A Bayesian Panel Data Unit Root Test: Re-examining the Economic Convergence Hypothesis

Loukia Meligkotsidou[*]   Elias Tzavalis[†]   Ioannis D. Vrontos[‡]

Abstract

In this paper a Bayesian approach to unit root testing for panel data models is proposed based on the comparison of stationary autoregressive models with and without individual deterministic trends, with their counterpart models with a unit autoregressive root. This is done under cross-sectional dependence among the units of the panel. Simulation experiments are conducted with the aim to assess the performance of the suggested inferential procedure, as well as to investigate if the Bayesian model comparison approach can distinguish unit root models from stationary autoregressive models under or without cross-sectional depenendence. The approach is applied to real GDP data for a panel of G7.

JEL Classification: C11, C22, G10

Keywords: Autoregressive models, Bayesian inference, Cross-sectional dependence, Model comparison, Panel data, Unit root testing.

[*]Department of Mathematics, University of Athens, Panepistemiopolis, Athens 157 84, Greece. *Email*: meligots@math.uoa.gr

[†]Department of Economics, Athens University of Economics and Business, Pattision 76, Athens 134 76, Greece. *Email*: etzavalis@aueb.gr

[‡]Department of Statistics, Athens University of Economics and Business, Pattision 76, Athens 134 76, Greece. *Email*: vrontos@aueb.gr

1 Introduction

There is growing interest recently in developing methods of investigating the presence of a unit root or any other form of nonstationarity in the autoregressive model based on panel data sets.[1] This stems from the fact that panel data, unlike single time series models, can considerably improve the power of the suggested methods to distinguish the hypothesis of unit roots against its alternative of stationarity. This happens due to the fact that panel data exploit both the cross-sectional (N) and the time-series (T) dimensions of the data.

Most of the above methods are based on frequentist analysis. In this paper, we suggest a Bayesian model selection method with the aim to detect the presence of unit roots in panel data models. This method has several interesting features. Within our framework, the initial conditions of the model (known as nuisance parameters) can be treated as unknown parameters and can be intergated out. The same can be done for the individual effects of the cross-sectional units of the panel and their contemporaneous cross-sectional dependence. The latter has been the focus of many recent studies in the literature of panel unit roots[2]. As aptly is noted in these studies, ignoring the cross-section dependence may lead to erroneous inference about the presence of unit roots in panel data models. To account for cross-section dependence, the frequentist methods restrict the variance-covariance matrix of the error terms of the panel to be determined by spatial effects or by a number K of unobserved (or observed) common factors across the N units of the panel, with $K < N$. In the Bayesian framework, we can relax this assumption. Instead, we assume that the variance-covariance matrix consists of $N(N+1)/2$ free parameters.

In our model comparison setting, we calculate the marginal likelihood of the data under different stationary and non-stationary models. In our analysis we consider autoregressive panel data models of lag order 1 with or without individual deterministic trends and with or without cross-sectional dependence. Competing models are their counterpart models with a unit autoregressive root. For each of the competing models the calculation of the marginal likelihood is done by integrating the nuisance parameters of the model out of the unnormalised posterior density, in order to compute the posterior model probabilities. For the models with a unit autoregressive root, all the integrations with respect to the nuisance parameters can be performed analytically, hence the calculation of the marginal likelihood is exact. For the stationary models, all the parameters except for the autoregressive coefficient can be integrated out anallytically. The latter integration is not feasible due to the prior constraints that ensure stationarity, but it can be performed numerically via Monte Carlo integration. Finally, we can estimate the most probable model's parameters either using exact inference or by constructing a Markov chain Monte Carlo algorithm.

The paper proceeds as follows. In Section 2 we present our model specification and we discuss its alternative reduced forms. In Section 3 we describe the Bayesian model comparison method that we suggest for chosing the best model to describe the data. We show how to estimate the marginal

---

[1] See, e.g. Maddala and Wu (1999), Harris and Tzavalis (1999), Phillips and Moon (1999), Hadri (2000), Levin, Lin and Chu (2002), Im, Pessaran and Shin (2003) and Breitung (2002), *inter alia*.

[2] See, e.g. O' Connell (1998), Phillips and Sul (2003), Chang (2004), Moon and Perron (2004), Pessaran (2004) and Kapetanios (2005), *inter alia*.

likelihoods for the model specifications considered and hence obtain the posterior model probabilities. In Section 4 we report the results of our empirical applications. Finally, we conclude in Section 5.

2 Panel Data Models

In this section, we present the models we consider for the analysis of panel data. We consider a stationary autoregressive model, as well as some reduced specifications of it, under which we derive the likelihood of the observed data integrating out the initial observation. This is also done for the case of a unit autoregressive root.

2.1 Stationary Models

Consider having observations for N cross-sectional units on T time periods, $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $t = 1, \ldots, T$. The N-variate observation at time t is modelled as

$$\mathbf{y}_t = \delta + \beta t + \phi \left(\mathbf{y}_{t-1} - \delta - \beta(t-1) \right) + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \Sigma), \tag{1}$$

where $\delta = (\delta_1, \ldots, \delta_N)'$ is a vector of constant terms, $\beta = (\beta_1, \ldots, \beta_N)'$ is a vector of trend parameters and ϕ is an autoregressive coefficient, common across units. We assume that the multivariate process is stationary, that is $-1 < \phi < 1$. Furthermore, we assume that the errors $\mathbf{u}_t$ of the panel at time t follow a multivariate normal distribution with zero mean and covariance matrix Σ. This assumption allows the N cross-sectional units to be correlated.

The cross-sectional dependence assumption can be relaxed by taking $\Sigma = diag(\sigma_1^2, \ldots, \sigma_N^2)$, where $\sigma_1^2, \ldots, \sigma_N^2$ are unit-specific error variances. In this case model (1) reduces to N independent autoregressive models with common autoregressive coefficient. Finally, note that the general model specification (1) nests models which do not allow for the presence of deterministic trends, if we take $\beta = \mathbf{0}$.

$$\text{Let} \quad \tilde{\boldsymbol{Y}} = \begin{bmatrix} \mathbf{y}'_1 - \phi \mathbf{y}'_0 \\ \mathbf{y}'_2 - \phi \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_T - \phi \mathbf{y}'_{T-1} \end{bmatrix}, \quad \tilde{\boldsymbol{X}} = \begin{bmatrix} 1 - \phi & 1 - \phi + \phi \\ 1 - \phi & 2 - 2\phi + \phi \\ \vdots & \vdots \\ 1 - \phi & T - T\phi + \phi \end{bmatrix}, \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} \delta' \\ \beta' \end{bmatrix}.$$

Then, the likelihood of the observed data $\mathbf{y}_1, \ldots, \mathbf{y}_T$, conditional on the initial observation $\mathbf{y}_0$ and the model parameters, is given by

$$L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \boldsymbol{B}, \Sigma, \phi, \mathbf{y}_0) = (2\pi)^{-NT/2} |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left[\Sigma^{-1} (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{B})'(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{B}) \right] \right\}. \tag{2}$$

The initial condition of the panel series $\{\mathbf{y_t}\}$, $\mathbf{y}_0 = (y_{10}, \ldots, y_{N0})'$, is not observed and, hence, under the Bayesian approach it is treated as a further set of parameters. Treating $\mathbf{y}_0$ as random is appropriate for unit root testing as shown in Lubrano (1995). Alternatively, the initial condition can be treated as fixed, i.e. be given specific values, however this imposes a degree of subjectivity into the analysis. In unit root

testing, dependence on fixed initial conditions may affect inferences concerning the unit root hypothesis [see Zivot (1994)].

We assume that the initial observation comes from the stationary process (1). Under this assumption, we assign a multivariate normal prior distribution on $\mathbf{y}_0$, i.e. $\mathbf{y}_0 \sim N(\delta, (1-\phi)^{-1}\Sigma)$. We integrate the conditional likelihood (2) with respect to this prior distribution on $\mathbf{y}_0$ to obtain the, so called, exact likelihood of the observed data (see Box and Jenkins, 1976; Hamilton, 1994). The exact likelihood is given by

$$L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \boldsymbol{B}, \Sigma, \phi) = (2\pi)^{-NT/2} |\Sigma|^{-T/2} (1-\phi^2)^{N/2} \exp\left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}) \right] \right\}, \quad (3)$$

where

$$\boldsymbol{Y} = \begin{bmatrix} \mathbf{y}'_1 \sqrt{1-\phi^2} \\ \mathbf{y}'_2 - \phi\mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_T - \phi\mathbf{y}'_{T-1} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{X} = \begin{bmatrix} \sqrt{1-\phi^2} & \sqrt{1-\phi^2} \\ 1-\phi & 2-2\phi+\phi \\ \vdots & \vdots \\ 1-\phi & T - T\phi + \phi \end{bmatrix}.$$

Note that the likelihood for the model without deterministic trends is given by (3) with $\boldsymbol{X} = (\sqrt{1-\phi^2}, 1-\phi, \ldots, 1-\phi)'$ and $\boldsymbol{B} = \delta'$.

2.2 Case of a Unit Autoregressive Root

In the case that $\phi = 1$, model (1) reduces to a random walk (RW) model with drift, i.e.

$$\mathbf{y}_t = \beta + \mathbf{y}_{t-1} + \mathbf{u}_t, \quad \mathbf{u}_t \sim N(\mathbf{0}, \Sigma). \quad (4)$$

Again, if $\Sigma = diag(\sigma_1^2, \ldots, \sigma_N^2)$, model (4) is equivalent to assuming N independent RW processes for the cross-sectional units. Finally, if $\beta = \mathbf{0}$, model (4) reduces to an N-variate pure RW process (without drift), if there is cross-sectional dependence, or to N independent univariate RW processes if not.

Although in the case of a unit autoregressive root the parameter δ in (1) cancels, δ still contributes to the initial level of the process, since $\mathbf{y}_0 = \delta + \mathbf{x}_0$. Thus there is a need to make an assumption about $\mathbf{y}_0$. As stated in the Bayesian unit root literature [for example, see Zivot (1994)], it is preferable to treat the initial condition as random both in the stationary case and in the presence of a unit root. In this way we can treat the effect of the initial condition symmetrically in order to remove the dependence of inferences on $\mathbf{y}_0$. A class of prior distributions which have been proposed (see Uhlig, 1994, Zivot, 1994, and Jarocinski and Marcet, 2005, Meligkotsidou et. al, 2008) as an appropriate choice for univariate RW models is the class of S-type priors. In our setting, the S-prior takes the form $\pi(\mathbf{y}_0) \equiv N(\delta, S\Sigma)$. S is interpreted as the start-up time of the process before starting being observed. In some applications the value of S may be known, however in most cases it will be unknown. One possible choice is to assume that the process starts at $t = 0$, which implies that $S = 1$. However, better practice would be to consider different values of S and check the robustness of the results to the choice of S. Reasonable choices of S are those values which are of the same magnitude with the quantity $\frac{1}{1-\phi_1^2}$ which is used under stationarity (see Meligkotsidou et. al, 2008).

4

The exact likelihood function for model (4) having integrated $\mathbf{y}_0$ under the above prior is given by

$$L(\mathbf{y}_1,\ldots,\mathbf{y}_T \mid \boldsymbol{B},\Sigma) = (2\pi)^{-NT/2}|\Sigma|^{-T/2}(S+1)^{-N/2}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(\boldsymbol{Y_0}-\boldsymbol{X_0B})'(\boldsymbol{Y_0}-\boldsymbol{X_0B})\right]\right\},$$
(5)

where

$$\boldsymbol{Y_0} = \begin{bmatrix} \mathbf{y}'_1(S+1)^{-1/2} \\ \mathbf{y}'_2 - \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_T - \mathbf{y}'_{T-1} \end{bmatrix}, \quad \text{and} \quad \boldsymbol{X_0} = \begin{bmatrix} (S+1)^{-1/2} & (S+1)^{-1/2} \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}.$$

3 Calculation of the Marginal Likelihood

3.1 Bayesian Model Comparison

The objective of the Bayesian approach, in the model comparison setting, is to determine how probable one model is relative to another or various other alternative models. Consider K competing models $m_1,\ldots,m_K$. The posterior probability of model m_k, $k = 1,\ldots,K$, is given by

$$p(m_k|\mathbf{y}) = \frac{L(\mathbf{y}|m_k)p(m_k)}{\sum_{j=1}^{K}L(\mathbf{y}|m_j)p(m_j)},$$
(6)

where

$$L(\mathbf{y}|m_k) = \int L(\mathbf{y}|\theta_k,m_k)p(\theta_k|m_k)d\theta_k$$
(7)

is the marginal likelihood of the vector of observations $\mathbf{y}$ under model m_k, θ_k denotes the model specific parameter for model m_k, $L(\mathbf{y}|\theta_k,m_k)$ is the likelihood function given model m_k, $p(\theta_k|m_k)$ is a prior distribution on θ_k, and $p(m_k)$ is the prior probability of model m_k. In this paper we assign equal prior probability to all competing models. It can be easily seen from (7) that the marginal likelihood under model m_k is just the likelihood function integrated over the specified prior distribution for that model, provided that the integration is feasible. Equivalently, it can be seen as the normalizing constant of the posterior distribution of θ_k, defined as the integral of the product likelihood times prior, which is known as the unnormalized posterior.

The marginal likelihoods $L(\mathbf{y}|m_k)$ defined in (7) are in general difficult to calculate. Kass and Raftery (1995) provide an extensive description of available numerical strategies. However, if the prior specification is conjugate to the likelihood function, some of the model parameters can be integrated out of the posterior distribution analytically. Under our choice of prior for the parameters of the stationary autoregressive models all of the integrations in (7) can be calculated analytically, except for the integration of the autoregressive coefficient which is done via Monte Carlo integration. The marginal likelihood of the random walk models can be evaluated analytically.

3.2 Marginal Likelihood Calculation for the Stationary Model

In this section we show how the marginal likelihood of the stationary autoregressive model (1) can be computed. First we need to specify prior distributions for the model parameters. The prior specification is very crucial in model comparison, as the choice of prior can affect the marginal likelihoods of the different models considered. As a general principle, note that flat priors tend to penalize more the models which are more complex [see Bernardo and Smith (1994), chapter 6]. In the unit root problem, choosing an appropriate prior distribution is not an easy task. Sims (1988) used a flat prior as a non-informative prior for testing the unit root hypothesis in autoregressive models of order one, while Phillips (1991) argued that a flat prior is actually informative and proposed an ignorance prior (or Jeffreys invariant prior).

We adopt proper prior distributions for the parameters of model (1). Specifically, we assume a matrix-variate normal prior distribution for $\boldsymbol{B}$, which will be denoted by $N_{2 \times N}\left(\boldsymbol{B_0}, \mathbf{H}, \boldsymbol{\Sigma}\right)$, where $\boldsymbol{B_0}$ is the $2 \times N$ matrix of means and $\mathbf{H}$ is a 2×2 symmetric positive definite matrix. The probability density function (pdf) of the matrix-variate normal prior distribution for $\boldsymbol{B}$ is given by

$$\pi(\boldsymbol{B}|\boldsymbol{\Sigma}) = (2\pi)^{-N}|\mathbf{H}|^{-N/2}|\boldsymbol{\Sigma}|^{-1}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{B}-\boldsymbol{B_0}\right)'\mathbf{H}^{-1}\left(\boldsymbol{B}-\boldsymbol{B_0}\right)\right]\right\}.$$

Now, for $\boldsymbol{\Sigma}$ we assume an inverted Wishart prior distribution, which will be denoted by $IW(\mathbf{Q}, v)$, where $\mathbf{Q}$ is a $N \times N$ hyperparameter matrix and v denotes the degrees of freedom. The pdf of the inverted Wishart prior distribution for $\boldsymbol{\Sigma}$ is given by

$$\pi(\boldsymbol{\Sigma}) = \left[2^{vN/2}\pi^{N(N-1)/4}\prod_{i=1}^{N}\Gamma\left(\frac{v+1-i}{2}\right)|\mathbf{Q}|^{-v/2}\right]^{-1}|\boldsymbol{\Sigma}|^{-(v+N+1)/2}\exp\left[-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\mathbf{Q}\right)\right],$$

where $v \geq N$. Finally, let $p(\phi)$ be an appropriate prior distribution on the autoregressive coefficient with support in the stationary region. Since in this paper we are dealing with unit root testing, we expect that in such applications the value of ϕ will be close to unity. Hence we consider a Beta prior on ϕ, i.e. $\phi \sim Be(a, b)$, concentrated to a region near unity.

The unnormalized posterior of the parameters of the stationary autoregressive model (1) can be integrated analytically with respect to the parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ under the above prior distributions. This results to the unnonrmalised marginal posterior of ϕ, i.e. $L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi)p(\phi)$, which is given by

$$\pi^{-\frac{NT}{2}}|\mathbf{K}|^{-\frac{N}{2}}|\mathbf{H}|^{-\frac{N}{2}}|\mathbf{Q}|^{\frac{v}{2}}|\mathbf{C}-\mathbf{M}'\mathbf{K}^{-1}\mathbf{M}+\mathbf{Q}|^{-\frac{T+v}{2}}\left\{\prod_{i=1}^{N}\frac{\Gamma\left((T+v+1-i)/2\right)}{\Gamma\left((v+1-i)/2\right)}\right\}(1-\phi^2)^{\frac{N}{2}}p(\phi), \quad (8)$$

where $\mathbf{M} = \boldsymbol{X}'\boldsymbol{Y} + \mathbf{H}^{-1}\boldsymbol{B_0}$, $\mathbf{K} = \boldsymbol{X}'\boldsymbol{X} + \mathbf{H}^{-1}$ and $\mathbf{C} = \boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{B_0}'\mathbf{H}^{-1}\boldsymbol{B_0}$.

The marginal likelihood of model (1) can be obtained by Monte Carlo integration of $L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi)p(\phi)$ with respect to ϕ. Following Kass and Raftery (1995), we can use Monte Carlo integration with importance sampling to evaluate the integral $\int L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi)p(\phi)d\phi$. After specifying an importance function, with domain in the stationary region $(-1, 1)$, denoted as $g(\phi)$, the importance weight of a random draw $\phi^{(j)}$ from $g(\phi)$ is defined as $w^{(j)} = p(\phi^{(j)})/g(\phi^{(j)})$. Then, an estimate of the marginal

likelihood $L(\mathbf{y}_1, \ldots, \mathbf{y}_T)$, denoted as $\hat{L}(\mathbf{y}_1, \ldots, \mathbf{y}_T)$, can be calculated as

$$\hat{L}(\mathbf{y}_1, \ldots, \mathbf{y}_T) = \frac{\sum_{j=1}^{J} w^{(j)} L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi^{(j)})}{\sum_{j=1}^{J} w^{(j)}},$$

with all normalizing constants included, where $\phi^{(j)}$ $j = 1, \ldots, J$, denotes a sample of J draws from the importance function $g(\phi)$.

The importance function must be carefully chosen in order to obtain an importance sampling scheme which is effective and efficient, thus leading to accurate Monte Carlo estimates of the integral of interest. The effective sample size (ESS) of a weighted sample of J draws from an importance sampling scheme is a measure of the efficiency of the scheme which is defined as

$$ESS(J) = \frac{J}{1 + var_g(w)},$$

where $var_g(w)$ is the variance of the normalized weights which can be estimated by the coefficient of variation of the unnormalized weights [see Liu (2001)]. The ESS shows how close to the target the importance function is. A sample of size J from a given importance function g corresponds to $ESS(J)$ i.i.d. draws from the target. Hence, the closer to J the $ESS(J)$ is, the more effective the importance sampling scheme is. A measure of the numerical accuracy of the Monte Carlo integration is the relative numerical efficiency (RNE) defined as

$$RNE = \frac{\sum_{j=1}^{J} w^{(j)} [L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi^{(j)}) - \hat{L}(\mathbf{y}_1, \ldots, \mathbf{y}_T)]^2}{\sum_{i=1}^{N} (w^{(i)})^2 [L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi^{(j)}) - \hat{L}(\mathbf{y}_1, \ldots, \mathbf{y}_T)]^2}.$$

Values of the RNE close to one show that the Monte Carlo integration is particularly accurate [see Geweke (1989), and DeJong and Whiteman (1996)].

In the case of an autoregressive model without deterministic trends, i.e. if in model (1) $\beta = \mathbf{0}$, we assume a multivariate normal prior on δ, i.e. $\delta \sim N(\xi, \tau^2 \Sigma)$. Again, all of the model parameters but ϕ can be integrated analytically and the above Monte Carlo approach can be followed in order to perform numerically the integration with respect to ϕ. The unnormalised marginal posterior of ϕ in this case is given by

$$\pi^{-\frac{NT}{2}} \left[1 + \tau^2 (1 - \phi)^2 (T - 1) + \tau^2 (1 - \phi^2) \right]^{-\frac{N}{2}} |\mathbf{Q}|^{\frac{v}{2}} |\mathbf{A} + \mathbf{Q}|^{-\frac{T+v}{2}} \left\{ \prod_{i=1}^{N} \frac{\Gamma\left((T + v + 1 - i)/2\right)}{\Gamma\left((v + 1 - i)/2\right)} \right\} (1 - \phi^2)^{\frac{N}{2}} p(\phi),$$

where

$$
\begin{aligned}
\mathbf{A} &= \tau^{-2} \xi \xi' + \mathbf{Y}_{(-1)}' \mathbf{Y}_{(-1)} + (1 - \phi^2) \mathbf{y_1} \mathbf{y_1}' - (\tau^{-2} + (1 - \phi)^2 (T - 1) + (1 - \phi^2))^{-1} \\
&\times \left(\tau^{-2} \xi + (1 - \phi) \mathbf{Y}_{(-1)}' \mathbf{1}_{T-1} + (1 - \phi^2) \mathbf{y_1} \right) \left(\tau^{-2} \xi + (1 - \phi) \mathbf{Y}_{(-1)}' \mathbf{1}_{T-1} + (1 - \phi^2) \mathbf{y_1} \right)',
\end{aligned}
$$

with $\mathbf{Y}_{(-1)}$ being the matrix $\mathbf{Y}$ excluding the first row, and $\mathbf{1}_{T-1}$ being the $(T - 1) \times 1$ vector of ones.

For the case of independent errors across units we assume independent priors for the parameters of the model for each cross-sectional unit. Specifically, we use a $N(\xi_{1i}, \tau_{1i}^2 \sigma_i^2)$ prior for δ_i, a $N(\xi_{2i}, \tau_{2i}^2 \sigma_i^2)$

prior for β_i and an inverted Gamma $IG(c_i, d_i)$ prior for σ_i^2, $i = 1, \ldots, N$. Under this prior specification, $L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi)$ can be calculated analytically and it is given by a product of the form

$$L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \phi) = \prod_{i=1}^{N} L(y_{i1}, \ldots, y_{iT} \mid \phi),$$

where $L(y_{i1}, \ldots, y_{iT} \mid \phi) = \int L(y_{i1}, \ldots, y_{iT} \mid \delta_i, \beta_i, \sigma_i^2, \phi) d\delta_i d\beta_i d\sigma_i^2$ if deterministic trends are present in the model and $L(y_{i1}, \ldots, y_{iT} \mid \phi) = \int L(y_{i1}, \ldots, y_{iT} \mid \delta_i, \sigma_i^2, \phi) d\delta_i d\sigma_i^2$ if not (for details on these calculations, see Meligkotsidou et. al., 2008).

3.2.1 Calculation of the Marginal Likelihood for the Random Walk Model

For the RW model (4) the priors on the model parameters $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are identical to those for the stationary model's parameters. In the case of a unit autoregressive root the unnormlised posterior distribution $L(\mathbf{y}_1, \ldots, \mathbf{y}_T \mid \boldsymbol{B}, \boldsymbol{\Sigma}) p(\boldsymbol{B} \mid \boldsymbol{\Sigma}) p(\boldsymbol{\Sigma})$ can be integrated analytically with respect to the model parameters and the marginal likelihood is given by

$$\pi^{-\frac{NT}{2}} |\mathbf{K_0}|^{-\frac{N}{2}} |\mathbf{H}|^{-\frac{N}{2}} |\mathbf{Q}|^{\frac{v}{2}} |\mathbf{C_0} - \mathbf{M_0'} \mathbf{K_0^{-1}} \mathbf{M_0} + \mathbf{Q}|^{-\frac{T+v}{2}} \left\{ \prod_{i=1}^{N} \frac{\Gamma\left((T+v+1-i)/2\right)}{\Gamma\left((v+1-i)/2\right)} \right\} (S+1)^{-\frac{N}{2}}, \quad (9)$$

where $\mathbf{M_0} = \boldsymbol{X_0'} \boldsymbol{Y_0} + \mathbf{H}^{-1} \boldsymbol{B_0}$, $\mathbf{K_0} = \boldsymbol{X_0'} \boldsymbol{X_0} + \mathbf{H}^{-1}$ and $\mathbf{C_0} = \boldsymbol{Y_0'} \boldsymbol{Y_0} + \boldsymbol{B_0'} \mathbf{H}^{-1} \boldsymbol{B_0}$.

In the case of a RW model without drift a $N(\xi, \tau^2 \Sigma)$ prior is assumed for δ, as in the stationary case, and the marginal likelihood of the model is given by

$$\pi^{-\frac{NT}{2}} |\mathbf{Q}|^{\frac{v}{2}} |\boldsymbol{Y_{0(-1)}'} \boldsymbol{Y_{0(-1)}} + (\tau^2 + S + 1)^{-1} (\mathbf{y_1} - \xi)(\mathbf{y_1} - \xi)' + \mathbf{Q}|^{-\frac{T+v}{2}} \left\{ \prod_{i=1}^{N} \frac{\Gamma\left((T+v+1-i)/2\right)}{\Gamma\left((v+1-i)/2\right)} \right\} (\tau^2 + S + 1)^{-\frac{N}{2}},$$

with $\boldsymbol{Y_{0(-1)}}$ being the matrix $\boldsymbol{Y_0}$ excluding the first row. Finally, if the errors across units are assumed to be independent, similarly to the stationary case the marginal likelihood is a product of the form

$$L(\mathbf{y}_1, \ldots, \mathbf{y}_T) = \prod_{i=1}^{N} L(y_{i1}, \ldots, y_{iT}),$$

where $L(y_{i1}, \ldots, y_{iT}) = \int L(y_{i1}, \ldots, y_{iT} \mid \delta_i, \beta_i, \sigma_i^2) d\delta_i d\beta_i d\sigma_i^2$ in the case of independent univariate RW models with drift and $L(y_{i1}, \ldots, y_{iT}) = \int L(y_{i1}, \ldots, y_{iT} \mid \beta_i, \sigma_i^2) d\beta_i d\sigma_i^2$ in the case of pure RW models (for details on these calculations, see Meligkotsidou et. al., 2008).

4 Applications to macroeconomic data

In this section we present an empirical application of the proposed Bayesian model comparison approach to unit root testing in panel data models. Our empirical study is concerned with the analysis of annual GDP data concerning a panel of G7 countries for a period from 1970 to 2004[3]. The group of seven (G7) countries consists of Canada, France, Germany, Italy, Japan, United Kingdom and United States. This group of countries is of particular interest since it is a representative subset of the OECD, which consists of the leading economies worldwide and which has formed a forum that deals with major economic issues.

---

[3]This data has been obtained from the Penn World Table of Alan Heston, Robert Summers and Bettina Aten, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.
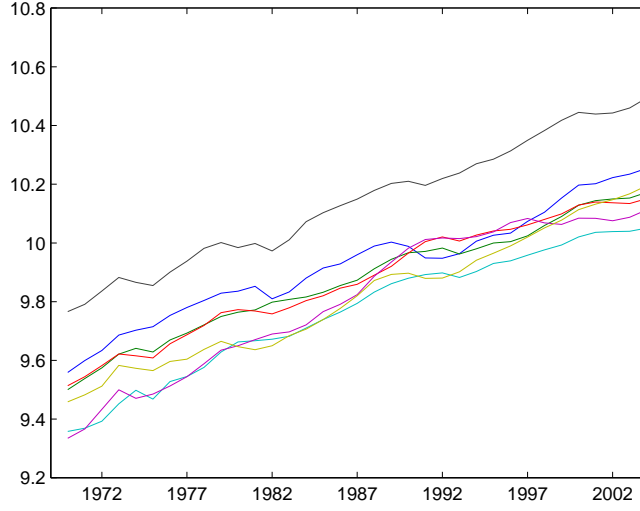
Figure 1: GDP series for G7 countries.

The aim of our empirical application is to detect the presence of a unit autoregressive root in GDP panel data, as well as to examine whether there is cross-sectional dependence among the G7 countries' GDP and estimate how strong this dependence is. It is sensible to believe that leading economies are likely to affect each other and therefore expect that there will be correlation of GDP among the G7. Indeed, looking at the plot of the analysed series (Figure 1) it can be seen that all countries follow a similar increasing pattern. In our Bayesian model comparison setting we consider the following competing models:

$$m_1 : y_{it} = \delta_i + \phi(y_{i,t-1} - \delta_i) + u_{it},\, u_{it} \sim N(0, \sigma_i^2),\, i = 1, \ldots, N$$
$$m_2 : \mathbf{y}_t = \delta + \phi\left(\mathbf{y}_{t-1} - \delta\right) + \mathbf{u}_t,\, \mathbf{u}_t \sim N(\mathbf{0}, \Sigma)$$
$$m_3 : y_{it} = \delta_i + \beta_i t + \phi(y_{i,t-1} - \delta_i - \beta_i(t-1)) + u_{it},\, u_{it} \sim N(0, \sigma_i^2),\, i = 1, \ldots, N$$
$$m_4 : \mathbf{y}_t = \delta + \beta t + \phi\left(\mathbf{y}_{t-1} - \delta - \beta(t-1)\right) + \mathbf{u}_t,\, \mathbf{u}_t \sim N(\mathbf{0}, \Sigma)$$
$$m_5 : y_{it} = y_{i,t-1} + u_{it},\, u_{it} \sim N(0, \sigma_i^2),\, i = 1, \ldots, N$$
$$m_6 : \mathbf{y}_t = \mathbf{y}_{t-1} + \mathbf{u}_t,\, \mathbf{u}_t \sim N(\mathbf{0}, \Sigma)$$
$$m_7 : y_{it} = \beta_i + y_{i,t-1} + u_{it},\, u_{it} \sim N(0, \sigma_i^2),\, i = 1, \ldots, N$$
$$m_8 : \mathbf{y}_t = \beta + \mathbf{y}_{t-1} + \mathbf{u}_t,\, \mathbf{u}_t \sim N(\mathbf{0}, \Sigma)$$

The stationary models $m_1 - m_3$ can be nested in the general AR(1) panel data model m_4, while the RW models $m_5 - m_7$ are nested in model m_8. For the parameters of the above competing models we use prior distributions of the type described in section 3. For the more general models, m_4 and m_8, we assume that $\boldsymbol{B} \sim N_{2 \times N}\left(\boldsymbol{B_0}, \mathbf{H}, \boldsymbol{\Sigma}\right)$, with the second row of the matrix $\boldsymbol{B_0}$ being equal to $\mathbf{0}$, the first row being a vector of arbitrary logical values in the domain of the data and $\mathbf{H} = diag(\tau_1^2, \tau_2^2)$, where $\tau_1 = 30$ and $\tau_2 = 0.1$. Furthermore, we assume that $\boldsymbol{\Sigma} \sim IW(\mathbf{Q}, v)$, with $v = N + 2$ and $\mathbf{Q}$ taken equal

9

to the sample covariance matrix in order to replicate the covarince structure of the data. Finally, for the autoregressive coefficient we adopt a Beta(a, b) prior, where $a = 5$ and $b = 0.5$, which is concentrated in a region near unity and results in the marginal likelihood of model m_1 being well defined. For the models which assume cross-sectional dependence but do not allow for deterministic trends, $\boldsymbol{B} = \delta'$ and we assign a $N(\xi, \tau^2 \Sigma)$ to δ, with $\xi = \mathbf{0}$ and $\tau = 30$. For the case of independent errors across units we assume that $\delta_i \sim N(\xi_{1i}, \tau_{1i}^2 \sigma_i^2)$, $\beta_i \sim N(\xi_{2i}, \tau_{2i}^2 \sigma_i^2)$ and $\sigma_i^2 \sim IG(c_i, d_i)$, $i = 1, \ldots, N$, with ξ_{1i} being an arbitrary logical value in the domain of the data from the ith country, $\tau_{1i} = 30$, $\xi_{2i} = 0$, $\tau_{2i} = 0.1$ and $c_i = d_i = 0.01$. Finally, we consider different values of S for the prior of $\mathbf{y}_0$ under the RW models, namely $S = 1, 5, 10, 50, 100$.

For all different values of S the most probable model, with posterior probability of about 0.55, is the RW model with drift allowing for cross-sectional dependence (model m_8). Under this model, the marginal posterior distribution of $\boldsymbol{\Sigma}$ is $IW(\mathbf{Q}^*, T + v)$, where $\mathbf{Q}^* = \mathbf{C_0} - \mathbf{M_0'} \mathbf{K_0^{-1}} \mathbf{M_0}$. The posterior mean of $\boldsymbol{\Sigma}$ is given by $E(\boldsymbol{\Sigma}) = (T + \nu - N - 1)^{-1} \mathbf{Q}^*$ while the posterior variance of the ij element of $\boldsymbol{\Sigma}$, denoted by σ_{ij} is

$$Var(\sigma_{ij}) = \frac{(T + \nu - N + 1)q_{ij}^{*2} + (T + \nu - N - 1)q_{ii}^* q_{jj}^*}{(T + \nu - N)(T + \nu - N - 1)^2(T + \nu - N - 3)},$$

where q_{ij}^* is the ij element of $\mathbf{Q}^*$. In Table 1 are shown the posterior means and standard deviations (in parenthesis) of the variances and covariances of the G7 countries' GDP, based on the RW model with drift assuming cross-sectional dependence. From the posterior mean of the variance-covariance matrix, an estimate of the correlation matrix can be easily obtained (see Table 2). It can be seen that all the correlations are positive and high, with a minimum of 0.72 between Canada and Japan and a maximum of 0.94 between Germany and France and between Italy and France. Our results indicate that there is high positive correlation of growth among the leading economies worldwide. This can be explained by the fact that ups and downs in one economy are transmitted to the other economies.

	Canada	France	Germany	Italy	Japan	UK	US
Canada	0.0014						
	(0.00034)						
France	0.0011	0.0012					
	(0.00029)	(0.00030)					
Germany	0.0011	0.0012	0.0013				
	(0.00030)	(0.00031)	(0.00033)				
Italy	0.0012	0.0013	0.0014	0.0016			
	(0.00033)	(0.00033)	(0.00035)	(0.00039)			
Japan	0.0013	0.0014	0.0016	0.0016	0.0022		
	(0.00037)	(0.00038)	(0.00041)	(0.00043)	(0.00054)		
UK	0.0014	0.0013	0.0013	0.0014	0.0016	0.0017	
	(0.00035)	(0.00033)	(0.00034)	(0.00037)	(0.00043)	(0.00040)	
US	0.0014	0.0012	0.0013	0.0014	0.0016	0.0016	0.0017
	(0.00036)	(0.00033)	(0.00035)	(0.00037)	(0.00044)	(0.00039)	(0.00042)

Table 1: Posterior means and standard deviations (in parenthesis) of the variances and covariances for the G7 countries' GDP, based on the RW model with drift and cross-sectional dependence.

	Canada	France	Germany	Italy	Japan	UK	US
Canada	1.00						
France	0.83	1.00					
Germany	0.79	0.94	1.00				
Italy	0.82	0.94	0.92	1.00			
Japan	0.72	0.88	0.92	0.87	1.00		
UK	0.88	0.90	0.86	0.86	0.84	1.00	
US	0.92	0.86	0.88	0.84	0.84	0.93	1.00

Table 2: Estimated correlations for the G7 countries' GDP, based on the RW model with drift and cross-sectional dependence.

5   Conclusions

In this paper, we have suggested a Bayesian approach to comparing stationary autoregressive panel models with and without individual deterministic trends, to their counterpart models with unit autoregressive roots. We have also allowed for cross-sectional dependence among the units of the panel. This dependence

seems to be reasonable and consistent with evidence from financial series. It can be attributed to the fact that changes to one economy affect other economies as well.

The Bayesian approach that we have followed in the paper employs analytic and Monte Carlo integration techniques for calculating the marginal likelihood of the panel models, which is necessary for calculating the posterior model probabilities. The empirical application of our method to GDP series of the G7 countries has shown that the Bayesian approach favors the RW model with drift and cross-sectional dependence among the units.

References

Arellano, M. (2003), Panel Data Econometrics, Oxford University Press.

Baltagi, B.H. and C. Kao (2000), "Nonstationary panels, cointegration in panels and dynamic danels, a survey" in B. Baltagi (ed.) Nonstationary Panels, Panel Cointegration and Dynamic Panels, *Advances in Econometrics*, 14, 7-52.

Bernardo, J.M., and A.F.M. Smith (1994). *Bayesian Theory*, Wiley Series in Applied Probability and Statistics.

Box, J.E.P., and G.M. Jenkins (1976). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.

Breitung, J. (2002), "Nonparametric tests for unit roots and cointegration", *Journal of Econometrics*, 108, 343-363.

Chang, Y. (2004), "Bootstrap unit root tests in panels with cross-sectional dependence", *Journal of Econometrics*, 120, 263-293.

DeJong, D.N., and C.H. Whiteman (1996). Bayesian Inference in Dynamic Equilibrium Models: An Application to the Rational Expectations Model of the Term Structure. *In Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, ed. D.A. Berry, K.M. Chaloner and J.K. Geweke, pp. 141-152, Wiley, New York.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**, 1317-1339.

Hadri, K. (2000), "Testing for stationarity in heterogenous panel data", *Econometrics Journal*, 3, 148-161.

Hamilton, J.D (1994). *Time Series Analysis*. Princeton University Press.

Harris, R.D.F. and F. Tzavalis (1999), "Inference for unit roots in dynamic panels where the time dimension is fixed", *Journal of Econometrics, 91, 201-226.*

Im, K.S., M.H. Pesaran and Y. Shin (2003), "Testing for unit roots in heterogenous panels", *Journal of Econometrics, 115, 53-74.*

Jarocinski, M., and Marcet, A. (2005). On the link between Bayesian estimators and classical bias corrections in time series, *working paper*.

Kapetanios, G. (2005), "Dynamic factor extraction of cross-sectional dependence in panel data unit root tests", Working Paper 509, Queen Mary, University of London.

Kass, R.E., and A.E. Raftery (1995). Bayes Factors, *Journal of the American Statistical Association*, **90**, 773-795.

Levin, A., C. Lin and C.J. Chu (2002), "Unit root tests in Panel data: asymptotic and finite-sample properties", *Journal of Econometrics, 108, 1-24.*

Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag.

Lubrano, M. (1995). Testing for unit roots in a Bayesian framework, *Journal of Econometrics*, **69**, 81-109.

Maddala, G.S. and S. Wu (1999), "A comparative study of unit root tests with panel data and a new simple test", Oxford Bulletin of Economics and Statistics, 61, 631-652.

Meligkotsidou, L., Tzavalis, E. and Vrontos, I. D. (2008). A Bayesian Analysis of Unit Roots and Structural Breaks in the Level and the Error Variance of Autoregressive Models of economic series. (*Submitted*).

Moon, H.R. and B. Perron (2005), "Testing for unit roots in panels with dynamic factors", *Journal of Econometrics, 122, 81-126.*

O'Connell, P. (1998), "The Overvaluation of Purchasing power parity", *Journal of International Economics, 44, 1-19.*

Phillips, P.C.B. (1991). To criticize the critics: An objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics*, **6**, 333-364.

Phillips, P.C.B. and H.R. Moon (1999), "Linear regression limit theory for nonstationary panel data", *Econometrica*, 67, 1057-1111.

Phillips, P.C.B. and D. Sul (2003), "Dynamic panel estimation and homogeneity testing under cross section dependence", *Econometrics Journal*, 6, 217-259.

Sims, C.A. (1988). Bayesian scepticism on unit root econometrics, *Journal of Economic dynamics and Control*, **12**, 463-474.

Uhlig, H. (1994). On Jeffreys prior when using the exact likelihood function, *Econometric Theory*, **10**, 633-644.

Zivot, E. (1994). A Bayesian analysis of the unit root hypothesis within an unobserved components model, *Econometric Theory*, **10**, 552-578.