Syracuse University

SURFACE

Center for Policy Research

Maxwell School of Citizenship and Public Affairs

6-2020

Growth Empirics: A Bayesian Semiparametric Model with Random Coefficients for a Panel of OECD Countries

Badi Baltagi
bbaltagi@maxwell.syr.edu

Georges Bresson
Université Paris II, georges.bresson@u-paris2.fr

Jean-Michel Etienne
Université Paris-Stud 11, jean-michel.etienne@u-psud.fr

CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Growth Empirics: A Bayesian Semiparametric Model with Random Coefficients for a Panel of OECD Countries

Badi H. Baltagi, Georges Bresson, and Jean-Michel Etienne

CENTER FOR POLICY RESEARCH – Summer 2020

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON	Jeffrey Kubik, ECON	Alexander Rothenberg, ECON
Robert Bifulco, PAIA	Yoonseok Lee, ECON	Rebecca Schewe, SOC
Leonard Burman, PAIA	Amy Lutz, SOC	Amy Ellen Schwartz, PAIA/ECON
Carmen Carrión-Flores, ECON	Yingyi Ma, SOC	Ying Shi, PAIA
Alfonso Flores-Lagunes, ECON	Katherine Michelmore, PAIA	Saba Siddiki, PAIA
Sarah Hamersma, PAIA	Jerry Miner, ECON	Perry Singleton, ECON
Madonna Harrington Meyer, SOC	Shannon Monnat, SOC	Yulong Wang, ECON
Colleen Heflin, PAIA	Jan Ondrich, ECON	Peter Wilcoxen, PAIA
William Horrace, ECON	David Popp, PAIA	Maria Zhu, ECON
Yilin Hou, PAIA	Stuart Rosenthal, ECON	
Hugo Jales, ECON	Michah Rothbart, PAIA	

GRADUATE ASSOCIATES

Rhea Acuña, PAIA	Mary Helander, SOC. SCI.	Huong Tran, ECON
Mariah Brennan, SOC. SCI.	Hyoung Kwon, PAIA	Joaquin Urrego, ECON
Ziqiao Chen, PAIA	Mattie Mackenzie-Liu, PAIA	Yao Wang, ECON
Yoon Jung Choi, PAIA	Maeve Maloney, ECON	Yi Yang, ECON
Dahae Choo, ECON	Austin McNeill Brown, SOC. SCI.	Xiaoyan Zhang, Human Dev.
Stephanie Coffey, PAIA	Qasim Mehdi, PAIA	Bo Zheng, PAIA
Giuseppe Germinario, ECON	Claire Pendergrast, SOC	Dongmei Zuo, SOC. SCI.
Myriam Gregoire-Zawilski, PAIA	Krushna Ranaware, SOC	
Jeehee Han, PAIA	Christopher Rick, PAIA	

STAFF

Joseph Boskovski, Manager, Maxwell X Lab	Candi Patterson, Computer Consultant
Katrina Fiacchi, Administrative Specialist	Samantha Trajkovski, Postdoctoral Scholar
Michelle Kincaid, Senior Associate, Maxwell X Lab	Laura Walsh, Administrative Assistant
Emily Minnoe, Administrative Assistant	

Abstract

This paper proposes semiparametric estimation of the relationship between growth rate of GDP per capita, growth rates of physical and human capital, labor as well as other covariates and common trends for a panel of 23 OECD countries observed over the period 1971-2015. The observed differentiated behaviors by country reveal strong heterogeneity. This is the motivation behind using a mixed fixed- and random-coefficients model to estimate this relationship. In particular, this paper uses a semiparametric specification with random intercepts and slopes coefficients. Motivated by Lee and Wand (2016), we estimate a mean field variational Bayes semiparametric model with random coefficients for this panel of countries. Results reveal nonparametric specifications for the common trends. The use of this flexible methodology may enrich the empirical growth literature underlining a large diversity of responses across variables and countries.

JEL No.: $C11, C14, C23, O47$

Keywords: GDP Per Capita, Growth Empirics, Mean Field Variational Bayes Approximation, Panel Data, Random Coefficients, Semiparametric Model

Authors: Badi H. Baltagi, Center for Policy Research and Department of Economics, Maxwell School, Syracuse University, bbaltagi@maxwell.syr.edu; Georges Bresson, Department of Economics, Université Paris II, France, georges.bresson@u-paris2.fr; Jean-Michel Etienne, Department of Economics, Université Paris-Stud 11, jean-michel.etienne@u-psud.fr

1. Introduction

There is a huge theoretical and empirical literature studying cross-country growth differentials. What explains differences in productivity and how does it enhance economic development? These are fundamental questions that continue to preoccupy many 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 both the level of Gross Domestic Product (GDP) per capita and the growth rate of GDP per capita. The term "Total Factor Productivity (TFP)" is used to refer to the "something else" (besides physical factor accumulation) that accounts for economic growth differences. This TFP, also known as the Solow residual, has been viewed by Abramowitz (1956) as a "measure of our ignorance". Different theories have provided alternative conceptions of TFP: variations in technology, externalities, changes in the sector composition of production, adoption of lower cost production methods, scale economies, variations in capacity utilization, institutions to mention a few. Besides the traditional inputs (labor and physical capital) in a Cobb-Douglas aggregate production function, a lot of other variables (technology, infrastructure, institutions, etc.) are often advanced to explain the evolution of the growth rate of GDP per capita and the country gaps in productivity. However, a vast majority of this empirical literature limit cross-country heterogeneity in production technology to the specification of total factor productivity.[1] Empirical evidence of the relative importance of each of these explanations is far from unanimous. There is even controversy between those like Mankiw et al. (1992) who argue that growth is mainly driven by factor accumulation and those like Easterly and Levine (2001), who argue that there is something else besides factor accumulation that explains differences in economic growth. The latter research recommends that economists devote more effort towards modeling and quantifying total factor productivity.

Besides the usual factors, like physical capital and labor, researchers include several indicators and estimate the model using standard panel data estimators. More variables lead to the curse of dimensionality especially with nonparametric estimation. As emphasized by Moral-Benito (2012), considerable effort in the last decade in the empirical growth literature has been spent on selecting appropriate variables to include in linear growth regressions. Using Bayesian model averaging (BMA) techniques, these authors select variables proposed as growth determinants from a total of more than 140 variables (see 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 MCMC methods, select the "best" model. One major criticism of BMA is that some important variables such as capital stock, R&D, education, etc., are eliminated in favor of other variables such as life expectancy, cultural or religion variables.

Another line of research use factor-augmented regressions (FAR) to avoid these problems. They focus attention on a small number of variables of interest such as capital and labor and consider the

---

[1] A majority of the growth empirics literature on aggregate data focuses on versions of the "convergence regression equation" (Barro and Sala-i Martin (1992, 1997, 2004), Mankiw et al. (1992), Islam (1995) to mention a few.)

other covariates merely as controls (see Anderson et al. (2006), Pesaran (2006), Bai and Ng (2008), Bai (2009), Eberhardt and Bond (2009), Eberhardt and Teal (2011), Kneip et al. (2012), Bresson et al. (2016) to mention a few).

Many empirical growth models in the applied literature assume common technology parameters. As emphasized by Eberhardt and Teal (2011), this homogeneous specification is not likely to hold across countries. Moreover, adding non-standard growth determinants is not sufficient. Some have explicitly introduced heterogeneity using the mean group and pooled mean group estimators proposed by Pesaran and Smith (1995) and Pesaran et al. (1999) (see also Bassanini and Scarpetta (2001), Bassanini et al. (2001)). Durlauf et al. (2001) explain that theoretical justifications for heterogeneous technology parameters can be found in the "new growth" literature which argues that production functions differ across countries and attempt to find the sources of this heterogeneity. As underlined by Brock and Durlauf (2001) (pp. 8-9) "*the assumption of parameter homogeneity seems particularly inappropriate when one is studying complex heterogeneous objects such as countries*".

Following the general ideas of Hsiao et al. (1999), Pesaran et al. (1999), Brock and Durlauf (2001), Durlauf et al. (2001), Hsiao and Pesaran (2008), Eberhardt and Teal (2011), Hsiao (2015), among others, we specify and estimate a random coefficient model of an aggregate Cobb-Douglas function including a common and country-specific TFP structure. Unlike the common factor approach in which the non observed TFP is modelled as the combination of a set a common factors and a vector of factor loadings, we use a semiparametric specification of these common factors. In our particular case, this set of common factors comes down to a trend (as in the Zhang et al. (2012)'s approach).

More specifically, we estimate a Bayesian semiparametric growth model of GDP per capita using growth rates of capital intensity, human capital and other control covariates such as R&D intensity, patent and trademark applications, government expenditures, trade, etc. In order to take into account the heterogeneity among countries, we allow random intercept and slope coefficients as well as a semiparametric function of time summarizing the dynamics of common factors. Inspired by the recent work of Lee and Wand (2016), we use a mean field variational Bayesian approach which has great advantages as compared to the Markov Chain Monte Carlo (MCMC) technique such as Gibbs sampling. In section 2, we briefly present the specification and the estimation method. In section 3, we describe our panel data using 23 OECD countries observed over the period 1971-2015. In Section 4, the estimation results are discussed. Finally, Section 5 concludes.

2. Growth empirics: the model and the estimation method

2.1. The model

Using a simplified setup, the production function takes the form:

$$Y_i(t) = F\left[K_i(t), A_i(t), L_i(t), H_i(t)\right], i = 1, ..., N, t = 1, ..., T \qquad (1)$$

where $Y_i(t)$ denotes output produced in country i at time t. We assume a one-sector production technology in which output is a homogeneous good that can be consumed, or invested to create new units of physical capital $K_i(t)$ and new units of human capital stock $H_i(t)$. $L_i(t)$ denotes labor, $A_i(t)$

is the level of economic and technological efficiency, the so-called TFP. We assume a Cobb-Douglas production function, so the production in country i at time t is given by:

$$Y_i(t) = F\left[K_i(t), A_i(t), L_i(t), H_i(t)\right] = K_i^{\alpha_i}(t) H_i^{\gamma_i}(t) \left[A_i(t)L_i(t)\right]^{\beta_i} \tag{2}$$

with decreasing returns to all capital ($\alpha_i + \gamma_i < 1$) and where $A_i(t)L_i(t)$ is the number of effective units of labor (*labor augmenting technical progress*). We assume constant returns to scale[2]: $\beta_i = 1 - (\alpha_i + \gamma_i)$ and we allow the elasticities with respect to capital and labor to vary by country. This is a heterogeneous by country model (as in Pesaran et al. (1999), Bassanini and Scarpetta (2001), Bassanini et al. (2001), Eberhardt and Teal (2011) to mention a few).

The level of economic and technological efficiency $A_i(t)$ is assumed to be the product of two components: economic efficiency (or innovation efficiency) $A_{1i}(t)$ depending upon R&D policies, economic policies, educational and health policies, institutions, openness, etc., and the level of technological progress $A_{2i}(t)$ (see amongst others, Cellini (1997), Bassanini et al. (2001) for a similar formulation). In turn, $A_{1i}(t)$ can be written as a product of policy variables $V_{ij}(t)$, while $A_{2i}(t)$ is assumed to grow at the rate $f_i(t)$.

$$A_i(t) = A_{1i}(t) A_{2i}(t) \text{ with } A_{1i}(t) = \prod_{j=1}^{J} V_{ij}^{\omega_{ij}}(t) \text{ and } A_{2i}(t) = A_{2i}(0)e^{f_i(t)} \tag{3}$$

Given the rapid technological changes due to the 3rd industrial revolution (telecommunications, the Internet, etc.), it is unrealistic to assume a constant rate of technological growth over time.[3,4] In this case, we can write:

$$\begin{aligned}
\log\left(\frac{Y_i(t)}{L_i(t)}\right) = {} & \alpha_i \log\left(\frac{K_i(t)}{L_i(t)}\right) + \gamma_i \log\left(\frac{H_i(t)}{L_i(t)}\right) + \left[1 - (\alpha_i + \gamma_i)\right] \sum_{j=1}^{J} \omega_{ij} \log V_{ij}(t) \\
& + \left[1 - (\alpha_i + \gamma_i)\right] \log A_{2i}(0) + \left[1 - (\alpha_i + \gamma_i)\right] f_i(t)
\end{aligned} \tag{4}$$

where $f_i(t)$ is a smooth function defined as:

$$f_i(t) = g_i(t) + \sum_{l=1}^{q_l} \eta_{il} W_{il}(t) \tag{5}$$

$\{W_l, 1 \leq l \leq q_l\}$ are spline bases of size q_l and $\{\eta_l, 1 \leq l \leq q_l\}$ are the spline coefficients. $g_i(t)$ is the parametric part of the smooth function which could be an empty set, a linear or a nonlinear function of the time trend. The specification (4)-(5) is a semiparametric model with varying parameters. As the main variables are non stationary (see section 4), the discrete version of the estimated model will be written in growth rate as follows:

$$\begin{aligned}
\Delta \log\left(\tfrac{Y}{L}\right)_{it} = {} & \alpha_i \Delta \log\left(\tfrac{K}{L}\right)_{it} + \gamma_i \Delta \log\left(\tfrac{H}{L}\right)_{it} + \sum_{j=1}^{J} \omega_{ij}^* \Delta \log V_{ijt} + \Delta f_i^*(t) \\
& \text{with } \Delta f_i^*(t) = \Delta g_i^*(t) + \sum_{l=1}^{q_l} \eta_{il}^* \Delta W_{il}(t) \text{ and } \Delta x_{it} = x_{it} - x_{i,t-1}
\end{aligned} \tag{6}$$

---

[2]If $\alpha_i + \gamma_i = 1$, there is no steady state for this model.

[3]In the standard Solow framework, Mankiw et al. (1992) assumes that $A_i(t) = A_{2i}(0)e^{f_i(t)} = A_{2i}(0)e^{gt}$, $\forall i$. Islam (1995) and Pesaran et al. (1999) allow for heterogeneity by introducing indexes for each country: $A_i(t) = A_{2i}(0)e^{f_i(t)} = A_0(t)e^{g_i t}$ where g_i is a country-specific constant.

[4]The common factor literature assumes that: $A_{2i}(t) = A_{2i}(0)e^{\lambda_i' f(t)}$ where $f(t)$ is set a common factors and λ_i is a vector of factor loadings.

where all the coefficients with a superscript $(*)$ are the initial coefficients multiplied by $[1 - (\alpha_i + \gamma_i)]$. The estimate of TFP growth[5] is then based on the standard discrete growth equation $\Delta \log A_{it} = \Delta \log \left(\frac{Y}{L}\right)_{it} - \alpha_i \Delta \log \left(\frac{K}{L}\right)_{it} - \gamma_i \Delta \log \left(\frac{H}{L}\right)_{it}$. In a similar way, the technical innovation change (or economic efficiency change) is given by $\Delta \log A_{1,it} = \sum_{j=1}^{J} \omega_{ij}^* \Delta \log V_{ijt}$ and the technical efficiency change (or the technological efficiency change) is given by $\Delta \log A_{2,it} = \Delta g_i^*(t) + \sum_{l=1}^{q_l} \eta_{il}^* \Delta W_{il}(t)$. This semiparametric model is introduced to describe the nonlinear trending phenomenon observed in cross-country growth models.

The varying coefficient model considered here is therefore part of the already long tradition of research on semiparametric estimation of partially linear varying coefficient panel data models[6] which allow flexibility to characterize trending phenomenon in nonlinear panel data analysis. Some use semiparametric profile likelihood methods (Chen et al. (2012), Li *et al.* (2013), kernel or averaged local linear estimation (Li et *et al.* (2011)). In the spirit of works on semiparametric partially linear model using series estimation methods (Huang et al. (2002), Fan and Li (2004), Qu and Li (2006), Fan et al. (2007), An et al. (2016) for instance), we use a Bayesian linear-mixed Gaussian model-based penalized spline specification with random coefficients proposed by Lee and Wand (2016). This approach is part of a series of Bayesian works that has been developing in recent years, such as the work of Park et al. (2015), Jeong and Park (2016) or Huang and Lu (2017) all of which use the Markov Chain Monte Carlo (MCMC) method. Unfortunately, the MCMC techniques such as Gibbs sampling could become computationally prohibitive and may suffer from poor mixing and do not scale well when applied to models that require inversion of large sparse covariance matrices (as in our case). But Lee and Wand (2016)'s approach, which uses a mean field variational Bayes approximation, has many advantages over MCMC.

---

[5]The literature on regression-based estimation of TFP growth is deep and extensive. There are a lot of contributions to the efficiency and productivity literature that offer different ways to estimate the canonical panel data model (production function, cost function or stochastic production frontier) to decompose TFP growth into an innovation component and a time trend technological component (see for instance Cornwell et al. (1990), Kumbhakar (1990), Baltagi and Griffin (1988), Coelli et al. (1998), Park et al. (1998, 2003, 2007), Kneip et al. (2012), Sickles et al. (2015), Duygun et al. (2017) and Isaksson et al. (2018), to mention a few).

[6]We follow Hsiao's volumunus work on random coefficient models, by applying a semiparametric model with random coefficients. Noteworthy are Cheng Hsiao's work on serial correlation tests in semiparametric panel data models (Li and Hsiao (1998)), semiparametric models based on series estimation methods (Li et al. (2003)), semiparametric nonlinear errors-in-variables models (Wang and Hsiao (2011)), semiparametric analysis of non-cointegrated process (Sun et al. (2011, 2015a)), semiparametric estimation of varying coefficient models (Li et al. (2015), An et al. (2016), Li et al. (2017)) to mention a few. Also, the work of Robinson (1988), Li and Stengos (1996), Li and Ullah (1998), Fan and Huang (2005), You et al. (2010) on semiparametric panel data models with random effects, to mention a few. Baltagi and Li (2002), Su and Ullah (2006), Sun et al. (2009), Zhang et al. (2011), Qian and Wang (2012) and Boneva et al. (2015), for semiparametric panel data models with fixed effects, to mention a few. The literature has also focused on estimating (time) varying coefficient trend functions using nonparametric estimation methods (see for instance Robinson (1989, 2012), Gao and Hawthorne (2006), Cai (2007), Chen et al. (2012), Zhang et al. (2012)) or on nonparametric models with fixed effects (Li et al. (2011), Lee and Robinson (2015)). The literature on the subject is important and is growing rapidly (see the surveys of Li and Racine (2007), Ai and Li (2008), Su and Ullah (2011), Chen et al. (2013), Sun et al. (2015b), Rodriguez-Poo and Soberon (2017) and also the special issue of the Journal of Econometrics, edited by Li and Li (2015), to mention a few).

The use of semiparametric methods for growth regressions is not recent. For instance, Liu and Stengos (1999) find non-linearities in cross-country growth regressions. Kumar and Ullah (2000) use a semiparametric varying parameter panel data models for an application to the estimation of speed of convergence. Park et al. (1998, 2003, 2007) considered linear stochastic frontier panel models in which the distribution of country-specific technical efficiency effects is estimated nonparametrically. They estimate robust standard errors for semi-nonparametric models based on adaptive estimation techniques for semiparametric efficient estimators. Chambers and Krause (2010) study the relationship between income inequality and economic growth using a semiparametric dynamic panel data model with fixed effects. Liu et al. (2017) use Bayesian treatments for panel data stochastic frontier models with time varying heterogeneity. But Bayesian semiparametric models with random coefficients based on series estimation methods could be more flexible and powerful.

In this paper, we do not focus on the conditional convergence path, defined as a partial adjustment process, the so-called Barro regressions. This error correction model may suffer from non-stationarity problems if the equilibrium error is not an $I(0)$ process (see Phillips and Sul (2003)[7]). In order to avoid these additional problems, we restrict ourselves to the case of a growth model that depends only on explanatory variables expressed in terms of growth rates (as in Zhang et al. (2012)).

Recall that economic efficiency $A_{1i}(t)$ depends on economic policies, institutions, etc. ($V_{ij}(t)$ for $j = 1, ..., J$). Many studies considered the policies and institutional dimensions in investigating the sources of economic growth. They focused on the transmission mechanism linking policy to growth as well as on cross-country differences in policy settings. The main idea is that production requires more than just direct investment in physical capital and labor but also investment in knowledge, human capital, research and development (R&D), infrastructure, etc. (see Temple (1999, 2001), Bassanini et al. (2001), Barro and Sala-i Martin (2004), Kappler (2004), Durlauf et al. (2005), Kim and Heshmati (2013) to mention a few). Very often, we find proxy variables of R&D and innovation (as private and public expenditures on R&D, patent and trademark applications, etc.). Since the work of Mankiw et al. (1992), it is a standard approach of empirical growth models to include a measure of human capital stock or a measure of its accumulation together with physical capital (sometimes measured by the average number of years of education among the working-age population, based on figures on educational attainment or the level of education of the labor force, the school enrollment, at primary, secondary or tertiary levels). Other studies introduce monetary and fiscal policy variables such as inflation, real government consumption expenditures as a proportion of GDP, tax ratio (proportion of total indirect taxes to total direct taxes). Another sphere of influence on growth to be considered is international trade as exposure to competition through openness or the diffusion of technology through trade can improve the economic growth of a country. To proxy for the trade openness of a country, some authors use the share of exports and imports to GDP, others use the trade ratio (as a percent of GDP). Another strand of the growth literature focuses on the relationship between financial market development and economic growth. Other studies consider demography, aging population, life expectancy, income inequality and their effects on growth. The purpose of this paper is not to embrace

---

[7]Phillips and Sul (2003) allow for the presence of heterogeneity across countries and regions as well as over time.

all potential variables explaining the sources of growth but to introduce some of them in addition to physical and human capital and labor.

Before studying the data base, let us return to the semiparametric specification (4)-(5) and highlight the estimation method used.

2.2. The Bayesian semiparametric model with random coefficients

In this section, we present the panel data semiparametric model with random coefficients. We are influenced by the work of Zao et al. (2006), Lee (2016) and Lee and Wand (2016) and we follow their methodology. Their model is a linear-mixed Gaussian specification given by:

$$y \sim N\left(X^P\beta^P + W^P\eta^P + f\left(X^S\right), \sigma_\varepsilon^2 I_{NT}\right) \tag{7}$$

where y is an $(NT \times 1)$ vector $(y_{11}, ..., y_{1T}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT})'$ for $i = 1, ..., N$ countries and $t = 1, ..., T$ time periods. X^P is an $(NT \times K^P)$ matrix of covariates, W^P is an $(NT \times NK^P)$ block-diagonal matrix of the X_i^P submatrices, and I_{NT} is an $(NT \times NT)$ identity matrix. $X_{it,1}^P$ is the intercept and $X_{it,k}^P$, $2 \leq k \leq K^P$ are the other control covariates. The dimensions of the vectors β^P and η^P are respectively $(K^P \times 1)$ and $(NK^P \times 1)$. The random intercept is defined by the sum $(\beta_1^P + \eta_{i,1}^P)$, the random slope for variable $X_{i,2}$ is the sum $(\beta_2^P + \eta_{i,2}^P)$, etc. The semiparametric additive function is given by (see also Ruppert et al. (2003))

$$
\begin{aligned}
f\left(X^S\right) &= X^S\beta^S + W^S\eta^S = \sum_{k=1}^{K^S} X_k^S\beta_k^S + \sum_{k=1}^{K^S} W_k^S\eta_k^S \\
&= \sum_{k=1}^{K^S} X_k^S\beta_k^S + \sum_{k=1}^{K^S}\sum_{l=1}^{q_k^S} \eta_{kl}^S W_{kl}^S\left(X_k^S\right) \text{ with } \eta_{kl}^S \sim N\left(0, \sigma_{\eta_k}^2\right)
\end{aligned}
\tag{8}
$$

X^S contains K^S covariates that are not already included in X^P. The $(NT \times q^{K^S})$ matrix W^S matrix, with $q^{K^S} = \sum_{k=1}^{K^S} q_k^S$, contains spline basis functions W_k^S of the same covariates, using q_k^S knots and η^S are the $(q^{K^S} \times 1)$ spline coefficient vectors. Then, for the covariate X_k^S, $W_k^S = \left\{W_{kl}^S\left(X_k^S\right), 1 \leq l \leq q_k^S\right\}$ is a set of a penalized spline functions[8] of size q_k^S and $\sigma_{\eta_k}^2$ is the penalized parameter for the spline coefficients $\left\{\eta_{k1}^S, ..., \eta_{kq_k^S}^S\right\}$.

Hence,

$$
\begin{aligned}
y &\sim N\left(X^P\beta^P + W^P\eta^P + X^S\beta^S + W^S\eta^S, \sigma_\varepsilon^2 I_{NT}\right) \\
&\sim N\left(X\beta + W\eta, \sigma_\varepsilon^2 I_{NT}\right)
\end{aligned}
\tag{9}
$$

where

$$
\begin{aligned}
X^P &= vec\left(X_1^P, ..., X_N^P\right), \ W^P = \underset{(1 \leq i \leq N)}{\text{blockdiag}}\left(X_i^P\right), \beta = \left(\beta^{P'}, \beta^{S'}\right)' \\
X &= \left(X^P, X^S\right), \eta = \left(\eta^{S'}, \eta^{P'}\right)', \text{ and } W = \left(W^S, W^P\right)
\end{aligned}
\tag{10}
$$

X and W are the fixed effects and random effects design matrices associated with the fixed effects and random effects vectors β and η.[9] The random effects vector η^P has an unstructured $(K^P \times K^P)$

---

[8]Penalized splines can be viewed as random effects as one penalizes the spline basis function coefficients by treating them as a random sample from a multivariate normal distribution to avoid overfitting (see Lee (2016)).

[9]This terminology is different from what the panel data literature dubs as "fixed" and "random" effects.

covariance matrix $I_N \otimes \Sigma^P$, with $\otimes$ denoting the Kronecker product. The spline coefficient vector η^S has a block-diagonal covariance matrix $\sigma^2_{\eta_k} I_{q^S_k}$ where $\sigma^2_{\eta_k}$ is the penalized parameter for the spline coefficients $\{\eta^S_{kl}, 1 \leq l \leq q^S_k\}$. Then, the random effects covariance matrix $Cov(\eta)$ is homoskedastic. Lee and Wand (2016) used a Bayesian approach to fit the model to the data. The full Bayesian model (with priors on parameters and hyperparameters) is given by:

$$y \quad | \quad \beta, \eta, \sigma^2_\varepsilon \sim N\left(X\beta + W\eta, \sigma^2_\varepsilon I_{NT}\right) \tag{11}$$

with

$$\begin{cases}
\beta & \sim & N\left(0, \sigma^2_\beta I_K\right), I_K = I_{K^P + K^S} \\
\eta \mid \Sigma^P, \sigma^2_{\eta_k} & \sim & N\left(0, \begin{pmatrix} \underset{(1 \leq k \leq K^S)}{\text{blockdiag}\left(\sigma^2_{\eta_k} I_{q^S_k}\right)} & 0 \\ 0 & I_N \otimes \Sigma^P \end{pmatrix}\right) \\
\Sigma^P \mid \psi^P_1, .., \psi^P_{K^P} & \sim & IW\left(\nu + K^P - 1, 2\nu \text{diag}\left(1/\psi^P_1, .., 1/\psi^P_{K^P}\right)\right) \\
\psi^P_p & \sim & IG\left(\frac{1}{2}, \Psi^{-2}_p\right), 1 \leq p \leq K^P \\
\sigma^2_\varepsilon \mid \psi_\varepsilon & \sim & IG\left(\frac{1}{2}, 1/\psi_\varepsilon\right) \\
\psi_\varepsilon & \sim & IG\left(\frac{1}{2}, \Psi^{-2}_\varepsilon\right) \\
\sigma^2_{\eta_k} \mid \psi_{\eta_k} & \sim & IG\left(\frac{1}{2}, 1/\psi_{\eta_k}\right), 1 \leq k \leq K^S \\
\psi_{\eta_k} & \sim & IG\left(\frac{1}{2}, \Psi^{-2}_{\eta_k}\right)
\end{cases}$$

where $IW(.)$ and $IG(.)$ denote the inverse-Wishart and inverse-Gamma distributions, respectively. [10],[11] The resulting joint posterior distribution does not have a known tractable distribution and the parameters have to be sampled using MCMC techniques such as Gibbs sampling. But, MCMC can be computationally prohibitive and do not scale well when applied to massive data sets and/or models that require storage and inversion of large sparse covariance matrices. Inference based on MCMC can be very slow for such models and MCMC methods may suffer from poor mixing. Instead we will use variational Bayesian inference.

2.3. The mean field variational Bayes approximation

Variational Bayesian inference is a scalable and computationally effective method for Bayesian computation. It uses a deterministic optimization approach to approximate the posterior distribution (see Bishop (2006) and Blei et al. (2017)). Let us give a quick overview of the mean field variational Bayes (hereafter MFVB) method and its application to the semiparametric panel data model with

---

[10]The standard deviation parameters have independent Half-Cauchy priors $\sigma_{\eta_k} \sim$ Half-Cauchy(Ψ_{η_k}), $\sigma_\varepsilon \sim$ Half-Cauchy(Ψ_ε) which are equivalent to the following statements: $\sigma^2_{\eta_k} \sim IG\left(\frac{1}{2}, 1/\psi_{\eta_k}\right)$ and $\psi_{\eta_k} \sim IG\left(\frac{1}{2}, \Psi^{-2}_{\eta_k}\right)$ and $\sigma^2_\varepsilon \sim IG\left(\frac{1}{2}, 1/\psi_\varepsilon\right)$ and $\psi_\varepsilon \sim IG\left(\frac{1}{2}, \Psi^{-2}_\varepsilon\right)$. Transformed cubic O'Sullivan splines with $q^S_k = 25$ knots are used for the estimation of W^S. O'Sullivan penalized splines are similar to P-splines, but have the advantage of being a direct generalization of smoothing splines (see also Wand and Ormerod (2008)). The quantity of spline basis functions has a minimal effect on the adequacy of (11) and, as Ruppert (2002) showed, the number of knots q^S_k is not a crucial parameter because smoothing is controlled by the penalty parameter.

[11]The values of the hyperparameters are those of diffuse priors with: $\sigma^2_\beta = 10^5$, $\Psi_\varepsilon = 10^5$, $\Psi_\eta = 10^5$ and $\nu = 2$. For the variance-covariance matrix Σ^P of the random part η^P_{ip}, $(p = 1, ..., K^P)$ of the coefficients, the hyperparameter of the inverse-gamma distribution of ψ^P_p was shrunk to $\Psi_P = 10$. It reduces the variability of the η^P_{ip}'s and avoids unplausible marginal effects $\left(\beta^P_p + \eta^P_{ip}\right)$. We thank Dek Terrell for this suggestion.

random coefficients. For a parameter vector φ that is continuous over the parameter space Φ, Bayes theorem allows one to define the posterior distribution as:

$$p\left(\varphi \mid y\right) = \frac{p\left(\varphi, y\right)}{p\left(y\right)} = \frac{p\left(y \mid \varphi\right) p\left(\varphi\right)}{p\left(y\right)} \text{ with } p\left(y\right) = \int_{\Phi} p\left(\varphi, y\right) d\varphi \tag{12}$$

Let q be an arbitrary density function over Φ. Then, the logarithm of the marginal likelihood satisfies (see Bishop (2006), Ormerod and Wand (2010)):

$$\begin{aligned}
\log p\left(y\right) &= \log p\left(y\right) \int_{\Phi} q\left(\varphi\right) d\varphi = \int_{\Phi} q\left(\varphi\right) \log p\left(y\right) d\varphi = \int_{\Phi} q\left(\varphi\right) \log \left\{ \frac{p\left(\varphi, y\right) / q\left(\varphi\right)}{p\left(\varphi \mid y\right) / q\left(\varphi\right)} \right\} d\varphi \\
&= \int_{\Phi} q\left(\varphi\right) \log \left\{ \frac{p\left(\varphi, y\right)}{q\left(\varphi\right)} \right\} d\varphi + \int_{\Phi} q\left(\varphi\right) \log \left\{ \frac{q\left(\varphi\right)}{p\left(\varphi \mid y\right)} \right\} d\varphi \\
&= \log \underline{p}\left(y, q\right) + KL(q, p)
\end{aligned} \tag{13}$$

where $KL(q, p)$ is the Kullback-Leibler divergence between $q\left(\varphi\right)$ and $p\left(\varphi \mid y\right)$. Furthermore, $\log \underline{p}\left(y, q\right)$ is a lower bound on the marginal log-likelihood. The Kullback-Leibler divergence becomes

$$\begin{aligned}
KL(q, p) &= E_{q(\varphi)}\left[\log q\left(\varphi\right)\right] - E_{q(\varphi)}\left[\log p\left(\varphi \mid y\right)\right] \\
&= E_{q(\varphi)}\left[\log q\left(\varphi\right)\right] - E_{q(\varphi)}\left[\log p\left(\varphi, y\right)\right] + \log p\left(y\right)
\end{aligned} \tag{14}$$

where the last term, $\log p\left(y\right)$, is a constant. The minimization of the Kullback-Leibler divergence is thus equivalent to maximizing the scalar quantity,

$$\log \underline{p}\left(y, q\right) = E_{q(\varphi)}\left[\log \left(\frac{p\left(\varphi, y\right)}{q\left(\varphi\right)}\right)\right] \tag{15}$$

which is usually referred as the evidence lower bound (ELBO).

MFVB approximation is analogous to Gibbs sampling for conjugate models (see Bishop (2006), Ormerod and Wand (2010), Pham et al. (2013), Lee and Wand (2016) and Blei et al. (2017) to mention a few). The parameters of the approximate distribution are chosen to minimize some measure of distance (as the Kullback-Leibler divergence) between the approximation and the posterior. For that, let $\{\varphi_1, ..., \varphi_M\}$ be a partition of the parameter vector φ. The MFVB approximates the posterior distribution $p\left(\varphi \mid y\right)$ by the product of the q-densities:

$$q\left(\varphi\right) = \prod_{j=1}^{M} q_j\left(\varphi_j\right) \tag{16}$$

The *optimal* q-densities which minimize the Kullback-Leibler divergence are given by

$$q_j^*\left(\varphi_j\right) \propto \exp\left[E_{q(-\varphi_j)}\left\{\log p\left(\varphi_j \mid \text{rest}\right)\right\}\right] \ , j = 1, ..., M \tag{17}$$

where $E_{q(-\varphi_j)}$ denotes expectation with respect to $\prod_{k \neq j} q_k\left(\varphi_k\right)$.

rest $\equiv \{y, \varphi_1, ..., \varphi_{j-1}, \varphi_{j+1}, ..., \varphi_M\}$ is the set containing the rest of the random vectors in the model, except φ_j and the distributions $\left(\varphi_j \mid \text{rest}\right)$ are the full conditionals in the MCMC literature. To obtain the optimal q-densities under product restriction (16), the iterative scheme consists firstly of initializing $q_j^*\left(\varphi_j\right)$, $1 \leq j \leq M$, then of executing cycles through updates

$$q_j^*\left(\varphi_j\right) \leftarrow \frac{\exp\left[E_{q(-\varphi_j)}\left\{\log p\left(y, \varphi\right)\right\}\right]}{\int \exp\left[E_{q(-\varphi_j)}\left\{\log p\left(y, \varphi\right)\right\}\right] d\varphi_j} \ , 1 \leq j \leq M \tag{18}$$

until the increase in $\log \underline{p}(y,q)$ is negligible.

We outline the MFVB approach of Lee and Wand (2016) to estimate the Bayesian semiparametric model with random coefficients. They derive the MFVB approximation of the linear-mixed Gaussian model-based penalized spline specification (11) on the following factorization:

$$
\begin{aligned}
p(\varphi \mid y) &= p(\varphi_1,...,\varphi_M \mid y) = p\left(\beta, \eta^P, \eta^S, \psi^P, \psi_\eta, \psi_\varepsilon, \Sigma^P, \sigma_\eta^2, \sigma_\varepsilon^2 \mid y\right) \qquad (19) \\
&\approx q\left(\beta, \eta^P, \eta^S, \psi^P, \psi_\eta, \psi_\varepsilon, \Sigma^P, \sigma_\eta^2, \sigma_\varepsilon^2\right) \\
&= q\left(\beta, \eta^P, \eta^S\right) q\left(\Sigma^P\right) q\left(\sigma_\varepsilon^2\right) q\left(\psi_\varepsilon\right) \prod_{k=1}^{K^P} q\left(\psi_p\right) \prod_{k=1}^{K^S} q\left(\psi_{\eta_k}\right) \prod_{k=1}^{K^S} q\left(\sigma_{\eta_k}^2\right) \\
&= q\left(\beta, \eta\right) q\left(\Sigma^P\right) q\left(\sigma_\varepsilon^2\right) q\left(\psi_\varepsilon\right) \prod_{r=1}^{K^P} q\left(\psi_p\right) \prod_{l=1}^{K^S} q\left(\psi_{\eta_k}\right) \prod_{k=1}^{K^S} q\left(\sigma_{\eta_k}^2\right)
\end{aligned}
$$

After tedious derivations[12], this leads to the following forms of the optimal q-densities

$$
\begin{cases}
q^*(\beta, \eta) &\sim N\left(\mu_{q(\beta,\eta)}, \Sigma_{q(\beta,\eta)}\right) \\
q^*\left(\Sigma^P\right) &\sim IW\left(\nu + N + K^P - 1, B_{q(\Sigma^P)}\right) \\
q^*\left(\sigma_\varepsilon^2\right) &\sim IG\left(\frac{1}{2}(T+1), B_{q(\sigma_\varepsilon^2)}\right) \\
q^*\left(\psi_\varepsilon\right) &\sim IG\left(1, B_{q(\psi_\varepsilon)}\right) \\
q^*\left(\sigma_{\eta_k}^2\right) &\sim IG\left(\frac{1}{2}\left(q_k^S + 1\right), B_{q(\sigma_{\eta_k}^2)}\right) \\
q^*\left(\psi_{\eta_k}\right) &\sim IG\left(1, B_{q(\psi_{\eta_k})}\right) \\
q^*\left(\psi_p\right) &\sim IG\left(\frac{1}{2}\left(\nu + K^P\right), B_{q(\psi_p)}\right)
\end{cases}
\qquad (20)
$$

where the parameters are updated according to Algorithm 1 (see Appendix 1). Convergence in Algorithm 1 is assessed using ELBO on the marginal log-likelihood:

$$
\log \underline{p}(y,q) = E_{q(\varphi)}\left[\log\left(\frac{p(\varphi, y)}{q(\varphi)}\right)\right] = E_{q(\varphi)}\left[\begin{array}{l} \log p\left(y, \beta, \eta, \psi^P, \psi_\eta, \psi_\varepsilon, \Sigma^P, \sigma_\eta^2, \sigma_\varepsilon^2\right) \\ -\log q\left(\beta, \eta, \psi^P, \psi_\eta, \psi_\varepsilon, \Sigma^P, \sigma_\eta^2, \sigma_\varepsilon^2\right) \end{array}\right]
$$

which is presented in Appendix 1. Updating parameters is stopped when the maximum ELBO is reached.[13]

Variational inference algorithms involve different implementation challenges from sampling algorithms[14] but the output is a distribution, rather than a sample and the accuracy scores of the MFVB approximation (as compared to MCMC) generally exceed $95 - 97\%$ in most papers on MFVB.[15] Last but not the least, the computing time gains afforded by MFVB algorithms, as compared to Gibbs sampling, are huge (see for instance Pham et al. (2013) and Lee and Wand (2016)).

---

[12]See Lee (2016) for the derivations of the optimal q-densities.

[13]The ELBO is judged to cease increasing when the tolerance criterion is less than 10^{-7}. This algorithm is part of the family of coordinate ascent variational inference (CAVI). It iteratively optimizes each factor of the mean field variational density, while holding the others fixed (see Bishop (2006) and Blei et al. (2017)).

[14]They are harder, in that they may require lengthy mathematical derivations to determine the update rules. However, once implemented, variational Bayes can be easier to test, because one can use the standard checks for optimization code (gradient checking, local optimum tests, etc.). Most variational inference algorithms converge to optima, which eliminates the need to check convergence diagnostics.

[15]See for instance Bishop (2006), Ormerod and Wand (2010), Faes et al. (2011), Pham et al. (2013), Ranganath et al. (2014), Lee and Wand (2016), Blei et al. (2017), Ranganath (2017) and Koop and Korobilis (2018) to mention a few.

In the supplementary material, we conduct a small Monte Carlo simulation study to assess the performance of the MFVB algorithm in terms of Bayesian inferential accuracy and computational speed.[16] We compare Algorithm 1 to an MCMC Gibbs sampling of the specification (11) for a couple of (N, T) $= (500, 20)$, $(25, 45)$ observations. The accuracy of the MFVB approximation is very high since the accuracy scores[17] range from 95% to 97% and the computational speed of MFVB is 1737 (resp. 297) times faster than the MCMC for $N = 500$, $T = 20$ (resp. for $N = 25$, $T = 45$). Thus, the MFVB approach has big advantages, as compared to MCMC technique such as Gibbs sampling, with a remarkable ability to perform high quality Bayesian inference for large panel data models faster than ever before.

3. The data

The data come from the World Bank's World Development Indicators and the OECD national accounts. The initial data sets are unbalanced with missing values. The percentage of missing values are between 0 and 26% depending on the variable considered (except for the average number of years of schooling of the population aged between 25 and 64 years which is a five-years census series). We focused on OECD countries and dropped variables and years for which the rate of missing values was larger than 15% (except for the average number of years of schooling). The remaining variables with missing data were processed using cubic B-splines to obtain satisfactory imputations.[18] The final data set utilizes a balanced panel data set for 23 OECD countries over 45 years (1971-2015) (see Appendix 2 for the description of the variables in the data set).

Figure 1 shows the individual means of the growth rate of GDP (constant 2010 PPP US$). The average value is 2.36% per year with a quite large range (from 1.57% for Switzerland to 4.10% for Ireland). But most of the 23 OECD countries have a growth rate lying between 2% and 3%.

Figure 2 shows the time means of the GDP growth rate and reveals a decreasing trend over

---

[16]We thank an anonymous referee for this suggestion.

[17]The MCMC samples of the full Bayesian model (11) are obtained *via* the `RStan` package. Some elements of the R code are available in the supplementary material of Lee and Wand (2016). The accuracy of MFVB approximation for a generic parameter φ is assessed using the accuracy score defined by Faes et al. (2011) as

$$\text{acc}\left(q^*\left(\varphi\right)\right) \quad = \quad 1 - \frac{1}{2}\int_{-\infty}^{\infty}\left|q^*\left(\varphi\right) - p_{MCMC}\left(\varphi \mid y\right)\right|d\varphi$$

where $p_{MCMC}\left(\varphi \mid y\right)$ is an accurate MCMC-based approximation to $p\left(\varphi \mid y\right)$. One proposed estimation of the $p_{MCMC}\left(\varphi \mid y\right)$ is based on the binned kernel density estimate with the direct plug-in bandwidth (see Wand and Jones (1995), Wand and Ripley (2009) and Lee and Wand (2016)).

[18]We have chosen a missingness rate of 15% that allows us to keep the maximum number of observations for the maximum number of countries. We also tried multiple imputation, using Bootstrap-based EM algorithms proposed by Honaker and King (2010) and Honaker et al. (2011) but we got implausible values, mainly for the oldest or most recent years. So, we prefer to use cubic B-spline interpolation. Moreover, our choice of using a cubic spline smoothing method rather than a multiple imputation method (like MICE or EM) is reinforced by the results of Yoon et al. (2017). They have compared the most familiar methods for estimating missing data (recurrent neural networks (RNN), cubic splines, multiple imputation, FIML, EM, ...). They show that RNN, just followed by cubic splines, give the best results (*i.e.*, smallest rmse) as compared to imputation (MICE (multiple imputation by chained equations) or EM). Increasing the missingness rate doesn't change significantly the rmse of RNN-based methods or cubic smoothing while the one of MICE or EM is increasing faster whatever the individual size and/or the time length.

the period 1971-2015 with some well-known troughs, most notably the 1974-75 oil crisis, the early 1980s, 1990s and 2000s recessions and the 2008-2009 crisis. The average rate (2.48%) ranges between 6.04% in 1973 and -3.81% in 2009 and, for the entire period, the GDP growth rate fell by 47.99%. Interestingly, this decrease does not appear to be at a constant rate. This is confirmed by the smooth local polynomial (of order 6). However, this common nonlinear trend hides different dynamics across countries. Figure 3 shows that the smooth local polynomial is different across countries. In other words, this confirms that a common trend is a simplifying hypothesis and that marked heterogeneities exist between countries.

The scatter plots of Figures B1, B2 and B3 given in the supplementary material[19] provide valuable insights into the expected signs of the relationships between the dependent variable and the explanatory variables. The productivity variable (Y/L) is defined as the ratio of the GDP (constant 2010 PPP US$) and the working age population (15-64 years). The capital intensity variable (K/L) is defined as the ratio of the physical capital stock (constant 2010 PPP US$) — obtained with the perpetual inventory method[20] — and the working age population (15-64 years). The human capital stock (H) is defined as the average number of years of schooling of the population aged between 25 and 64 years.

Recall from section 2 that economic efficiency depends on several variables (V) related to R&D, innovation, economic policies, etc. Here, we use business sector expenditure on R&D (% of GDP) for R&D intensity, patent applications, trademark applications, government final consumption expenditure (% of GDP), trade ratio (% of GDP) for openness, inflation, stock market capitalisation (% of GDP) for financial development, income share held by the highest 10%, for inequality and life expectancy at birth. Figure B1 shows a positive relationship between growth rate of GDP per capita and capital intensity while that with human capital appears to be more tenuous. Similarly, the relationship between growth rate of GDP per capita, R&D intensity or patent applications do not seem to be of the right sign or significant. Figure B2 shows a strong positive relationship between the growth rate of GDP per capita and the growth rate of trademark applications and trade, and a strong negative relationship between the growth rate of GDP per capita and the growth rate of government expenditures. On the other hand, the link between financial development and GDP is more diffuse. In Figure B3, the inflation-GDP link is weakly negative while that between inequality and GDP appears to be significantly increasing. On the other hand, the increase in life expectancy at birth does not seem to "play" a role in GDP growth.

The graphs show strong heterogeneity as well as different trends across countries and years. We carry out unit root tests on these variables both in logs and in growth rates. We apply the Pesaran (2007) cross-sectional augmented IPS (CIPS) test. Like the IPS test, this relaxes the assumption of a common autoregressive parameter in the augmented Dickey-Fuller (ADF) specification contrary to other standard tests such as Levin-Lin-Chu, Harris-Tzavalis or Breitung tests. In addition, it allows cross-section dependence across the countries.[21] As cultural, institutional, environmental and

---

[19]We report graphs and tables in the supplementary material for the sake for brevity.

[20]See Appendix 2.

[21]Cross-sectionally augmented Dickey-Fuller (CADF) test augments the standard Dickey-Fuller regressions with cross-sectional averages of lagged levels and its first differences.

other factors make the common autoregressive parameter assumption tenuous for both macro- and microeconometric panel data sets, the CIPS tests seem appropriate. Pesaran (2007)'s CIPS statistics test the null hypothesis of panel unit roots against the possibility of some countries being stationary while others are non-stationary.

Table 1 gives the results of the CIPS tests. For the variables in logs levels, we cannot reject the null hypothesis of panel unit roots except for the life expectancy (with individual-specific intercepts) and for trademark applications (with individual-specific intercepts and trend). In contrast, for the growth rates, we always reject the null hypothesis of panel unit roots. These statistics allow us to design a growth rate model with the presence of a trend. This supports taking into account the heterogeneity of the intercepts and especially the slopes of the coefficients of the model. These figures and statistics show that there are fundamental differences among countries. The motivation of a mixed fixed- and random-coefficients model is then conditioned on these individual specific effects. Such a specification allows us to draw inference on certain population characteristics through the imposition of *a priori* constraints on the coefficients (see Hsiao (2003), Hsiao and Pesaran (2008), Hsiao (2015) among others). These different elements lead us to choose a semiparametric specification with random intercepts and slopes coefficients.

4. The results

The Bayesian Gaussian semiparametric model (with random intercepts and slopes coefficients) (6) proposed in section 2 may be written as:

$$\Delta \log \left(\frac{Y}{L}\right)_{it} \quad \sim \quad N \left(\sum_{k=1}^{K^P} \left(\beta_k^P X_{it,k}^P + \eta_{i,k}^P W_{it,k}^P \right) + \Delta f_i \left(\frac{t}{T}\right), \sigma_\varepsilon^2 I_{NT} \right) \qquad (21)$$

$$\text{where } \Delta f_i \left(\frac{t}{T}\right) \quad = \quad \beta_i^S \left(\frac{t}{T}\right) + \sum_{k=1}^{q_k^S} \eta_{ik}^S \Delta W_{ik}^S \left(\frac{t}{T}\right)$$

$$\text{and } W^P \quad = \quad \operatorname*{blockdiag}_{(1 \leq i \leq N)} \left(X_i^P\right)$$

for $i = 1, ..., N(= 23)$, $t = 1, ..., T(= 45)$ where $\Delta \log y_{it} \left(\equiv \log y_{it} - \log y_{it-1}\right)$, $X_{it,1}^P$ is the intercept and $X_{it,k}^P$, $2 \leq k \leq K^P$ are the other covariates in growth rates for the capital intensity, the human capital, R&D, patents, trademarks, etc. $\left(\beta_k^P + \eta_{i,k}^P\right)$ is the random coefficient of the covariate $X_{it,k}$ for country i while $\left(\beta_1^P + \eta_{i,1}^P\right)$ corresponds to the random intercept.[22] The smoothed function (*i.e.*, the semiparametric part of the model)[23] consists only of a trend $\left\{\frac{t}{T}\right\}_{t=1}^T = \left(\frac{1}{T}, \frac{2}{T}, .., 1\right)'$. In view of Figure 3, we assume that the β^S coefficients and the spline coefficients η_k^S vary across countries allowing for country-specific trend effects.[24] This model encompasses the case with common trends in

---

[22] In our specification $y \sim N \left(X^P \beta^P + W^P \eta^P + X^S \beta^S + W^S \eta^S, \sigma_\varepsilon^2 I_{NT} \right)$, y is the growth rate of GDP per capita, $\left[X^S, W^S\right]$ are restricted to the time trends and $\left[X^P, W^P\right]$ are the other covariates including the intercept. In our case, X^S contains the N country-specific time trends (*i.e.*, $S = N$) or a common time trend (*i.e.*, $S = 1$).

[23] The smoothed function written in first differences comes from: $f_i \left(\frac{t}{T}\right) = g_i \left(\frac{t}{T}\right) + \sum_{k=1}^{q_k^S} \eta_{ik}^S W_{ik}^S \left(\frac{t}{T}\right)$ where $g_i \left(\frac{t}{T}\right) = \sum_{r=1}^R \zeta_{i,r} \left(\frac{t}{T}\right)^r$, $R = 2, 3, ...$ Then, $\Delta g_i \left(\frac{t}{T}\right) = \frac{2\zeta_{i,2}}{T} \left(\frac{t}{T}\right) + \sum_{r=1}^R \mathcal{O}\left(T^{-r}\right) \simeq \beta_i^S \left(\frac{t}{T}\right)$ and $\Delta f_i\left(t/T\right) = \beta_i^S \left(t/T\right) + \sum_{k=1}^{q_k^S} \eta_{ik}^S \Delta W_{ik}^S \left(t/T\right)$.

[24] More specifically: $X^S = I_N \otimes \left\{\frac{t}{T}\right\}_{t=1}^T$.

which $f_i\left(\frac{t}{T}\right) = f\left(\frac{t}{T}\right), \forall i$. For the spline bases of the smooth function $\Delta f_i(.)$, we follow the approach suggested by Baltagi and Li (2002) in the case of first difference models or dynamic models. The spline bases are approximated with the difference of spline bases : $\Delta W_{ik}^S\left(\frac{t}{T}\right) \approx W_{ik}^S\left(\frac{t}{T}\right) - W_{ik}^S\left(\frac{t-1}{T}\right)$.

Thus, it may be asked whether the hypothesis of country-specific trends $f_i\left(\frac{t}{T}\right)$ is justified or whether it is necessary to introduce common trends $f\left(\frac{t}{T}\right)$, *i.e.*, a function depending on common coefficients $\beta_i^S = \beta^S, \forall i$, on common spline bases $W_{ik}^S = W_k^S, \forall i$ and on common spline coefficients $\eta_{ik}^S = \eta_k^S, \forall i$. In order to test common trends *versus* country-specific trends, we use a bootstrap nonparametric R^2 test inspired from Zhang et al. (2012) (see also Kneip et al. (2012)). It allows one to test a semiparametric model with random intercept and slope coefficients with common trends (RCM_CT) against a semiparametric model with random intercept and slope coefficients with country-specific trends (RCM_CST).[25]

Using 100 bootstrap resamples and the Epanechnikov kernel associated with different bandwidths, results in Table 2 show that we cannot reject the null hypothesis of a semiparametric model with random intercept and slope coefficients with common trends (RCM_CT) since the bootstrapped p-values are more than 80%. Results are close no matter what bandwidth was used, *i.e.*, whether one uses the robust or the standard estimators of the Silverman (1986)'s rule of thumb. We have also applied this test to compare the semiparametric one-way error component models with common trends (OWEC_CT) or with country-specific trends (OWEC_CST).[26] Table 2 shows that the bootstrapped p-values are lower than 6% which allow us to strongly reject the null hypothesis of common trends. An interesting conclusion that can be drawn from these tests is that, with an RCM_CT specification, the assumption of common trends is not rejected, whereas with an OWEC_CT specification, the hypothesis of common trends is strongly rejected in favor of that of country-specific trends. We can discriminate between these two specifications: random intercept and slope coefficients model with semiparametric common trends (RCM_CT) *versus* random intercept and constant slope coefficients model with semiparametric country-specific trends (OWEC_CST). Using a bootstrap non parametric test inspired from Henderson et al. (2008) (see Appendix 4), and running 100 bootstrap resamples, we cannot reject the null of a random intercept and slope coefficients model with semiparametric common trends (RCM_CT) since the p-value is 100%. Hence, the introduction of random coefficients associated with common trends seems sufficient to take into account the various trends of GDP per capita growth rates between countries that have been observed in Figure 3.

Given the tests results, we estimate the random intercept and slope coefficients model with semiparametric common trends (RCM_CT) both with MCMC and MFVB approximation.[27,28] MCMC samples of size $11,000$ were generated. The first $1,000$ values were discarded as burn-in and the

---

[25]See Appendix 3.

[26]Zhang et al. (2012) estimate the following specification: $\Delta \log Y_{it} = \beta_0 + \beta_1 \Delta \log L_{it} + \beta_2 \Delta \log K_{it} + \beta_3 \Delta \log H_{it} + m\left(\frac{t}{T}\right) + \alpha_i + \varepsilon_{it}$. It is a constant slope coefficients model with fixed effects and nonparametric common trends $m\left(\frac{t}{T}\right)$. They use profile least squares on quarterly interpolated data for 16 OECD countries over 1975Q4 - 2010Q3. They reject the null hypothesis of common trends at the 10% level.

[27]We thank an anonymous referee for this suggestion.

[28]Estimation was conducted using R version 3.4 on a MacBook Pro, 2.8 GHz core i7 16Go MGz DDR3 ram. Some elements of the R code are available in the supplementary material of Lee and Wand (2016).

remaining $10,000$ were thinned by a factor of 10. The MCMC approach required longer computational time (8169 seconds). Table 3 gives the posterior mean, posterior standard deviation, the 95% credible set and the Gelman-Rubin convergence diagnostic $\widehat{R}$. As $\widehat{R} < 1.2$ for all the parameters, convergence has been reached. This Table is provided for information purposes. We will focus on the results of the MFVB in Table 4. The computational time of the MFVB algorithm was very short. It took only 10.28 second and stopped only after 175 cycles. The computational speed of MFVB was 795 times faster than that of MCMC! Convexity properties of the MFVB algorithm guarantees quick convergence of such an algorithm to at least a local optima (see Figure 4 for the plot of the ELBO). Table 4 gives the estimated coefficients and the standard deviations for β^P, β^S and σ_ε^2 given by Algorithm 1. In addition, from the optimal q-densities (20), we generated $1,000$ draws of these parameters and calculated their posterior means, posterior standard deviations and their 95% credible sets as in typical Bayesian presentation.[29] To assess the performance of the MFVB algorithm in terms of Bayesian inferential accuracy, we have computed the accuracy scores (last column of Table 4). The accuracy of MFVB approximation is very high since the scores range from 89% to 98%, except for σ_ε^2. Figure 5 gives approximate posterior density functions obtained *via* MFVB and MCMC. Each pair of density function corresponds to a model parameter. The accuracy scores (in percent) on the topright of each plot show the accuracy of MFVB approximation compared against the MCMC benchmark and confirm MFVB's suitability for MCMC. In Table 4, we can see that most of the coefficients are significantly different from zero whatever the risk level, except for R&D intensity, patent and trademark applications and life expectancy growth rates. Elasticity of growth in GDP per capita relative to the stock of physical capital is around 37% which is quite standard for cross-country comparisons. The elasticity of growth in GDP per capita relative to human capital is also positive (29%) leading to decreasing returns to all capital as suggested in (2). The results also show a strong negative (-24.10%) elasticity of growth in GDP per capita relative to government consumption expenditures. Also, a positive (4.05%) elasticity of growth in GDP per capita relative to openness. One can note a negative (-10.46%) elasticity of growth in GDP per capita relative to inflation, as expected. Elasticity of growth in GDP per capita relative to the financial development is around 1.15%. The effect of the increase in inequality is significant and positive (5.4%). This positive elasticity may be a surprise at first sight, but it complements a long list of studies that report positive effects of inequality on GDP in the most developed countries[30]. Both dummy variables have a significant and negative effect, with the

---

[29]We thank Dek Terrell for this suggestion.

[30]The previous literature suggests different effects through different channels for the relationship between inequality and economic growth. Negative effects of inequality on growth may refer to the link between inequality and education level on the one hand and the effect of education on economic growth on the other hand. Furthermore, inequality may harm political and economic stability and thus limit the attractiveness of the economy for investment. Likewise, according to the endogenous fiscal policy theory, greater inequality may become unacceptable to voters. As a result, their preferences move towards higher taxation and redistribution (see Perotti (1996), Alesina and Perotti (1996)).

However, positive effects of inequality on growth may come from incentives for innovation and entrepreneurship. If the rates of return on investment (*e.g.*, in education) are high, then inequality may motivate more people to undertake studies and increase their level of education. Moreover, as savings rates tend to be higher in upper income classes, then saving and investment and thus economic growth should be higher in more unequal societies, *ceteris paribus* (see Kaldor

effect of the shock associated with the 2008-2009 crisis being more important. Last, the parametric part β^S of the common time trend effect is negative (-2.38%), which corresponds roughly to the 2.05 points lost between 1971 (4.278%) and 2015 (2.225%) of the observed time means of the GDP growth rates, corroborating what has been observed in Figure 2. Figure 6 shows the MFVB spline fit of time means of growth rates of GDP per capita and the narrow pointwise 95% credible set. It illustrates the good quality of the adjustment. The estimated MFVB spline fit reveals the nonlinearity of the trend on the whole period and the economic recovery after the collapse observed in 2009. Figures B4 and B5 in the supplementary material also confirm for some countries the good fit of the growth rates of GDP per capita even if the profiles of evolution are very different from each other. The semiparametric approach seems to be well justified.[31]

So far, we have discussed average marginal effects. One of the advantages of the Bayesian Gaussian semiparametric model with random coefficients is that it is possible to discriminate between these marginal effects $(\beta_k^P + \eta_{i,k}^P)$ for the $i = 1, ..., N(= 23)$ OECD countries. Figures B6 to B14 in the supplementary material plot bar charts of those marginal effects.[32] There are significant differences between countries. Thus the elasticities with respect to the physical capital stock vary between 19.80% (New Zealand) and 51.70% (Greece). Elasticities with respect to human capital mainly range from 23.38% (Portugal) and 37.82% (New Zealand). Elasticities with respect to public spending range from -8.38% (Luxembourg) and -13.99% (Norway). The elasticities of growth in GDP per capita relative to the trade ratio are mainly positive and range between 2.93% (UK) and 9.96% (Belgium), except for five countries (Ireland (-0.35%), Iceland (-2.13%), Norway (-3.62%), New Zealand (-4.14%) and especially Greece (-13.41%)). We also observe different elasticities of the growth in GDP per capita relative to inflation (from -17.29% (New Zealand) to -4.48% (Italy)). The elasticities with respect to the financial development vary between 0.41% (France) and 3.38% (Greece) just preceded by Ireland and Luxembourg. The growth of inequality positively affects that of GDP per capita varying

---

(1957), Lazear and Rosen (1981), Barro (2000)).

Some influential empirical studies (*e.g.*, Cingano (2014), Ostry et al. (2014), see also Neves et al. (2016)) found evidence in favour of a negative effects of inequality on growth. These studies have in common that they all rely on the System GMM dynamic panel estimator. But, they suffer from a weak instrument problem in the inequality-growth setting because lagged differences of inequality have practically no explanatory power for current inequality levels. In fact, using fixed effects models or Difference GMM estimators, Kolev and Niehues (2016) show that the effect of income inequality on growth is negative only for less-developed countries and for countries with high levels of inequality, and rather positive otherwise.

[31] For information purposes, we give in the supplementary appendix the estimation of the growth rate model for GDP per capita with random intercept and slope coefficients and semiparametric country-specific trends (see Table B1). This time, inequality growth and the dummy variable associated with the oil shock become non significant. The elasticity of growth in GDP per capita relative to human capital seems excessive (50%) leading to an unrealistic labor elasticity less than 12%. The country-specific time trend effects range between -4.68% (Japan), -3.91% (Greece) and (-1.45%) (Iceland) with countries like France (-3.51%), Germany (-2.91%) and the UK (-3.02%) being in the middle of the distribution. For some countries, this trend effect is not significantly different from zero . This is the case of Iceland, Switzerland and New Zealand. Given the 95% credible sets, it is highly likely that all these country-specific trend coefficients are not statistically different from each other since all of them overlap with each other. This confirms our choice of a random intercept and slope coefficients model with semiparametric common trends.

[32] We give these bar charts for only the variables with significant $\left(\beta_k^P + \eta_{i,k}^P\right)$ coefficients.

from 2.91% (UK), followed by 3.66% (Switzerland) and 3.81% (Germany) to 7.31% (Spain), 8.42% (Norway) and 9.29% (Greece). Also notable is that growth rates of GDP per capita for the 23 OECD countries have not responded in a homogeneous and similar manner to the oil crisis (1973-1975) and to the 2008-2009 crisis.

As stated in section 2.1, we can estimate the TFP growth based on the estimated discrete growth equation with $\Delta \log \widehat{A}_{it} = \Delta \log \left(\frac{Y}{L} \right)_{it} - \widehat{\alpha}_i \Delta \log \left(\frac{K}{L} \right)_{it} - \widehat{\gamma}_i \Delta \log \left(\frac{H}{L} \right)_{it}$ or by summing the estimated technical innovation change $\Delta \log \widehat{A}_{1,it} = \sum\limits_{j=1}^{J} \widehat{\omega}_{ij}^* \Delta \log V_{ijt}$ and the estimated technical efficiency change $\Delta \log \widehat{A}_{2,it} = \widehat{\beta}_i^S \left(\frac{t}{T} \right) + \sum_{k=1}^{q_k^S} \widehat{\eta}_{ik}^S \Delta W_{ik}^S \left(\frac{t}{T} \right)$. Figure B15 in the supplementary material gives the boxplot of the TFP growth for the 23 OECD countries. The mean value for all 23 OECD countries is around 1.097%. In mean, countries with the highest TFP growth rates are Norway, Ireland, Japan and Iceland, while countries with the lowest TFP growth rates are Italy, Greece, Spain and Switzerland. But the time variabilities, represented by the intervals between the lower and upper whiskers, differ greatly across countries. Countries such as Australia, Canada, France, the Netherlands, the United Kingdom and the United States have experienced relatively lower time variability of TFP growths than countries such as Greece, Ireland, Luxembourg or Portugal. Some extreme (outliers) observations (located outside the fences) underline the sensitivity of the TFP during some periods: strong positive growth for the years around 1985 and strong negative growth after the first oil shock and for the 2008-2009 crisis. These temporal evolutions of the TFP growth are traced in Figure B16 for some countries. The change in technical efficiency (Figure B17) has experienced a decrease of -0.053% over 1971-2015. This result is close to those of Sickles et al. (2015) or Duygun et al. (2017) who found a -0.06% decrease for 24 OECD countries over the period 1970-2000. The boxplot of the change in technical innovation (economic efficiency change) looks a lot like the TFP (see Figure B18). The mean value for all 23 OECD countries is around 1.30% which is higher than the value (0.84%) found by Sickles et al. (2015) on a shorter period. As compared to the TFP growth (Figure B15), we can note a quite similar ranking of the countries with the highest changes in technical innovation namely, Ireland, Iceland, Japan and United Kingdom, and countries with the lowest changes in technical innovation namely, Spain, Greece, Italy and Switzerland. Again, there are fairly strong time variations of the change in technical innovation for all 23 OECD countries. Some of them are illustrated in Figure B19. One of the flexibilities of the MFVB approach is therefore to easily estimate this TFP growth as well as the changes in the technical and innovation efficiencies. This sensitivity applied to covariates adds to the richness of the random coefficients model.[33]

Compared to other panel data studies on growth empirics, our approach has several advantages. First, it rejects the too restrictive assumption of the implicit hypothesis that all the slope coefficients

---

[33]In the supplementary material, we give the estimation of a random intercept and slope coefficients model with nonparametric common trends *i.e.*, $\Delta f \left(t/T \right) = \sum_{k=1}^{q_k^S} \eta_k^S \Delta W_k \left(t/T \right)$. Even if this model seems to fit well the time means of the growth rate of GDP per capita (see Figure B20), some variables have now no effect (see Table B2). This is the case of inflation and the dummy 1973-1975 in addition to the already irrelevant variables in Table 4. Moreover, the marginal effects also change as compared to those previously estimated (see Figures B21 to B24). The introduction of a semiparametric specification for common trends therefore seems more relevant than the use of a pure nonparametric specification.

are in fact identical for all countries. It allows one to estimate country-specific intercepts and slope coefficients as advocated by Islam (1995), Pesaran et al. (1999), Eberhardt and Teal (2011), Durlauf et al. (2005), among others. Second, the use of a Bayesian linear-mixed Gaussian model-based penalized spline specification allows us flexibility in the functional form of the common trends. In a field such as growth empirics, such a semiparametric specification can be promising. We provide a Bayesian approach that allows one to obtain easily computable estimators. In the class of approximate Bayesian computation methods, the mean field variational Bayes approximation is important for the researcher's toolkit. It gives results faster than the usual MCMC methods which can take in some cases hours or days to achieve the same results.[34]

5. Conclusion

This paper proposed semiparametric estimation of the relationship between growth rates of GDP per capita, capital intensity, human capital and other covariates including patent and trademark applications, government consumption expenditures, and trade, to mention a few. This is estimated using a panel of 23 OECD countries observed over the period $1971 - 2015$. We cannot reject the null hypothesis of a random intercept and slope coefficients specification with semiparametric common trends. The motivation of a mixed fixed- and random-coefficients model has been conditioned on these country specific effects. Following a recent approach proposed by Lee and Wand (2016), we have specified and estimated a mean field variational Bayes semiparametric model with random coefficients for this panel of countries. This approach has numerous advantages as compared to the MCMC technique such as Gibbs sampling. Results reveal different marginal effects across countries. The inclusion of random coefficients in a mixed model-based penalized spline basis function enriches the estimates and their interpretations, given the large diversity of responses across variables and countries.

---

[34]The application on children's birthweight of Lee and Wand (2016) took 4 days with MCMC and few minutes with MFVB.

6. Appendix

6.1. Appendix 1 - The mean field variational Bayes algorithm and the variational evidence lower bound

Algorithm 1. Mean field variational Bayes algorithm (see Lee and Wand (2016), pp. 882).

---

1. Initialize $\mu_{q(1/\sigma_\varepsilon^2)} > 0$, $\mu_{q(1/\psi_\varepsilon)} > 0$, $\mu_{q\left(1/\sigma_{\eta_k}^2\right)} > 0$, $\mu_{q\left(1/\psi_{\eta_k}\right)} > 0$, $1 \le k \le K^S$, $\mu_{q(1/\psi_p)} > 0$, $1 \le p \le K^P$, $M_{q\left((\Sigma^P)^{-1}\right)}$ positive definite.

2. Cycle through updates:

   (a) $R \leftarrow 0$, $r \leftarrow 0$, for $i = 1, ..., N$:

   $G_i \leftarrow \mu_{q(1/\sigma_\varepsilon^2)} \left(C_i^S\right)' X_i^P$; $H_i \leftarrow \left[\mu_{q(1/\sigma_\varepsilon^2)} \left(X_i^P\right)' X_i^P + M_{q\left((\Sigma^P)^{-1}\right)}\right]^{-1}$ with $C_i^S = \left[X_i^S,\ W_i^S\right]$

   $R \leftarrow R + G_i H_i \left(G_i\right)'$; $r \leftarrow r + G_i H_i \left(X_i^P\right)' y_i$

   (b) $\Sigma_{q(\beta,\eta^S)} \leftarrow \left[\mu_{q(1/\sigma_\varepsilon^2)} \left(C^S\right)' C^S + \begin{bmatrix} \sigma_\beta^{-2} I_K & 0 \\ 0 & \begin{array}{c} \text{blockdiag}\left(\mu_{q\left(1/\sigma_{\eta_k}^2\right)} I_{q_k^S}\right) \\ {\scriptstyle (1 \le l \le L)} \end{array} \end{bmatrix} - R\right]^{-1}$

   (c) $\mu_{q(\beta,\eta^S)} \leftarrow \mu_{q(1/\sigma_\varepsilon^2)} \Sigma_{q(\beta,\eta^S)} \left[\left(C^S\right)' y - r\right]$ with $C^S = \left[X^S,\ W^S\right]$

   (d) for $i = 1, ..., N$:

   $\Sigma_{q\left(\eta_i^P\right)} \leftarrow H_i + H_i \left(G_i\right)' \Sigma_{q(\beta,\eta^S)} G_i H_i$

   $\mu_{q\left(\eta_i^P\right)} \leftarrow H_i \left[\mu_{q(1/\sigma_\varepsilon^2)} \left(X_i^P\right)' y_i - \left(G_i\right)' \mu_{q(\beta,\eta^S)}\right]$

   $B_{q(\sigma_\varepsilon^2)} \leftarrow \mu_{q(1/\psi_\varepsilon)} + \frac{1}{2} \begin{bmatrix} \| D \|^2 + \text{tr}\left[\left(C^S\right)' C^S \Sigma_{q(\beta,\eta^S)}\right] \\ + \sum_{i=1}^N \text{tr}\left[\left(X_i^P\right)' X_i^P \Sigma_{q\left(\eta_i^P\right)}\right] \\ -2\mu_{q(1/\sigma_\varepsilon^2)}^{-1} \sum_{i=1}^N \text{tr}\left[G_i H_i \left(G_i\right)' \Sigma_{q(\beta,\eta^S)}\right] \end{bmatrix}$

   with $D = y - C^S \mu_{q(\beta,\eta^S)} \begin{pmatrix} X_1^P \mu_{q\left(\eta_1^P\right)} \\ \vdots \\ X_N^P \mu_{q\left(\eta_N^P\right)} \end{pmatrix}$

   (e) $\mu_{q(1/\sigma_\varepsilon^2)} \leftarrow \frac{1}{2} (T + 1) / B_{q(\sigma_\varepsilon^2)}$; $\mu_{q(1/\psi_\varepsilon)} \leftarrow 1 / \left[\mu_{q(1/\sigma_\varepsilon^2)} + \Psi_\varepsilon^{-2}\right]$

   (f) for $k = 1, ..., K^P$:

   $B_{q(\psi_p)} \leftarrow \nu \left(M_{q\left((\Sigma^P)^{-1}\right)}\right)_{kk} + \Psi_p^{-2}$; $\mu_{q(1/\psi_p)} \leftarrow \frac{1}{2} \left(\nu + K^P\right) / B_{q(\psi_p)}$

   $B_{q(\Sigma^P)} \leftarrow \sum_{i=1}^N \left(\mu_{q\left(\eta_i^P\right)} \mu_{q\left(\eta_i^P\right)}' + \Sigma_{q\left(\eta_i^P\right)}\right) + 2\nu \text{diag}\left(\mu_{q\left(1/\psi_1^P\right)}, ..., \mu_{q\left(1/\psi_{K^P}^P\right)}\right)$

   (g) $M_{q\left((\Sigma^P)^{-1}\right)} \leftarrow \left(\nu + N + K^P - 1\right) B_{q(\Sigma^P)}^{-1}$

   (h) for $k = 1, ..., K^S$:

   $\mu_{q\left(1/\psi_{\eta_k}\right)} \leftarrow 1 / \left[\mu_{q\left(1/\sigma_{\eta_k}^2\right)} + \Psi_{\eta_k}^{-2}\right]$; $B_{q(\psi_{\eta_k})} \leftarrow 1 / \mu_{q\left(1/\psi_{\eta_k}\right)}$

   $\mu_{q\left(1/\sigma_{\eta_k}^2\right)} \leftarrow \frac{q^S + 1}{2\mu_{q\left(1/\psi_{\eta_k}\right)} + \|\mu_{q\left(\eta_k^S\right)}\|^2 + \text{tr}\left[\Sigma_{q\left(\eta_k^S\right)}\right]}$; $B_{q(\sigma_{\eta_k}^2)} \leftarrow \frac{1}{2} \left(q_k^S + 1\right) / \mu_{q\left(1/\psi_{\eta_k}\right)}$

   (i) for $i = 1, ..., M$:

   $\Lambda_{q\left(\beta,\eta^S,\eta_i^P\right)} \equiv E_q \left[\left(\begin{bmatrix} \beta \\ \eta^S \end{bmatrix} - \mu_{q(\beta,\eta^S)}\right) \left(\eta_i^P - \mu_{q\left(\eta_i^P\right)}\right)'\right] \leftarrow \Sigma_{q(\beta,\eta^S)} G_i H_i$

   $\Sigma_{q\left(\beta,\eta^S,\eta_i^P\right)} \equiv Cov \begin{pmatrix} \beta \\ \eta^S \\ \eta_i^P \end{pmatrix} \leftarrow \begin{pmatrix} \Sigma_{q(\beta,\eta^S)} & \Lambda_{q\left(\beta,\eta^S,\eta_i^P\right)} \\ \Lambda'_{q\left(\beta,\eta^S,\eta_i^P\right)} & \Sigma_{q\left(\eta_i^P\right)} \end{pmatrix}$

(j) $\Sigma_{q(\beta,\eta)} \leftarrow \begin{pmatrix} \Sigma_{q(\beta,\eta^S)} & \Lambda_{q(\beta,\eta^S,\eta_1^P,...,\eta_N^P)} \\ \Lambda'_{q(\beta,\eta^S,\eta_1^P,...,\eta_N^P)} & \Sigma_{q(\eta_1^P,...,\eta_N^P)} \end{pmatrix}$

until the increase in the ELBO $\log \underline{p}(y,q)$ is negligible.

The variational lower bound on the marginal log-likelihood has the following expression (see Lee and Wand (2016), pp. 893):

$$
\begin{aligned}
\log \underline{p}(y,q) =\ & \frac{1}{2}K^P\left(\nu + K^P - 1\right)\log 2\nu - \frac{T}{2}\log 2\pi - \left(\frac{1}{2}K^P + K^S + 1\right)\log\pi \\
& -\frac{P}{2}\log\sigma_\beta^2 - \frac{\sigma_\beta^{-2}}{2}\left[\|\mu_{q(\beta)}\|^2 + \text{tr}\left[\Sigma_{q(\beta)}\right]\right] + \frac{1}{2}\left(\sum_{k=1}^{K^S} q_k^S + P + N\right) \\
& -\frac{1}{2}\sum_{i=1}^N \log\left|\mu_{q(1/\sigma_\varepsilon^2)}\left(X_i^P\right)' X_i^P + M_{q((\Sigma^P)^{-1})}\right| - \frac{1}{2}\log\left|\Sigma_{q(\beta,\eta^S)}^{-1}\right| \\
& -\log\left(\mathcal{C}_{K^P,\nu+K^P-1}\right) + \log\left(\mathcal{C}_{K^P,\nu+N+K^P-1}\right) - \frac{1}{2}\left(\nu + N + K^P - 1\right)\log\left|B_{q(\Sigma^P)}\right| \\
& +\sum_{k=1}^{K^S}\log\Gamma\left(\frac{q_k^S+1}{2}\right) - \frac{1}{2}\sum_{k=1}^{K^S}\left(q_k^S+1\right)\log B_{q(\sigma_{\eta_k}^2)} + K^S\log\Gamma\left(N(T+1)\right) \\
& -\frac{1}{2}(T+1)\log B_{q(\sigma_\varepsilon^2)} - \sum_{p=1}^{K^P}\log\Psi_p + K^P\log\Gamma\left(\frac{q^S+\nu}{2}\right) \\
& +\sum_{p=1}^{K^P}\nu M_{q((\Sigma^P)^{-1})}\mu_{q(1/\psi_p)} - \frac{1}{2}\left(K^P+\nu\right)\sum_{p=1}^{K^P}\log B_{q(\psi_p)} \\
& -\sum_{k=1}^{K^S}\left[\log\psi_{\eta_k} + \log B_{q(\psi_{\eta_k})} + \mu_{q(1/\psi_{\eta_k})}\mu_{q(1/\sigma_{\eta_k}^2)}\right] + \log\psi_\varepsilon \\
& -\log B_{q(\psi_\varepsilon)} + \mu_{q(1/\psi_\varepsilon)}\mu_{q(1/\sigma_\varepsilon^2)}
\end{aligned}
$$

where $\mathcal{C}_{a,b}$ is the normalizing factor: $\mathcal{C}_{a,b} = 2^{ab/2}\pi^{a(a-1)/4}\prod_{j=1}^{a}\Gamma\left(\frac{b+1-j}{2}\right)$ and Γ is the Gamma function.

6.2. Appendix 2 - Definition of the variables

Variable	Description
GDP (Y)	GDP (constant 2010 purchasing power parities (PPP) US$)
labor (L)	working age population (15-64 years)
physical capital (K)	physical capital (constant 2010 PPP US$) obtained from perpetual inventory method (see below)
human capital (H)	average number of years of schooling of the population aged between 25 and 64 years
RD intensity	business sector expenditure on R&D (% of GDP)
patents	patent applications, residents
trademark	trademark applications, total
gov.expenditure	general government final consumption expenditure (% of GDP)
trade ratio	trade (% of GDP)
inflation	growth rate of private final consumption deflator
financial development	stock market capitalisation (% of GDP)
inequality	income share held by the highest 10%
life expectancy	life expectancy at birth (years)

The net capital stock at the beginning of period t, K_t, can be written as a function of the net capital stock at the beginning of the previous period $t-1$, K_{t-1}, gross investment in the previous period, I_{t-1}, and consumption of fixed capital, D_{t-1}: $K_t = K_{t-1} + I_{t-1} - D_{t-1}$. Assuming geometric depreciation at a constant rate δ, we can rewrite the capital stock as: $K_t = (1-\delta)K_{t-1} + I_{t-1}$. Following Griliches (1980), De la Fuente and Domenech (2006) and Berlemann and Wesselhöft (2012), the estimation of the initial capital stock can be written as: $K_{t-1} = I_t / (g_I + \delta)$ where g_I is the growth rate of gross investment (gross fixed capital formation (constant 2010 PPP US$) in the database). It relies on the assumption that the economy under consideration is outside its long-term equilibrium. When g_I is replaced by g_{DGP}, the growth rate of GDP, the economy is supposed to be at its steady state. In our case, we fix $g_I = 0.04$ and the depreciation rate at $\delta = 0.15$. This depreciation rate is standard in recent empirical studies and even undervalued for some R&D intensive industries (such as telecommunications, chemical and pharmaceutical industry, electrical equipment, etc.) where the R&D investment depreciation rate is estimated at around $35-40\%$ per year (see Li (2012)).

6.3. Appendix 3 - A bootstrapped nonparametric test for common trends

The idea behind the test is the following. Under the null hypothesis H_0 of common trends, $\widehat{\xi}_{it}$ is a consistent estimate for the augmented residuals $\xi_{it} = \varepsilon_{it} + \eta_{1i}^P$ where η_{1i}^P is the specific effect coming from the random intercept $(\beta_1^P + \eta_{1i}^P)$. Under H_0, there is no time trend in $\{\xi_{it}\}_{t=1}^T$ for each cross-sectional unit i. Nevertheless, under H_1, ξ_{it} includes an individual-specific time trend component $f_i(t/T)$. This motivates us to consider a residual-based test for common trends in the spirit of the bootstrap version of the nonparametric test of Zhang et al. (2012).[35] For each cross-section i, they propose to run the nonparametric regression of $\left\{\widehat{\xi}_{it}\right\}_{t=1}^T$ on $\{t/T\}_{t=1}^T$: $\widehat{\xi}_{it} = m_i(t/T) + \nu_{it}$. Under H_0, we have $m_i(\tau) = 0$ for $\tau \in [0,1]$, then $\left\{\widehat{\xi}_{it}\right\}_{t=1}^T$ contains no useful trending information so that the nonparametric R^2 should be close to 0 for each individual i. Zhang et al. (2012) develop the asymptotic distribution of a normalized nonparametric R^2 which follows a $N(0,1)$ under the null hypothesis and a sequence of Pitman local alternatives. The procedure is the following:

1. obtain the augmented residuals $\widehat{\xi}_{it} = \widehat{\varepsilon}_{it} + \widehat{\eta}_{1i}^P$ where $\widehat{\varepsilon}_{it} = y_{it} - \left(X_{it}^P\widehat{\beta}^P + W_{it}^P\widehat{\eta}_i^P + \widehat{f}(t/T)\right)$ where $\widehat{\beta}^P$, $\widehat{\eta}_i^P$ and $\widehat{f}$ are obtained from the MFVB approximation of this restricted model and calculate the test statistics $\overline{\Gamma}_{NT}$ (see below).

2. Obtain the bootstrap error ξ_t^* by random sampling with replacement from $\left\{\widehat{\xi}_j, j = 1, 2, ..., T\right\}$ where $\widehat{\xi}_j = \left(\widehat{\xi}_{1j}, ..., \widehat{\xi}_{Nj}\right)'$.

3. Generate the bootstrap analogue of y_{it} by holding X_{it}^P, W_{it}^P, X_{it}^S, W_{it}^S as fixed: $y_{it}^* = X_{it}^P\widehat{\beta}^P + W_{it}^P\widehat{\eta}_i^P + \widehat{f}(t/T) + \xi_{it}^*$ for $i = 1, ..., N$ and $t = 1, ..., T$ where ξ_{it}^* is the ith element in the N-vector ξ_t^*.

4. Based on the bootstrap resample $\{y_{it}^*, X_{it}^P, W_{it}^P, X_{it}^S, W_{it}^S\}$, run the MFVB approximation of the model to obtain the bootstrapped residuals $\left\{\widehat{\xi}_{it}^*\right\}$.

---

[35]Zhang et al. (2012) estimate the following fixed effects specification: $y_{it} = X_{it}\beta + m\left(\frac{t}{T}\right) + \alpha_i + \varepsilon_{it}$. It is a constant intercept and slope coefficients model with (correlated) fixed effects and nonparametric common trends $m\left(\frac{t}{T}\right)$. They use profile least squares to estimate this semiparametric specification.

5. Based on $\left\{\widehat{\xi}_{it}^*\right\}$, compute the bootstrap test statistics $\overline{\Gamma}_{NT}^*$ defined analogously to $\overline{\Gamma}_{NT}$ but with $\widehat{\xi}_{it}$ being replaced by $\widehat{\xi}_{it}^*$.

6. Repeat step 2-5 for B times and index the bootstrap statistics as $\left\{\overline{\Gamma}_{NT,l}^*\right\}_{l=1}^{B}$.

7. The bootstrap p-value is calculated by $p^* = B^{-1}\sum_{l=1}^{B} 1\left\{\overline{\Gamma}_{NT}^* > \overline{\Gamma}_{NT}\right\}$ where $1\{.\}$ is the indicator function.

The test statistics $\overline{\Gamma}_{NT}$ proposed by Zhang et al. (2012) is given by:

$$\overline{\Gamma}_{NT} = \frac{1}{\sqrt{\widehat{\Omega}_{NT}}}\sqrt{\frac{b}{N}}\sum_{i=1}^{N}\frac{ESS_i - \widehat{\xi}_i'MQM\widehat{\xi}_i}{TSS_i/T}$$

where

$$\widehat{\Omega}_{NT} = \frac{2b}{T^2}\sum_{1\leq t\neq s\leq T}\alpha_{ts}^2\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\rho}_{ij}^2\right), \widehat{\rho}_{ij} = \widehat{\omega}_{ij}/\left(\widehat{\sigma}_i\widehat{\sigma}_j\right)$$

$$\widehat{\omega}_{ij} = T^{-1}\sum_{t=1}^{T}\left(\widehat{\xi}_{it} - \overline{\overline{\xi}}_i\right)\left(\widehat{\xi}_{jt} - \overline{\overline{\xi}}_j\right), \widehat{\sigma}_i^2 = T^{-1}\sum_{t=1}^{T}\left(\widehat{\xi}_{it} - \overline{\overline{\xi}}_i\right)^2, \overline{\overline{\xi}}_i = T^{-1}\sum_{t=1}^{T}\widehat{\xi}_{it}$$

with

$$TSS_i = \widehat{\xi}_i'M\widehat{\xi}_i, ESS_i = \widehat{\xi}_i'\left(\overline{H} - \overline{J}_T\right)\widehat{\xi}_i, M = \left(I_T - \overline{J}_T\right), \overline{J}_T = \iota_T\iota_T'/T$$

$$\alpha_{ts} = T\overline{H}_{ts} - 1 , Q = T^{-1}diag(\alpha_{11},...,\alpha_{TT})$$

and

$$\overline{H} = \int_0^1 H\left(\tau\right)d\tau, \tau \in [0,1], H\left(\tau\right) = V_b(\tau)z_b^{[1]}(\tau)\left(z_b^{[1]}(\tau)'V_b(\tau)z_b^{[1]}(\tau)\right)^{-1}z_b^{[1]}(\tau)'V_b(\tau)$$

$$z_b^{[1]} = \left(z_{b,1}^{[1]},...,z_{b,T}^{[1]}\right)', z_{b,t}^{[1]} = \left(1, \frac{t/T - \tau}{b}\right) \text{ with } V_b(\tau) = diag(\overline{w}_{b,1}\left(\tau\right),...,\overline{w}_{b,T}\left(\tau\right))$$

$$\text{where } \overline{w}_{b,t}\left(\tau\right) = w_b\left(t/T - \tau\right)/\lambda_{t/T} \text{ and } \lambda_{t/T} = \int_0^1 w_b\left(t/T - s\right)ds$$

where $w_b\left(.\right) = w\left(./b\right)/b$ and ι_T is a $(T \times 1)$ vector of ones. $w\left(.\right)$ is the Epanechnikov kernel $w\left(v\right) = 0.75\left(1 - v^2\right)1\{|\ v\ | \leq 1\}$ and b is the bandwidth with $v = x - \tau$ and $x = \{t/T\}_{t=1}^{T}$. Zhang et al. (2012) show that $\lambda_{t/T} = 1$ for $t/T \in [b, 1 - b]$ and is larger that $1/2$ otherwise.

When implementing density estimates, the choice of the kernel $w\left(.\right)$ and the bandwidth b are essential. Popular choices are based on Epanechnikov, Parzen or Tukey windows and b is based on the integrated square error or its expected value. Generally, $b = \phi T^{-1/5}$ where $\phi = 1.06\sigma_x$ or $\phi = 0.79R$ or $\phi = 0.9A$ where σ_x is the standard deviation of the series $x = \{t/T\}_{t=1}^{T}$, R is the interquartile range of x and $A = \min\left(\sigma_x, R/1.34\right)$. In our case, $\sigma_x = \sqrt{\frac{T^2-1}{12T^2}} \simeq 1/\sqrt{12}$. So, Zhang et al. (2012) choose $b = c\sqrt{1/12}T^{-1/5}$ with $c = 0.5, 1.0$ or 1.5. We can also define the bandwidth parameter b using unbiased cross-validation techniques (see Pagan and Ullah (1999)).

6.4. Appendix 4 - A bootstrapped nonparametric specification test for RCM_CT against OWEC_CST

The idea behind the test is the following. We use H_0 to denote the null hypothesis of the random intercept and slope coefficients model with semiparametric common trends (RCM_CT), against H_1:

the corresponding alternative is the random intercept and constant slope coefficients model with semiparametric country-specific trends (OWEC_CST).

$$\begin{cases} H_0: & y_{it} = \beta_1^P + X_{it}^P \beta^P + W_{it}^P \eta_i^P + f(t/T) + \eta_{1i}^P + \varepsilon_{it} \\ H_1: & y_{it} = \beta_1^P + X_{it}^P \beta^P + f_i(t/T) + \eta_{1i}^R + \varepsilon_{it} \end{cases}$$

Let $\hat{y}_{it} = \hat{\beta}_1^P + X_{it}^P \hat{\beta}^P + W_{it}^P \hat{\eta}_i^P + \hat{f}(t/T)$ and $\tilde{y}_{it} = \tilde{\beta}_1^P + X_{it}^P \tilde{\beta}^P + \tilde{f}_i(t/T)$ the fitted values of y_{it} under the null H_0 and the alternative H_1. The test statistic for testing H_0 is:

$$J_N = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{y}_{it} - \tilde{y}_{it})^2$$

Under H_0 (resp. H_1), J_N converges to 0 in probability (resp. to a positive constant). Since it is well known that asymptotic theory does not provide good approximations for nonparametric kernel-based tests in finite sample applications, we use the following bootstrap procedure inspired from Henderson et al. (2008) to approximate the finite sample null distribution of J_N.

1. Let $\hat{\xi}_{P,it} = \hat{\eta}_{1i}^P + \hat{\varepsilon}_{it}$, the augmented residuals from the estimation of the random intercept and slope coefficients model with semiparametric common trends (RCM_CT). Obtain the bootstrap error $\xi_{P,i}^*$ by random sampling with replacement from $\left\{ \widehat{\xi}_{P,i}, i = 1, 2, ..., N \right\}$ where $\widehat{\xi}_{P,i} = \left(\widehat{\xi}_{P,i1}, ..., \widehat{\xi}_{P,iT} \right)'$.

2. Generate the bootstrap analogue of y_{it} by holding $X_{it}^P, W_{it}^P, X_{it}^S, W_{it}^S$ as fixed: $y_{it}^* = \hat{\beta}_1^P + X_{it}^P \widehat{\beta}^P + W_{it}^P \widehat{\eta}_i^P + \widehat{f}(t/T) + \xi_{P,it}^*$ for $i = 1, ..., N$ and $t = 1, ..., T$ where $\xi_{P,it}^*$ is the tth element in the T-vector $\xi_{P,i}^*$.

3. Based on the bootstrap resample $\left\{ y_{it}^*, X_{it}^P, W_{it}^P, X_{it}^S, W_{it}^S \right\}$, first run the MFVB approximation of the RCM_CT model to obtain the bootstrapped fitted values residuals $\{ \hat{y}_{it}^* \}$. Second, run the MFVB approximation of the OWEC_CST model to obtain the bootstrapped fitted values residuals $\{ \tilde{y}_{it}^* \}$.

4. Based on $\{ \hat{y}_{it}^* \}$ and $\{ \tilde{y}_{it}^* \}$, compute the bootstrap test statistics J_N^* defined analogously to J_N but with $\hat{y}_{it}$ and $\tilde{y}_{it}$ being replaced by $\{ \hat{y}_{it}^* \}$ and $\{ \tilde{y}_{it}^* \}$.

5. Repeat step 2-4 for B times and index the bootstrap statistics as $\left\{ J_{N,l}^* \right\}_{l=1}^{B}$.

6. The bootstrap p-value is calculated by $p^* = B^{-1} \sum_{l=1}^{B} 1 \left\{ J_{N,l}^* > J_N \right\}$ where $1 \{.\}$ is the indicator function.

References

Abramowitz, M., 1956. Resource and output trends in the United States since 1870. American Economic Review 46, 2, 5–23.

Ai, C., Li, Q., 2008. Semi-parametric and non-parametric methods in panel data models, in: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice. Springer, pp. 451–478.

Alesina, A., Perotti, R., 1996. Income distribution, political instability and investment. European Economic Review 40, 1203–1228.

An, Y., Hsiao, C., Li, D., 2016. Semiparametric estimation of partially linear varying coefficient panel data models, in: González-Rivera, G., Carter Hill, R., Lee, T. (Eds.), Advances in Econometrics, Essays in Honor of Aman Ullah. Emerald Publishing Ltd, Bingley, UK. volume 36, pp. 47–65.

Anderson, H., Issler, J., Vahid, F., 2006. Common features. Journal of Econometrics 132, 1–15.

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

Bai, J., Ng, S., 2008. Large dimensional factor analysis, in: Greene, W. (Ed.), Foundations and Trends in Econometrics. volume 3, 2, pp. 89–163.

Baltagi, B.H., Griffin, J.M., 1988. A general index of technical change. Journal of Political Economy 96, 20–41.

Baltagi, B.H., Li, D., 2002. Series estimation of partially linear panel data models with fixed effects. Annals of Economics and Finance 3, 103–116.

Barro, R., 2000. Inequality and growth in a panel of countries. Journal of Economic Growth 5, 1, 5–32.

Barro, R., Sala-i Martin, X., 1992. Convergence. Journal of Political Economy 100, 223–251.

Barro, R., Sala-i Martin, X., 1997. Technological diffusion, convergence and growth. Journal of Economic Growth 2, 1–26.

Barro, R., Sala-i Martin, X., 2004. Economic Growth. MIT Press, 2nd edition, MA, USA.

Bassanini, A., Scarpetta, S., 2001. Does human capital matter for growth in OECD countries? Evidence from pooled mean-group estimates, in: OECD Economics Department Working Papers, No. 282, OECD Publishing, Paris, pp. 1–30.

Bassanini, A., Scarpetta, S., Hemmings, P., 2001. Economic growth: the role of policies and institutions. panel data evidence from OECD countries, in: OECD Economics Department Working Papers, No. 283, OECD Publishing, Paris, pp. 1–70.

Berlemann, M., Wesselhöft, J., 2012. Estimating aggregate capital stocks using the perpetual inventory method, new empirical evidence for 103 countries. Working paper No. 125. Department of Economics, Helmut Schmidt University, Hamburg.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer, New York, USA.

Blei, D., Kucukelbir, A., McAuliffe, J., 2017. Variational inference: A review for statisticians. Journal of the American Statistical Association 112, 518, 859–877.

Boneva, L., Linton, O., Vogt, M., 2015. A semiparametric model for heterogeneous panel data with fixed effects. Journal of Econometrics 188, 2, 327–345.

Bresson, G., Etienne, J.M., Mohnen, P., 2016. How important is innovation? A Bayesian factor-augmented productivity model on panel data. Macroeconomic Dynamics 20, 8, 1987–2009.

Brock, W., Durlauf, S., 2001. Growth economics and reality. World Bank Economic Review 15, 2, 229–272.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Cellini, R., 1997. Growth empirics: evidence from a panel of annual data. Applied Economic Letters 4, 347–351.

Chambers, D., Krause, A., 2010. Is the relationship between inequality and growth affected by physical and human capital accumulation? The Journal of Economic Inequality 8, 2, 53–72.

Chen, J., Gao, J., Li, D., 2012. Semiparametric trending panel data models with cross-sectional dependence. Journal of Econometrics 171, 1, 71–85.

Chen, J., Li, D., Gao, J., 2013. Non- and semi-parametric panel data models: a selective review. Working paper 18/13. Department of Econometrics and Business Statistics, Monash University, Australia.

Cingano, F., 2014. Trends in income inequality and its impact on economic growth, in: OECD SEM Working Paper 163, OECD Publishing, Paris, pp. 1–64.

Coelli, T.J., Rao, D.S.P., Battese, G.E., 1998. An Introduction to Efficiency and Productivity Analysis. Kluwer Academic Publishers, Boston, MA, USA.

Cornwell, C., Schmidt, P., Sickles, R.C., 1990. Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46, 185–200.

Durlauf, S., Johnson, P., Temple, J., 2005. Growth econometrics, in: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. Elsevier, Amsterdam, The Netherlands. volume 1, pp. 555–677.

Durlauf, S., Kourtellos, A., Minkin, A., 2001. The local Solow growth model. European Economic Review 45, 4-5, 928–940.

Duygun, M., Hao, J., Isaksson, A., Sickles, R., 2017. World productivity growth: a model averaging approach. Pacific Economic Review 22, 4, 587–619.

Easterly, W., Levine, R., 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, M., Bond, S., 2009. Cross-section dependence in nonstationary panel models: a novel estimator. MPRA Paper 01/2009. University Library of Munich, Germany.

Eberhardt, M., Teal, F., 2011. Econometrics for grumblers: a new look at the literature on cross-country growth empirics. Journal of Economic Surveys 25, 1, 109–155.

Faes, C., Ormerod, J., Wand, M., 2011. Variational Bayesian inference for parametric and non-parametric regression with missing predictor data. Journal of the American Statistical Association 106, 959–971.

Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031–1057.

Fan, J., Huang, T., Li, R., 2007. Analysis of longitudinal data with semiparametric estimation of covariance function. Journal of the American Statistical Association 102, 632–641.

Fan, J., Li, R., 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. Journal of the American Statistical Association 99, 710–723.

Fernandez, C., Ley, E., Steel, M., 2001. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics 16, 563–576.

De la Fuente, A., Domenech, R., 2006. Human capital in growth regressions: how much difference does data quality make? Journal of the European Economic Association 4, 1, 1–36.

Gao, J., Hawthorne, K., 2006. Semiparametric estimation and testing of the trend of temperature series. The Econometrics Journal 9, 332–355.

Griliches, Z., 1980. R&D and the productivity slowdown. The American Economic Review 70, 2, 343–348.

Henderson, D., Carroll, R., Li, Q., 2008. Nonparametric estimation and testing of fixed effects panel data models. Journal of Econometrics 144, 257–275.

Honaker, J., King, G., 2010. What to do about missing values in time-series cross-section data. American Journal of Political Science 54, 561–581.

Honaker, J., King, G., Blackwell, M., 2011. Amelia II: a program for missing data. Journal of Statistical Software 7, 45, 1–47.

Hsiao, C., 2003. Analysis of Panel Data. Econometric Society monograph n°34, 2nd edition, Cambridge University Press, New York.

Hsiao, C., 2015. Random coefficients models in panels, in: Baltagi, B. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, Oxford, UK, pp. 402–417.

Hsiao, C., Pesaran, M., 2008. Random coefficients models, in: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice. Springer, pp. 187–216.

Hsiao, C., Pesaran, M., Tahmiscioglu, A., 1999. Bayes estimation of short-run coefficients in dynamic panel data models, in: Hsiao, C., Lahiri, K., Lee, L.F., Pesaran, M. (Eds.), Analysis of Panels and Limited Dependent Variable Models. Cambridge University Press, Cambridge, UK, pp. 268–296.

Huang, J., Lu, T., 2017. Bayesian inference on partially linear mixed-effects joint models for longitudinal data with multiple features. Computational Statistics 32, 179–196.

Huang, J., Wu, C., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89, 1, 111–128.

Isaksson, A., Shang, C., Sickles, R., 2018. Non-structural analysis of productivity growth for the industrialized countries: a jackknife model averaging approach. Working paper, June, 7, 2018. Rice University, Houston, TX, USA.

Islam, N., 1995. Growth empirics: a panel data approach. Quarterly Journal of Economics 110, 4, 1127–1170.

Jeong, S., Park, T., 2016. Bayesian semiparametric inference on functional relationships in linear mixed models. Bayesian Analysis 11, 4, 1137–1163.

Kaldor, N., 1957. A model of economic growth. The Economic Journal 67, 268, 591–627.

Kappler, L., 2004. Determination of potential growth using panel techniques. ZEW discussion paper 04-69. Centre for European Economic Research, Mannheim University.

Kim, T., Heshmati, A., 2013. Economic Growth: The New Perspectives For Theory and Policy. Springer, New York.

Kneip, A., Sickles, R., Song, W., 2012. A new panel data treatment for heterogeneity in time trends. Econometric Theory 28, 590–628.

Kolev, G., Niehues, J., 2016. The inequality-growth relationship: an empirical reassessment. IW-Report 7/2016. Institut der deutschen Wirtschaft Koln, Germany.

Koop, G., Korobilis, D., 2018. Variational Bayes inference in high-dimensional time-varying parameter models. MPRA Paper No. 87972, 15 july. MPRA (Munich Personal RePEc Archive).

Kumar, S., Ullah, A., 2000. Semiparametric varying parameter panel data models: an application to estimation of speed of convergence, in: Fomby, T., Carter Hill, R. (Eds.), Advances in Econometrics, Applying Kernel and Nonparametric Estimation to Economic Topics. Emerald Publishing Ltd, Bingley, UK. volume 14, pp. 109–128.

Kumbhakar, S.C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46, 201–211.

Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89, 5, 841–864.

Lee, C.Y.Y., 2016. Fast Approximate Inference for Longitudinal and Multilevel Data Analysis. Ph.D. thesis. University of Technology Sydney, Australia.

Lee, C.Y.Y., Wand, M., 2016. Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. Biometrical Journal 58, 4, 868–895.

Lee, J., Robinson, P., 2015. Panel nonparametric regression with fixed effects. Journal of Econometrics 188, 2, 346–362.

Li, D., Chen, J., Gao, J., 2011. Non-parametric time-varying coefficient panel data models with fixed effects. The Econometrics Journal 14, 3, 387–408.

Li, H., Lin, Z., Hsiao, C., 2015. Testing purchasing power parity hypothesis: a semiparametric varying coefficient approach. Empirical Economics 48, 1, 427–438.

Li, K., Li, D., Liang, Z., Hsiao, C., 2017. Estimation of semi-varying coefficient models with nonstationary regressors. Econometric Reviews 36, 1-3, 354–369.

Li, Q., Hsiao, C., 1998. Testing serial correlation in semiparametric panel data models. Journal of Econometrics 87, 2, 207–237.

Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. Journal of Econometrics 112, 2, 295–325.

Li, Q., Li, T., 2015. Heterogeneity in panel data and in nonparametric analysis. Journal of Econometrics 188, 2, 346–362.

Li, Q., Racine, J., 2007. Nonparametric Econometrics. Princeton University Press, Princeton, New Jersey, USA.

Li, Q., Stengos, T., 1996. On the root-n-consistent semiparametric estimation of partially linear models. Journal of Econometrics 71, 389–397.

Li, Q., Ullah, A., 1998. Estimating partially linear panel data models with one-way error components. Econometric Reviews 17, 2, 145–166.

Li, W., 2012. Depreciation of business R&D capital. Technical Report. Bureau of Economic Analysis, U.S. Department of Commerce, New York.

Liu, J., Sickles, R., Tsionas, E., 2017. Bayesian treatments for panel data stochastic frontier models with time varying heterogeneity. Econometrics 5, 33, 1–21.

Liu, Z., Stengos, T., 1999. Non-linearities in cross-country growth regressions: a semiparametric approach. Journal of Applied Econometrics 14, 5, 527–538.

Mankiw, N., Romer, D., Weil, D., 1992. The contribution to the empirics of economic growth. Quarterly Journal of Economics 107, 2, 407–437.

Moral-Benito, E., 2012. Determinants of economic growth: a Bayesian panel data approach. Review of Economics and Statistics 94, 2, 566–579.

Neves, P., Afonso, O., Silva, S., 2016. A meta-analytic reassessment of the effects of inequality on growth. World Development 78, 386–400.

Ormerod, J., Wand, M., 2010. Explaining variational approximations. The American Statistician 64, 2, 140–153.

Ostry, J., Berg, A., Tsangarides, C., 2014. Redistribution, inequality, and growth. IMF Staff discussion Note, February. IMF, Washington.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, New York.

Park, B.U., Mammen, E., Lee, Y.K., Lee, E.R., 2015. Varying coefficient regression models: a review and new developments. International Statistical Review 83, 1, 36–64.

Park, B.U., Sickles, R., Simar, L., 1998. Stochastic panel frontiers: a semiparametric approach. Journal of Econometrics 84, 273–301.

Park, B.U., Sickles, R., Simar, L., 2003. Semiparametric-efficient estimation of AR(1) panel data models. Journal of Econometrics 117, 279–309.

Park, B.U., Sickles, R., Simar, L., 2007. Semiparametric-efficient estimation of dynamic panel data models. Journal of Econometrics 136, 281–301.

Perotti, R., 1996. Growth, income distribution, and democracy: what the data say. Journal of Economic Growth 1, 149–187.

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

Pesaran, M.H., 2007. A simple panel unit root test in the presence of cross-section dependence. Journal of Applied Econometrics 22, 265–312.

Pesaran, M.H., Shin, Y., Smith, R., 1999. Pooled mean group estimation of dynamic heterogeneous panels. Journal of the American Statistical Association 94, 621–634.

Pesaran, M.H., Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79–113.

Pham, T., Ormerod, J., Wand, M., 2013. Mean field variational Bayesian inference for nonparametric regression with measurement error. Computational Statistics and Data Analysis 68, 375–387.

Phillips, P., Sul, D., 2003. The elusive empirical shadow of growth convergence. Economics Working Papers. Economics Department, University of Aukland, New Zealand.

Qian, J., Wang, L., 2012. Estimating semiparametric panel data models by marginal integration. Journal of Econometrics 167, 483–493.

Qu, A., Li, R., 2006. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics 62, 379–391.

Ranganath, R., 2017. Black Box Variational Inference: Scalable, Generic Bayesian Computation and its Applications. Ph.D. thesis. Princeton University, NJ., USA.

Ranganath, R., Gerrish, S., Blei, D., 2014. Black box variational inference, in: Kaski, S., Corander, J. (Eds.), Artificial Intelligence and Statistics, PMLR. pp. 814–822.

Robinson, P., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 4, 931–954.

Robinson, P., 1989. Nonparametric estimation of time-varying parameters, in: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer, Berlin, Germany, pp. 164–253.

Robinson, P., 2012. Nonparametric trending regression with cross-sectional dependence. Journal of Econometrics 169, 1, 4–14.

Rodriguez-Poo, J., Soberon, A., 2017. Nonparametric and semiparametric panel data models: recent developments. Journal of Economic Surveys 31, 4, 923–960.

Ruppert, D., 2002. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 11, 4, 735–757.

Ruppert, D., Wand, M., Caroll, R., 2003. Semiparametric Regression. Cambridge University Press, New York, USA.

Sickles, R., Hao, J., Shang, C., 2015. Panel data and productivity measurement, in: Baltagi, B. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, Oxford, UK, pp. 517–547.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, London.

Solow, R., 1957. Technical change and the aggregate production function. Review of Economics and Statistics 39, 312–320.

Su, L., Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with fixed effects. Economics Letters 92, 75–81.

Su, L., Ullah, A., 2011. Nonparametric and semiparametric panel econometric models: estimation and testing, in: Ullah, A., Giles, D. (Eds.), Handbook of Empirical Economics and Finance. Taylor and Francis Group, New York, pp. 455–497.

Sun, Y., Carroll, R., Li, D., 2009. Semiparametric estimation of fixed effects panel data varying coefficient models, in: Li, Q., Racine, J. (Eds.), Advances in Econometrics, Nonparametric Econometric Methods. Emerald Publishing Ltd, Bingley, UK. volume 25, pp. 101–129.

Sun, Y., Hsiao, C., Li, Q., 2011. Measuring correlations of integrated but not cointegrated variables: a semiparametric approach. Journal of Econometrics 164, 2, 252–267.

Sun, Y., Hsiao, C., Li, Q., 2015a. Volatility spillover effect: a semiparametric analysis of non-cointegrated process. Econometric Reviews 34, 1-2, 127–145.

Sun, Y., Zhang, Y., Li, Q., 2015b. Nonparametric panel data regression models, in: Baltagi, B. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, Oxford, UK, pp. 285–324.

Temple, J., 1999. The new growth evidence. Journal of Economic Literature 37, 112–156.

Temple, J., 2001. Growth effects of education and social capital in the OECD countries, in: OECD Economic Studies, 33, 57-101, OECD Publishing, Paris, pp. 57–101.

Wand, M., Ormerod, J., 2008. On semiparametric regression with O'sullivan penalized splines. Australian and New Zealand Journal of Statistics 50, 179–198.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman and Hall/CRC Press, London.

Wand, M.P., Ripley, B.D., 2009. Kernsmooth 2.23. Functions for kernel smoothing. R package. R Core Team. URL http://www.R-project.org/.

Wang, L., Hsiao, C., 2011. Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. Journal of Econometrics 165, 1, 30–44.

Yoon, J., Zame, W., van der Schaar, M., 2017. Multi-directional recurrent neural networks: A novel method for estimating missing data, in: Precup, D., Teh, Y. (Eds.), Proceedings of Machine Learning Research, PMLR. pp. 3958–3966.

You, J., Zhou, X., Zhou, Y., 2010. Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors. Journal of Multivariate Analysis 101, 1079–1101.

Zao, Y., Staudenmayer, J., Coull, B., Wand, M., 2006. General design Bayesian generalized linear mixed models. Statistical Science 21, 35–51.

Zhang, J., Feng, S., Li, G., Lian, H., 2011. Empirical likelihood inference for partially linear panel data models with fixed effects. Economics Letters 113, 165–167.

Zhang, Y., Su, L., Phillips, P.C.B., 2012. Testing common trends in semi-parametric panel data model with fixed effects. The Econometrics Journal 15, 56–100.

Table 1 - CIPS unit root tests.

CIPS tests	logs		growth rate	
	intercepts	intercepts+trend	intercepts	intercepts+trend
Y/L	-1.749	-1.974	-4.729	-4.875
K/L	-0.806	-1.209	-2.765	-2.833
human capital	-1.118	-1.517	-3.736	-4.218
R&D intensity	-1.405	-1.112	-3.827	-3.909
patent applications	-1.849	-1.869	-5.453	-5.840
trademark applications	-1.827	-2.873	-5.997	-6.191
government expenditures	-2.045	-2.208	-5.138	-5.170
trade ratio	-2.207	-2.452	-5.252	-5.297
financial development	-2.185	-2.626	-5.857	-6.001
inflation			-3.568	-3.544
inequality	0.355	-1.674	-4.787	-5.240
life expectancy	-2.548	-2.563	-6.056	-6.320

critical values:

CIPS with individual-specific intercepts: -2.23 (1%), -2.11 (5%), -2.04 (10%)

CIPS with individual-specific intercepts and trend: -2.73 (1%), -2.61 (5%), -2.54 (10%)

Table 2 - Bootstrapped p-values of the nonparametric test for common trends.

bandwidth $b = \phi T^{-1/5}$		p-value	p-value
		H_0 : RCM_CT	H_0 : OWEC_CT
		vs H_1 : RCM_CST	vs H_1 : OWEC_CST
$\phi = 0.5\sigma_x$	$(b = 0.0067)$	0.82	0.04
$\phi = 0.9A$	$(b = 0.1213)$	0.89	0.01
$\phi = \sigma_x$	$(b = 0.1348)$	0.88	0.04
$\phi = 0.79R$	$(b = 0.1803)$	0.88	0.02
$\phi = 1.5\sigma_x$	$(b = 0.2022)$	0.81	0.06
UCV	$(b = 0.7782)$	0.84	0.03

R is the interquartile range of $x = \{t/T\}_{t=1}^{T}$, $A = \min(\sigma_x, R/1.34)$ with $\sigma_x = 1/\sqrt{12}$ being the standard deviation of the series x. UCV uses unbiased cross-validation to select the bandwidth.

RCM_CT (resp. RCM_CST): random intercept and slope coefficients with semiparametric common trends (resp. country-specific trends).

OWEC_CT (resp. OWEC_CST) : random intercept and constant slope coefficients with semiparametric common trends (resp. country-specific trends).

Table 3 - MCMC results for growth rate model for GDP per capita with semiparametric common trend.

$\Delta \log(Y/L)$	post.mean	post. sd	2.5%	97.5%	$\widehat{R}$
intercept	0.0302	0.0025	0.0252	0.0351	0.9994
$\Delta \log(K/L)$	0.3704	0.0342	0.3072	0.4416	0.9992
$\Delta \log(\text{human capital})$	0.3020	0.0901	0.1337	0.4852	0.9991
$\Delta \log(\text{R\&D intensity})$	-0.0097	0.0111	-0.0310	0.0128	0.9990
$\Delta \log(\text{patents})$	-0.0073	0.0044	-0.0154	0.0014	0.9997
$\Delta \log(\text{trademark})$	0.0061	0.0046	-0.0032	0.0146	1.0007
$\Delta \log(\text{gov.expenditures})$	-0.2406	0.0261	-0.2937	-0.1893	0.9997
$\Delta \log(\text{trade ratio})$	0.0405	0.0160	0.0110	0.0735	1.0002
inflation	-0.1050	0.0254	-0.1570	-0.0547	0.9993
$\Delta \log(\text{financial dvlpt.})$	0.0115	0.0030	0.0056	0.0170	0.9994
$\Delta \log(\text{inequality})$	0.0545	0.0344	-0.0102	0.1218	0.9995
$\Delta \log(\text{life expectancy})$	-0.1053	0.1464	-0.3866	0.1678	0.9991
dummy 1973-1975	-0.0090	0.0035	-0.0159	-0.0019	1.0017
dummy 2008-2009	-0.0133	0.0039	-0.0208	-0.0062	0.9992
common trend $\left\{ \frac{t}{T} \right\}_{t=1}^{T}$	-0.0237	0.0027	-0.0291	-0.0184	0.9993
$\sigma_{\varepsilon}^2 \times 10^{-3}$	0.1764	0.0087	0.1598	0.1937	0.9990

Table 3 gives posterior means, posterior standard deviations, 95% credible sets and Gelman-Rubin convergence diagnostics for β^P, β^S and σ_{ε}^2.

Table 4 - MFVB results for growth rate model for GDP per capita with semiparametric common trend.

$\Delta \log(Y/L)$	coeff.	sd	post.mean	post.sd	2.5%	97.5%	accuracy scores (%)
intercept	0.0300	0.0023	0.0300	0.0024	0.0251	0.0346	94
$\Delta \log(K/L)$	0.3710	0.0296	0.3716	0.0303	0.3151	0.4280	92
$\Delta \log$(human capital)	0.2926	0.0733	0.2916	0.0755	0.1438	0.4332	89
$\Delta \log$(R&D intensity)	-0.0088	0.0106	-0.0090	0.0106	-0.0293	0.0113	94
$\Delta \log$(patents)	-0.0074	0.0046	-0.0072	0.0047	-0.0166	0.0019	96
$\Delta \log$(trademark)	0.0062	0.0047	0.0063	0.0047	-0.0024	0.0156	98
$\Delta \log$(gov.expenditures)	-0.2409	0.0223	-0.2408	0.0226	-0.2856	-0.1964	93
$\Delta \log$(trade ratio)	0.0405	0.0152	0.0408	0.0154	0.0100	0.0723	95
inflation	-0.1046	0.0225	-0.1059	0.0222	-0.1501	-0.0622	94
$\Delta \log$(financial dvlpt.)	0.0115	0.0030	0.0115	0.0029	0.0058	0.0174	98
$\Delta \log$(inequality)	0.0543	0.0322	0.0540	0.0320	-0.0075	0.1177	94
$\Delta \log$(life expectancy)	-0.1023	0.1398	-0.1062	0.1389	-0.3755	0.1739	95
dummy 1973-1975	-0.0090	0.0037	-0.0090	0.0038	-0.0164	-0.0020	98
dummy 2008-2009	-0.0132	0.0036	-0.0133	0.0035	-0.0199	-0.0066	96
common trend $\left\{ \frac{t}{T} \right\}_{t=1}^{T}$	-0.0238	0.0026	-0.0238	0.0026	-0.0289	-0.0188	94
$\sigma_{\varepsilon}^2 \times 10^{-3}$	0.1702		0.1702	0.0075	0.1561	0.1858	71

Table 4 gives estimated coefficients, standard deviations, posterior means, posterior standard deviations, 95% credible sets and accuracy scores for β^P, β^S and σ_{ε}^2.
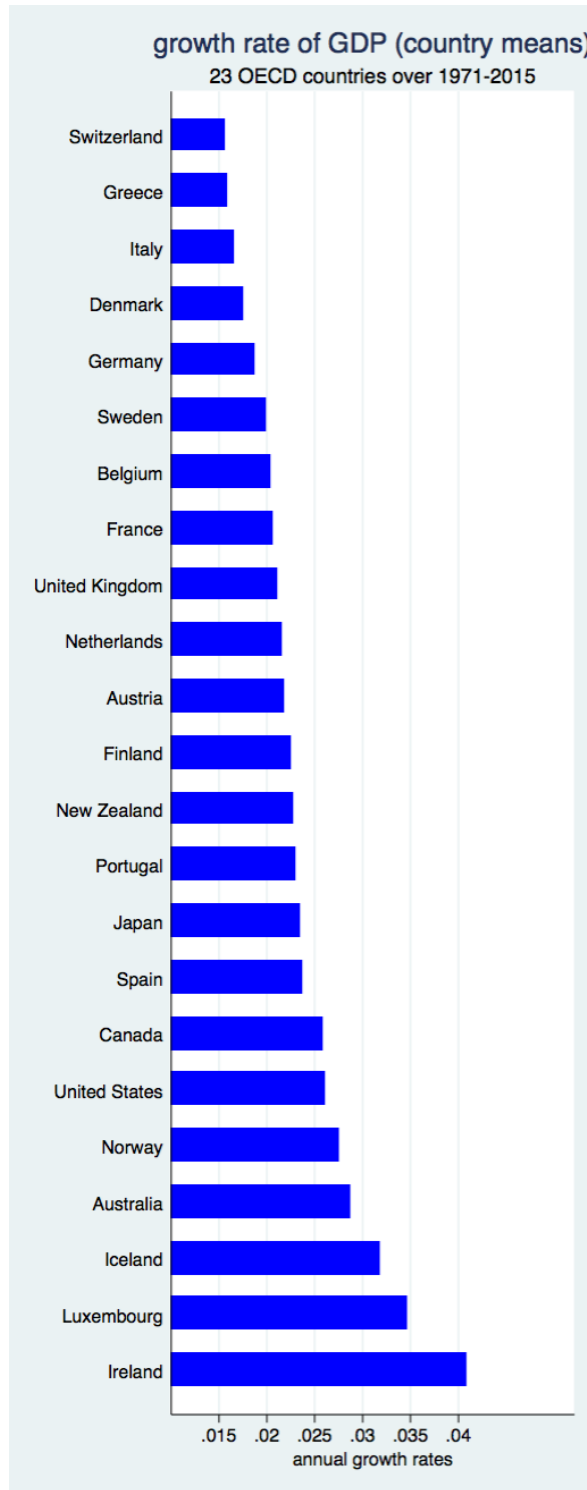
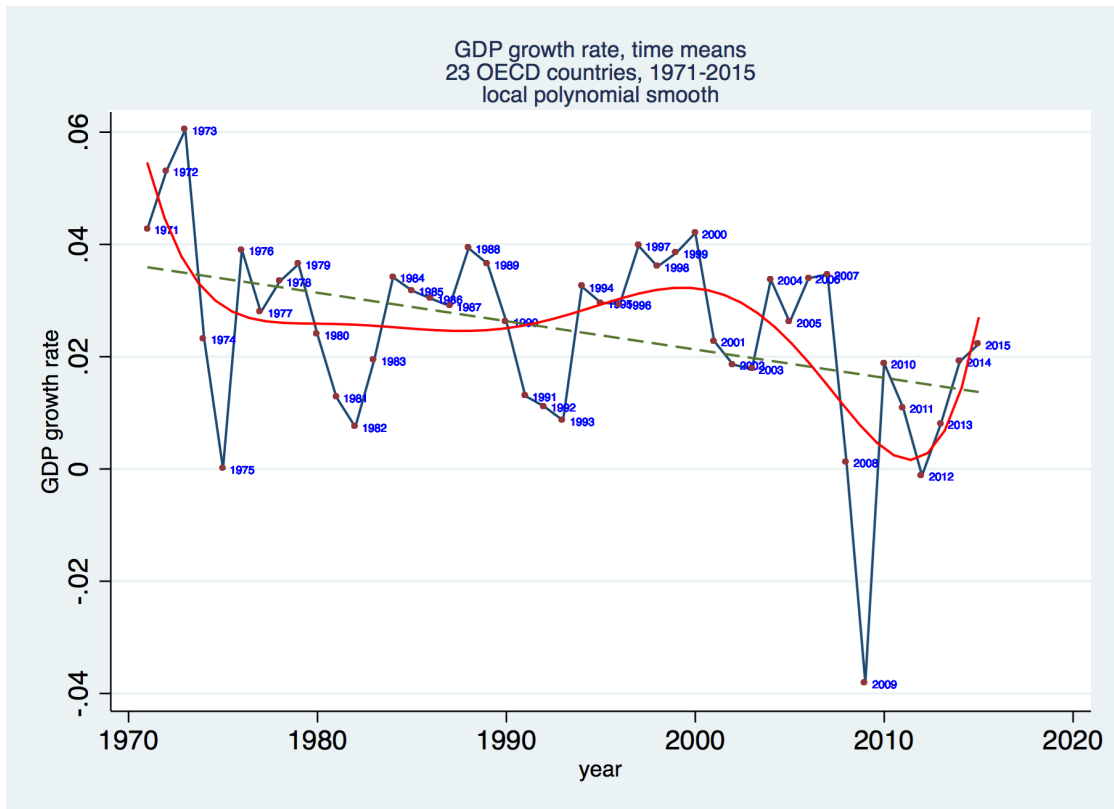Figure 1. Country means of the growth rate of GDP (23 OECD countries, 1971-2015).

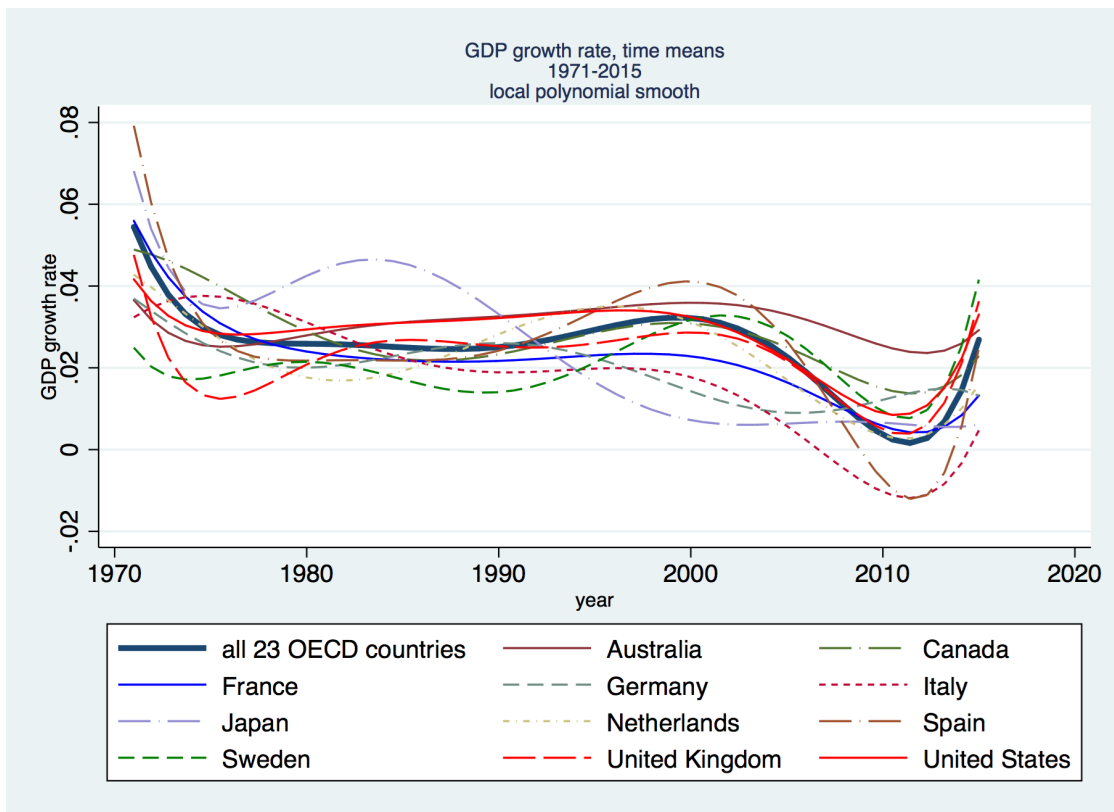Figure 2. Time means of the growth rate of GDP (23 OECD countries, 1971-2015).

Figure 3. Time means of the growth rate of GDP for some OECD countries (1971-2015).

Evidence Lower Bound (ELBO)
23 OECD countries 1971–2015
Random intercept and random slopes model, common trend
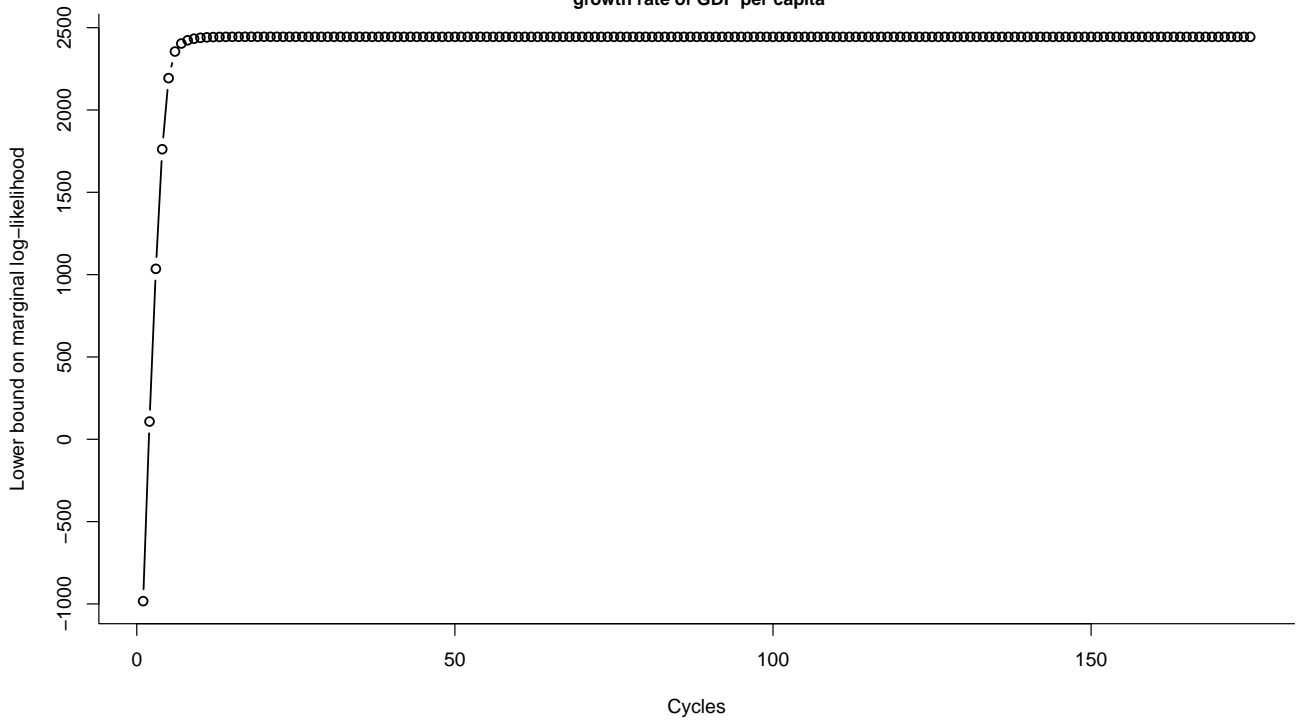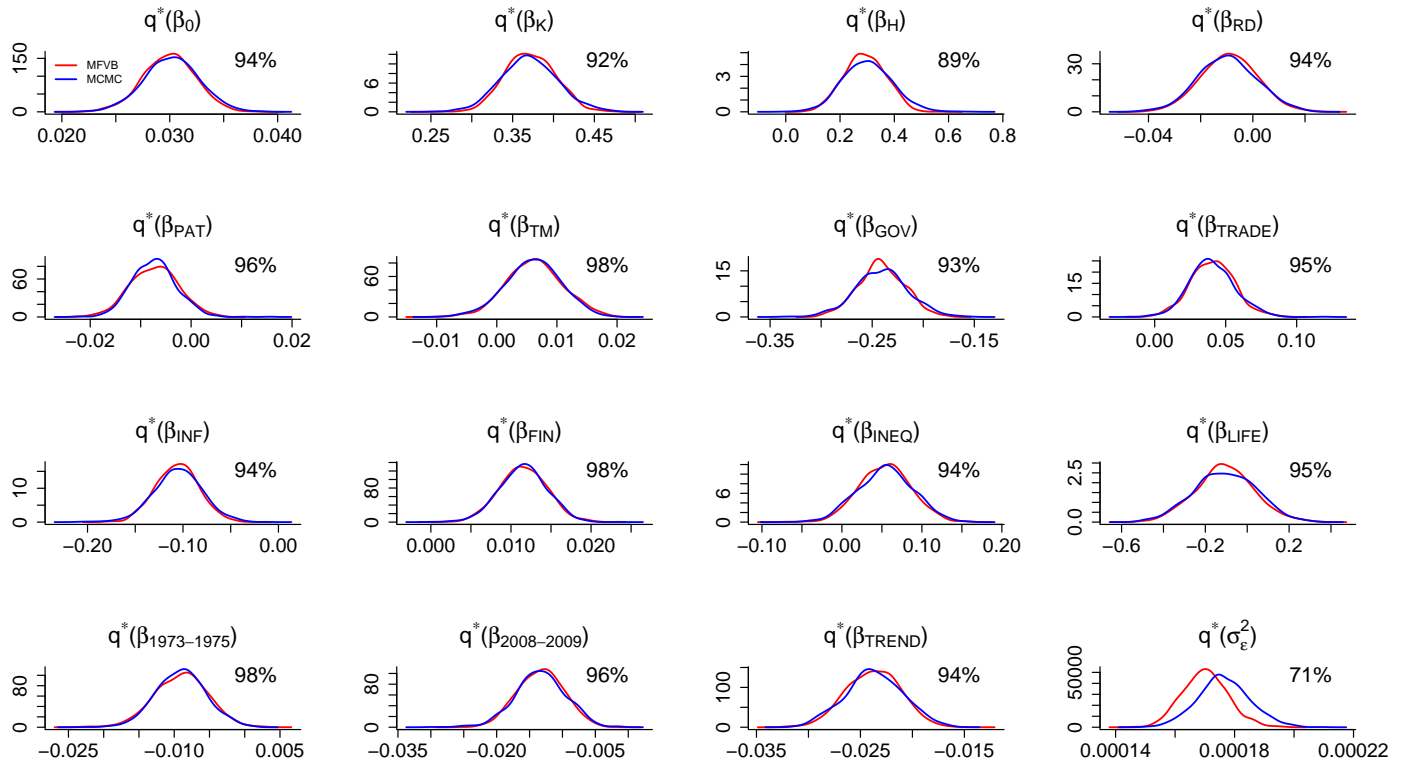**growth rate of GDP per capita**

Figure 4.

Figure 5. MCMC and MFVB approximate posterior density functions and accuracy scores (in percent) for β^P, β^S and σ_ε^2.

Figure 6. Time means of the MFVB spline fit of the growth rate of GDP per capita.