origin: An experimental assessment in the Paris suburbs
Emmanuel Duguet, Yannick L'Horty, Pascale Petit

11-5. The Fateful Triangle : Complementarities between product, process and organisational innovation in the UK and France
Gérard Ballot, Fathi Fakhfakh, Fabrice Galia, and Ammon Salter

The TEPP Institute

The CNRS **Institute for Labor Studies and Public Policies** (the TEPP Institute, FR n°3126 CNRS) gathers together research centres specializing in economics and sociology:

- the **Centre d'Etudes de l'Emploi** (Centre for Employment Studies), **CEE**, Public Administrative Organization, under the leadership of the Ministers of Work and Research
- l'**Equipe de Recherche sur les Marchés, l'Emploi et la Simulation** (Research Team on Markets, Employment and Simulation), **ERMES**, University of Paris II Panthéon-Assas
- the **Centre d'Etudes des Politiques Economiques de l'université d'Evry** (Research Centre focused on the analysis of economic policy and its foundations and implications), **EPEE**, University of Evry Val d'Essonne
- the **Centre Pierre Naville** (Research on Work and Urban Policies), **CPN**, University of Evry Val d'Essonne
- l'**Equipe de Recherche sur l'Utilisation des Données Temporelles en Economie** (Research Team on Use of Time Data in Economics), **ERUDITE**, University of Paris-Est Créteil and University of Paris-Est Marne-la-Vallée
- the **Groupe d'Analyse des Itinéraires et des Niveaux Salariaux** (The Group on Analysis of Wage Levels and Trajectories), **GAINS**, University of the Maine

The TEPP Institute brings together 147 researchers and research professors and 100 PhD students who study changes in work and employment in relation to the choices made by firms and analyse public policies using new evaluation methods.