

Mesters, Geert; Koopman, Siem Jan

**Working Paper**

## Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time

Tinbergen Institute Discussion Paper, No. 12-009/4

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Mesters, Geert; Koopman, Siem Jan (2012) : Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time, Tinbergen Institute Discussion Paper, No. 12-009/4, Tinbergen Institute, Amsterdam and Rotterdam, <http://nbn-resolving.de/urn:NBN:nl:ui:31-1871/38471>

This Version is available at:

<http://hdl.handle.net/10419/87227>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TI 2012-009/4

Tinbergen Institute Discussion Paper



# Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time

*Geert Mesters<sup>a,b</sup>*

*Siem Jan Koopman<sup>b,c</sup>*

<sup>a</sup> *Netherlands Institute for the Study of Crime and Law Enforcement;*

<sup>b</sup> *Department of Econometrics, VU University Amsterdam;*

<sup>c</sup> *Tinbergen Institute.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900  
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 525 8579

# Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time

*G. Mesters<sup>(a,b)\*</sup> and S.J. Koopman<sup>(b,c)</sup>*

<sup>(a)</sup> Netherlands Institute for the Study of Crime and Law Enforcement,

<sup>(b)</sup> Department of Econometrics, VU University Amsterdam,

<sup>(c)</sup> Tinbergen Institute, Amsterdam

February 6, 2012

## Abstract

An exact maximum likelihood method is developed for the estimation of parameters in a nonlinear non-Gaussian dynamic panel data model with unobserved random individual-specific and time-varying effects. We propose an estimation procedure based on the importance sampling technique. We evaluate the method in a Monte Carlo study for Student's  $t$  dynamic panel data models. We finally present an extensive empirical study into the interrelationships between the economic growth figures of countries listed in the Penn World Tables. It is shown that our dynamic panel data model can provide an insightful analysis of common and heterogeneous features in world-wide economic growth.

*JEL classification:* C33; C51

*Some keywords:* Panel data; non-Gaussian; Importance Sampling; Random Effects; Student's  $t$ ; Economic Growth.

---

\*Corresponding author: G. Mesters, Department of Econometrics, VU University Amsterdam, The Netherlands. Contact address: NSCR, PO box 71304, 1008 BH Amsterdam, The Netherlands, tel. nr. +31 20 59 83951, email: gmesters@nscr.nl, Technical Appendix: <http://personal.vu.nl/s.j.koopman>

# 1 Introduction

In this paper we develop a Monte Carlo maximum likelihood procedure for the estimation of parameters in a generalized dynamic panel data model. We assume that available data  $y$  stems from a possibly unbalanced panel of  $N$  individuals. For each individual  $i$ , we have  $T_i$  observations over time, with  $i = 1, \dots, N$ . The term “individual” can refer to countries, firms, groups, persons or other separately definable entities. Our generalized dynamic panel data model consists of a nonlinear non-Gaussian density for the observations conditional on a latent signal. We decompose the latent signal into a fixed component and a stochastic component. The fixed component is defined as a linear function of explanatory variables and lagged observations, whereas the stochastic component includes random individual-specific effects and time-varying effects. The two effects are assumed to come from mutually independent Gaussian densities. When the density of the observations is considered to be conditionally Gaussian with mean equal to the latent signal and some arbitrary variance, the model reduces to the linear Gaussian random effects panel data model as studied in Hsiao (2003) and Baltagi (2005).

Maximum likelihood estimation is complicated for the proposed model because the likelihood does not exist in closed form. The nonlinearity of the observation density together with the stochastic component of the latent signal prohibit closed form solutions. For the simultaneous analysis of random individual-specific and time-varying effects we extend the methods of Shephard & Pitt (1997) and Durbin & Koopman (1997). Their methods are based on Monte Carlo simulation methods. In particular, they adopt an importance sampler for which an approximating linear Gaussian state space model is used to draw samples. We need to extend their method because our model includes random individual-specific effects. We construct a sequence of conditional importance densities that sequentially integrates out random effects from the joint distribution. We disentangle the integration over the cross-section dimension (for the individual-specific effects) and the time series dimension (for the time-varying effects). The constructed importance densities are based on a linear

Gaussian dynamic panel data model which sufficiently approximates the true model. Our proposed methodology for obtaining the linearized Gaussian model takes into account the developments reported in So (2003) and Jungbacker & Koopman (2007).

We further show that the panel of time series can be collapsed into two low-dimensional vector series. Each vector series follows an approximating linear Gaussian panel data model. These low-dimensional vector series are used to sample random individual-specific and time-varying effects from the importance densities. In particular, the first transformation collapses the cross-sectional dimension of  $y$  without compromising the information needed to sample the time-varying effects. This transformation is introduced in Jungbacker & Koopman (2008). The second transformation collapses the time series dimension of the panel without compromising the information needed to sample the individual-specific effects. The transformations are easy to implement and lead to large computational savings when evaluating the Monte Carlo likelihood. We document the possible savings that can be achieved by our approach.

For non-Gaussian dynamic panel data models without time-varying effects, other Monte Carlo estimation methods are considered in the literature. Examples are simulated maximum likelihood approaches based on the Geweke-Hajivassiliou-Keane (GHK) sampler, see Geweke (1991), Hajivassiliou (1990) and Keane (1994) and the more general Markov Chain Monte Carlo methods including Gibbs sampling and the Metropolis-Hastings algorithm, see Geweke & Keane (2001). Richard & Zhang (2007) and Liesenfeld & Richard (2008) show that simulation based inference is possible using their efficient importance sampling (EIS) method for non-Gaussian dynamic panel data models with individual-specific and time-varying effects. Our method differs from the Liesenfeld & Richard (2008) approach in three ways. First, we disentangle the Monte Carlo integration over the individual-specific and time-varying effects by conditioning on the posterior modal values of the time-varying and individual-specific effects, respectively. Second, our importance samplers fall in the class of importance samplers proposed by Shephard & Pitt (1997) and Durbin & Koopman (1997), which are considerably faster compared to the EIS importance samplers, as shown in Koopman, Lucas & Scharth

(2011) for a large variety of models. Third, we sample random effects from our importance densities, after transforming the data panel  $y$  into two low-dimensional vector series. Increasing the panel dimensions while keeping the number of random effects constant, has almost no impact on the overall computational efficiency of our proposed estimation method.

The new estimation method for the general model provides several additional benefits. First, when only individual-specific effects are included in our model, our sampler remains highly accurate despite the length of the time series dimension. In this respect we improve on the GHK sampler based simulation method, whose performance is shown to deteriorate as the time series dimension becomes large, see Lee (1997). Second, our framework allows for the treatment of heterogeneous regression parameters. This is useful as in empirical panel data studies parameter homogeneity is often hard to establish a priori. Heterogeneity can be imposed with respect to the individuals as well as the time periods, by following the implementation described in Hsiao & Pesaran (2008). They discuss heterogeneous parameters in the context of linear regression models. Our estimation procedure requires only minor modifications to adopt their model specifications, while retaining our non-Gaussian framework. Third, the estimation method can be computationally modified to handle missing values and unbalanced panels. Additional methods are not necessary and it contrasts with the two-step procedures as developed by for example Stock & Watson (2002).

The remainder of the paper is organized as follows. The next section reviews some known examples of studies from different areas of research, for which the model specification can be cast into our generalized dynamic panel data model framework. Section 3 formally describes the generalized dynamic panel data model in detail. In Section 4 we develop our Monte Carlo maximum likelihood method for the general model. We provide steps for efficient implementation and explain our treatment of unbalanced panels. Section 5 evaluates the performance of our estimation method in a simulation study concerning Student's  $t$  dynamic panel data models, which are further investigated in Section 6, where we present an empirical study concerning economic growth rates for countries listed in the Penn World tables. The study extends existing panel data applications for economic growth by allowing

for Student's  $t$  densities, missing values, parameter heterogeneity and multiple time-varying effects. Section 7 summarizes our findings and presents some directions for future research.

## 2 Some Examples

Due to the general formulation of our model many different classes of models can be considered for the estimation methods of Section 4. Before we present the generalized dynamic panel data model in detail in Section 3 we discuss three illustrations from different fields of research. Each illustration implies a different non-Gaussian conditional density for variable  $y_{i,t}$ , which is associated with individual  $i$  and time period  $t$ . Each density is defined conditional on signal  $z_{i,t}$ , given by

$$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i + \xi_t, \quad (1)$$

where  $\gamma$  is a parameter measuring the effect of the previous outcome variable  $y_{i,t-1}$ ,  $\beta$  is a parameter vector measuring explanatory variables  $x_{i,t}$ ,  $\mu_i$  is a random individual-specific effect and  $\xi_t$  is a random time-varying effect. Equation (1) is discussed in Hsiao (2003) and Baltagi (2005) for the case where  $y_{i,t}$  follows a conditionally Gaussian density, with mean given by  $z_{i,t}$  and some arbitrary variance. When relaxing the Gaussian assumption many additional phenomena can be cast into this framework.

### 2.1 Credit risk model

An important topic in empirical finance is the dynamical modeling of default rates. The basic framework that underlies the works of Duffie, Saita & Wang (2007) and Koopman & Lucas (2008), is that the default risk of firms can be decomposed into a firm-specific risk component and a systematic, economy-wide risk component. Their framework can be described using observed variables  $y_{i,t}$  and a partially unobserved signal  $z_{i,t}$ , given by (1). Let  $y_{i,t}$  denote the number of firms that default from a group of firms  $i$  in time period  $t$ .



Koopman & Lucas (2008) construct groups of firms by matching current ratings, industries and age cohorts. The corresponding density, relating  $y_{i,t}$  to signal  $z_{i,t}$ , is then given by

$$y_{i,t} \sim \text{Binomial } \{k_{i,t}, [1 + \exp(-z_{i,t})]^{-1}\},$$

where  $k_{i,t}$  is the number of firms in group  $i$  at time  $t$  and  $[1 + \exp(-z_{i,t})]^{-1}$  is the logit transformed default probability. The logit transformation leaves  $z_{i,t}$  unrestricted and keeps the default probability between zero and one. For the interpretation of the components of signal  $z_{i,t}$  it holds that  $\gamma$  captures the structural causal effect of firms defaulting in the previous time period,  $x_{i,t}$  is the vector of explanatory variables including observed sector-specific and macro-economic variables, see Duffie et al. (2007),  $\mu_i$  is the group-specific risk component, capturing unobserved rating, industry and age cohort effects, and finally  $\xi_t$  is the unobserved common systematic risk component. Additional lags of  $y_{i,t}$  can also be included in this specification, which can be important depending on chosen time-grid. The unobserved systematic risk factor is often taken as the weighted average of several economy-wide factors.

## 2.2 Binary panel data model

In situations where person  $i$  has to make yes/no decisions for multiple time periods  $t$ , we obtain binary panel data sets. Examples are decisions concerning employment, crime, children, consumption and union participation. The modeling of binary panels has attracted much attention in the micro-econometric literature; see Heckman (1981*a, b*) for early contributions and Baltagi (2005, Chapter 11) for a textbook treatment. A special case concerns the dynamic analysis of labor supply of females; see for example Hyslop (1999) and Keane & Sauer (2009). For this particular example let  $y_{i,t} = 1$  denote the outcome that female  $i$  is working in time period  $t$  and  $y_{i,t} = 0$  that she is not working. A logistic model for variable  $y_{i,t}$  can be given by

$$y_{i,t} \sim \text{Binary } \{[1 + \exp(-z_{i,t})]^{-1}\},$$

where signal  $z_{i,t}$  is interpretable as the latent net-utility resulting from decision  $y_{i,t}$ . For the components of the signal  $z_{i,t}$  in (1) it holds that,  $y_{i,t-1}$  captures the effect of the decision in the previous time period,  $x_{i,t}$  is a vector of observed human capital, demographic and family structure variables,  $\mu_i$  is the person-specific effect representing time invariant unobserved human capital and taste factors, and  $\xi_t$  is the time-varying effect capturing demand side aspects of the labor market, such as the business cycle. This model is an example. Different simplifications and extensions of this model can be framed in our set-up.

## 2.3 Economic growth model

In the empirical economic growth literature much effort has been devoted towards testing variants of the neo-classical growth model of Solow (1956). Prominent examples include Mankiw, Romer & Weil (1992) and Islam (1995). The neo-classical growth model relates the gross domestic product of a country in an arbitrary time period to the savings rate, the population growth rate and the gross domestic product of the previous time period. In the majority of studies the corresponding parameters are estimated by using data from the Penn World Tables. However, a concern is that the data is subject to large outliers; see Temple (1999) and Durlauf, Johnson & Temple (2005). It implies that a selection of observations, that are distant from the rest data, may act as influential outliers or leverage points. For example, De Long & Summers (1991) find that, within their sample of countries, the observations of Botswana and Zambia have large effects on the coefficient estimates and their precision. To address this issue Juarez & Steel (2010) assume a Student's  $t$  density for the growth variable variable  $y_{i,t} = \log Y_{i,t} - \log Y_{i,t-1}$ , where  $Y_{i,t}$  is the per capita output of country  $i$  for period  $t$ . The growth rate is then modeled by

$$y_{i,t} \sim t(z_{i,t}, \sigma_\zeta, \nu),$$

where  $t(z_{i,t}, \sigma_\zeta, \nu)$  is the Student's  $t$  density with mean  $z_{i,t}$ , scaling  $\sigma_\zeta$  and degrees of freedom  $\nu$ . Juarez & Steel (2010) find strong evidence in favor of heavier tails for a sample of OECD

countries<sup>1</sup>. We investigate this economic growth illustration further and more formally in Section 6.

### 3 Generalized dynamic panel data model

We formally define the generalized dynamic panel data model for observations of variable  $y_{i,t}$ , that is associated with individual  $i$  and time  $t$ . Data is available for  $N$  individuals. For each individual  $i$ , the time series dimension is  $T_i$ , for  $i = 1, \dots, N$ . Each time period is indexed by  $t$ . The entire time span of the unbalanced panel is restricted between some arbitrary starting period  $t = 1$  and the final period  $t = T$ . The model for variable  $y_{i,t}$  is given by

$$y_{i,t} \stackrel{i.i.d.}{\sim} p(y_{i,t}|z_{i,t}; \psi), \quad (2)$$

where  $z_{i,t}$  is the signal for  $y_{i,t}$  and  $p(y_{i,t}|z_{i,t}; \psi)$  is a density that depends on the parameter vector  $\psi$ . We assume that  $p(y_{i,t}|z_{i,t}; \psi)$  is possibly non-Gaussian and is correctly specified. The latent signal  $z_{i,t}$  incorporates all dynamics, covariates and stochastic processes driving the density  $p(y_{i,t}|z_{i,t}; \psi)$ . A general decomposition of signal  $z_{i,t}$  is given by

$$z_{i,t} = w_{i,t} + \epsilon_{i,t}, \quad (3)$$

where  $w_{i,t}$  is a fixed component and  $\epsilon_{i,t}$  is a stochastic component. The fixed component  $w_{i,t}$  is specified by the linear function

$$w_{i,t} = x'_{i,t}\beta + \gamma(B)y_{i,t}, \quad (4)$$

where  $x_{i,t}$  is a  $k \times 1$  vector of observable explanatory variables,  $\beta$  is a  $k \times 1$  parameter vector and  $\gamma(B) = \gamma^1 B + \dots + \gamma^{p_y} B^{p_y}$  is the backshift polynomial, with unknown coefficients  $\gamma^j$  for  $j = 1, \dots, p_y$  and for some non-negative integer  $p_y$ . The backshift operator  $B$  is defined such that  $B^s y_{i,t} = y_{i,t-s}$ , for any integer  $s$ . The polynomial  $\gamma(B)$  incorporates past outcomes to

---

<sup>1</sup>OECD is the Organisation for Economic Co-operation and Development.

affect the current signal in a structural way; see the discussion in Baltagi (2005, Chapter 8). Initially we will assume that parameters  $\beta$  and  $\gamma$  are common for all individuals and time periods, and that explanatory variables  $x_{i,t}$  are exogenous and uncorrelated with  $\epsilon_{j,t}$ , for all  $i, j = i, \dots, N$  and common time periods  $t$ . Section 6 discusses options for relaxing these assumptions.

The stochastic component  $\epsilon_{i,t}$  is given by

$$\epsilon_{i,t} = a'_{i,t}\mu_i + b'_{i,t}\xi_t, \quad \mu_i \sim NID(\delta, \Sigma_\mu), \quad (5)$$

where  $\mu_i$  is a  $q \times 1$  vector of individual-specific effects, which is weighted for individual  $i$  in time period  $t$  by  $q \times 1$  vector  $a_{i,t}$  and  $\xi_t$  is a  $r \times 1$  vector of time-varying effects, which is weighted for individual  $i$  in time period  $t$  by  $r \times 1$  vector  $b_{i,t}$ . The individual effects  $\mu_i$  are assumed normally and independently distributed, with  $q \times 1$  common mean vector  $\delta$  and  $q \times q$  variance matrix  $\Sigma_\mu$ , which are both considered fixed. Both weight vectors,  $a_{i,t}$  and  $b_{i,t}$ , are considered fixed and may depend on the parameter vector  $\psi$ . Time-varying effects  $\xi_t$ , are assumed to be generated from a linear dynamic process given by

$$\xi_t = G\alpha_t, \quad \alpha_{t+1} = H\alpha_t + R\eta_t, \quad \eta_t \sim NID(0, \Sigma_\eta), \quad t = 1, \dots, T, \quad (6)$$

where  $r \times p$  dimensional matrix  $G$  relates the generating linear autoregressive process  $\alpha_t$  to the time-varying effects  $\xi_t$ ,  $H$  is a  $p \times p$  transition matrix,  $R$  is a  $p \times l$  disturbance selection matrix and  $\eta_t$  is a  $l \times 1$  vector of disturbances with variance matrix  $\Sigma_\eta$ . These system matrices are fixed and known, although some elements may depend on parameter vector  $\psi$ . The initial state vector  $\alpha_1$  is assumed normally distributed with mean zero and variance matrix  $P$ . The corresponding initial time-varying effect is normally distributed with mean zero and variance  $\Sigma_\xi = GPG'$ . Individual-specific effects  $\mu_i$  and  $\mu_j$  are assumed mutually uncorrelated and independent from the time-varying effects,  $\xi_t$ , for all  $i, j = 1, \dots, N$  and  $t = 1, \dots, T$ .

Many studies based on panel data models are dynamic in nature and the occurrence of a

particular outcome often appears to be related to past outcomes. The dynamic panel data model, given by equations (2), (3), (4), (5) and (6), allows us to distinguish between two sources capable of explaining these dynamics, see Heckman (1981*a, b*). The first source is the presence of “true state dependence”, which is the phenomenon that past outcomes provide explanatory power for future outcomes. This is represented in our model by term  $\gamma(B)y_{i,t}$ . The second source, referred to by Heckman (1981*a*) as “spurious state dependence”, explains dynamics as resulting from serial correlation in stochastic component  $\epsilon_{i,t}$ . We aim to capture serial correlation in  $\epsilon_{i,t}$  by including individual-specific effects  $\mu_i$  and time-varying effects  $\xi_t$ .

The general model contains many parameters. To identify these parameters in the model different strategies can be considered. In general we need to restrict either the distribution of  $\mu_i$ ,  $\xi_t$  or a combination of both. Further, as only a limited number of elements of weight vectors  $a_{i,t}$  and  $b_{i,t}$  can be identified, some hierarchical constraints must be imposed. Many different restrictions can be considered, of which the appropriateness needs to be evaluated on a case by case basis.

The initial signal of the first time period is given, for  $i = 1, \dots, N$ , by

$$z_{i,1} = x'_{i,1}\beta + \gamma(B)y_{i,1} + a'_{i,1}\mu_i + b'_{i,1}\xi_1, \quad \mu_i \sim NID(\delta, \Sigma_\mu), \quad \xi_1 \sim N(0, \Sigma_\xi), \quad (7)$$

where we assume  $y_{i,t}$ , for  $t < 1$ , to be fixed and known constants for all  $i = 1, \dots, N$ . For a more elaborate treatment of the initial conditions, the methods of Woolridge (2005) can be considered in our framework, but this is not further explored in this paper. The generalized dynamic panel data model of this paper is fully defined by equations (2), (3), (4), (5), (6) and (7). All parameters are collected in vector  $\psi$  and typically contain parameters affecting signal  $z_{i,t}$ . Under the assumption that the model is correctly specified, the density of the observations  $y = \{y_{i,t}\}$  conditional on signal  $z = \{z_{i,t}\}$  is given by

$$p(y|z; \psi) = \prod_{i=1}^N \prod_{t=1}^{T_i} p(y_{i,t}|z_{i,t}; \psi) = \prod_{i=1}^N p(y_{i,1}|z_{i,1}; \psi) \prod_{t=2}^{T_i} p(y_{i,t}|\mu_i, \xi_t; x_{i,t}, \mathcal{Y}_{i,t-1}, \psi), \quad (8)$$

where  $\mathcal{Y}_{i,t} = \{y_{i,1}, \dots, y_{i,t}\}$ . The last equality is partly the result of the prediction error decomposition.

## 4 Likelihood evaluation by Monte Carlo integration

This section discusses the method of Monte Carlo maximum likelihood for the estimation of the parameter vector  $\psi$ . We first consider the generalized dynamic panel data model for balanced panels,  $T_i = T$  for all  $i = 1, \dots, N$ . In Section 4.4 we provide the necessary alterations for the treatment of unbalanced panels. The loglikelihood for observation vector  $y$  is defined by  $\ell(\psi) = \log p(y; \psi)$ , where  $p(y; \psi)$  denotes the joint density of all observations for parameter vector  $\psi$ . In the remainder of this section we drop the dependence on parameter vector  $\psi$  for notational convenience and define  $\log p(y) \equiv \log p(y; \psi)$ .

In the presence of unobserved random individual-specific and time-varying effects,  $\mu = \{\mu_i\}$  and  $\xi = \{\xi_t\}$ , density  $p(y)$  can be expressed as

$$p(y) = \int_z p(y, z) \, dz = \int_\xi \int_\mu p(y, \mu, \xi; x) \, d\mu \, d\xi = \int_\xi \int_\mu p(y|\mu, \xi; x) p(\mu, \xi) \, d\mu \, d\xi, \quad (9)$$

where the second equality holds as  $x = \{x_{i,t}\}$  is non-stochastic and where  $p(\mu, \xi) = p(\mu)p(\xi)$ , since  $\mu$  and  $\xi$  are independent.

The evaluation of the high dimensional integral (9) is complicated because an analytical solution is not available for the nonlinear non-Gaussian density  $p(y|\mu, \xi; x) = p(y|z)$ . We propose to solve the integral by Monte Carlo integration. A frequency based estimator of this type is given by  $M^{-1} \sum_{i=1}^M p(y|\mu^{(i)}, \xi^{(i)}; x)$ , where draws  $\mu^{(i)}$  and  $\xi^{(i)}$ , for  $i = 1, \dots, M$ , are obtained from density  $p(\mu, \xi)$ . This estimator is based on sampler  $p(\mu, \xi)$  and is consistent but it requires a very large number of draws before convergence to  $p(y)$  is achieved. More efficiency is obtained when an adequate importance sampler can be used; see Ripley (1987). The implementation relies on constructing an importance density, sampling from it and adjusting the density of interest to correct for the use of the “incorrect” sampler. The

importance density is usually conditioned on  $y$  so that the simulations from this density are efficient as it accounts for the observations directly.

A general importance sampling representation for  $p(y)$  is given by

$$p(y) = \int_{\xi} \int_{\mu} \frac{p(y|\mu, \xi; x)p(\mu, \xi)}{g(\mu, \xi|y)} g(\mu, \xi|y) d\mu d\xi, \quad (10)$$

where  $g(\mu, \xi|y)$  denotes the importance density. Integral (10) can be solved by Monte Carlo integration for which we sample  $\mu^{(i)}$  and  $\xi^{(i)}$  from the importance density and compute estimate  $M^{-1} \sum_{i=1}^M p(y, \mu^{(i)}, \xi^{(i)}; x) / g(\mu^{(i)}, \xi^{(i)}|y)$ .

For any choice of the density  $g(\mu, \xi|y)$ , sampling from it is likely to be complicated as the covariance matrix of  $y$  has an inconvenient structure, as a result of correlation between all individuals (due to  $\xi_t$ ) and time periods (due to  $\mu_i$ ). To circumvent the problem, we propose the use of two importance samplers; one for the cross-section dimension (for the integral of  $\mu$ ) and one for the time series dimension (for the integral of  $\xi$ ). Since  $g(\mu, \xi|y) = g(\mu|y)g(\xi|y)$ , by the independence between  $\mu$  and  $\xi$ , we can sample  $\mu$  and  $\xi$  separately from  $g(\mu|y)$  and  $g(\xi|y)$ , respectively. We notice that densities  $g(\mu|y)$  and  $g(\xi|y)$  still depend on  $\xi$  and  $\mu$ , respectively, by means of  $y$ . We therefore integrate out  $\mu$ , by keeping  $\xi$  fixed at its posterior modal value and we integrate out  $\xi$  by keeping  $\mu$  fixed at its posterior modal value. The posterior model values are chosen for computational convenience. Other sufficient statistics can also be considered. Their performance needs to be evaluated on a case by case basis. For density  $p(y)$  from the generalized dynamic panel data model we propose the following importance sampling representation

$$p(y) = \int_{\xi} \int_{\mu} \frac{p(y|\mu, \xi; x)p(\mu)p(\xi)}{g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu})} g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) d\mu d\xi, \quad (11)$$

where  $g(\mu|y; \hat{\xi})$  and  $g(\xi|y; \hat{\mu})$  are the importance densities. We define  $\hat{\mu}$  and  $\hat{\xi}$  as the posterior

modal values of  $p(\mu, \xi|y; x)$ , that is

$$\{\hat{\mu}, \hat{\xi}\} \equiv \operatorname{argmax}_{\mu, \xi} p(\mu, \xi|y; x). \quad (12)$$

The posterior modal values  $\hat{\mu}$  and  $\hat{\xi}$  can be found iteratively, which we discuss Section 4.1. When applying Bayes rule twice to the right hand side of equation (11) we obtain

$$p(y) = g(y; \hat{\xi})g(y; \hat{\mu}) \int_{\xi} \int_{\mu} \frac{p(y|\mu, \xi; x)}{g(y|\mu; \hat{\xi})g(y|\xi; \hat{\mu})} g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi, \quad (13)$$

where we have retained the marginal properties of  $\mu$  and  $\xi$  by imposing  $g(\xi) = p(\xi)$  and  $g(\mu) = p(\mu)$ . Densities  $g(y; \hat{\xi}) = g(\mu, y; \hat{\xi})/g(\mu|y; \hat{\xi})$  and  $g(y; \hat{\mu}) = g(\xi, y; \hat{\mu})/g(\xi|y; \hat{\mu})$  can be interpreted as the joint densities of the observations conditional on the posterior modal values  $\hat{\mu}$  and  $\hat{\xi}$ , respectively. We notice that  $g(\mu; \hat{\xi}) = p(\mu)$  and  $g(\xi; \hat{\mu}) = p(\xi)$ , as  $\mu$  and  $\xi$  are independent. Under the assumption that the modes  $\hat{\mu}$  and  $\hat{\xi}$  are well defined and can be computed, we define  $\hat{p}(y)$  as the Monte Carlo estimate of (13) and given by

$$\hat{p}(y) = g(y; \hat{\xi})g(y; \hat{\mu}) \sum_{i=1}^M \frac{p(y|\mu^{(i)}, \xi^{(i)}; x)}{g(y|\mu^{(i)}; \hat{\xi})g(y|\xi^{(i)}; \hat{\mu})}, \quad (14)$$

where samples  $\{\mu^{(1)}, \dots, \mu^{(M)}\}$  are drawn independently from importance density  $g(\mu|y; \hat{\xi})$  and samples  $\{\xi^{(1)}, \dots, \xi^{(M)}\}$  from  $g(\xi|y; \hat{\mu})$ . Density  $p(y|\mu^{(i)}, \xi^{(i)}; x)$  is evaluated using equation (8).

The quality of the estimate in equation (14) depends on how well the product of  $g(\mu|y; \hat{\xi})$  and  $g(\xi|y; \hat{\mu})$  approximates  $p(y, \mu, \xi; x)$ , which needs to be evaluated on a case by case basis. In practice we take both importance densities from the Gaussian distribution and adjust their mean and variance to ensure that the product is close in proportionality to  $p(y, \mu, \xi; x)$ . For any importance density there holds that  $\hat{p}(y) \rightarrow p(y)$  as  $M \rightarrow \infty$ , which is implied by Kolmogorov's strong law of large numbers. However, the efficiency of the importance density relies on the rate of convergence in terms of  $M$ . The Lindeberg-Levy central limit theorem implies a  $\sqrt{M}$  convergence rate if draws from the importance sampler are independent and



if importance weights

$$w^{(i)} = p(y|\mu^{(i)}, \xi^{(i)}; x) / \left[ g(y|\xi^{(i)}; \hat{\mu}) g(y|\mu^{(i)}; \hat{\xi}) \right], \quad (15)$$

have finite mean and variance, as argued in Geweke (1989). The last condition can be examined empirically using extreme value theory based tests proposed in Monahan (2001) and Koopman, Shephard & Creal (2009). In the simulation study of Section 5 we consider diagnostic test statistics for the existence of a variance in a sequence of importance weights drawn from Student's  $t$  dynamic panel data models.

## 4.1 Constructing the importance density

Next we consider the construction of importance densities  $g(\mu|y; \hat{\xi})$  and  $g(\xi|y; \hat{\mu})$ , proposed for evaluating estimate  $\hat{p}(y)$ , given in equation (14). We choose both densities to follow Gaussian distributions and modify their means and variances such that their modes are equal to the modes of the original posterior density  $p(\mu, \xi|y; x)$ . Similar strategies are followed for models without random individual-specific effects; see for example, Shephard & Pitt (1997) and Durbin & Koopman (1997, 2000). So (2003) and Jungbacker & Koopman (2007) argue that this strategy can be implemented by numerically maximizing  $\log p(\mu, \xi|y; x) = \log p(y|\mu, \xi; x) + \log p(\mu, \xi) - \log p(y; x)$  with respect to  $\mu$  and  $\xi$ .

The instrumental basis to facilitate this numerical maximization is given, for variable  $y_{i,t}$ , by the linear Gaussian panel data model

$$y_{i,t} = c_{i,t} + \epsilon_{i,t} + u_{i,t}, \quad u_{i,t} \sim NID(0, d_{i,t}^2), \quad (16)$$

where  $c_{i,t}$  is a fixed constant, stochastic component  $\epsilon_{i,t}$  is given by equation (5) and  $u_{i,t}$  is a random variable with mean zero and fixed variance  $d_{i,t}^2$ . The stochastic component  $\epsilon_{i,t}$  is the same as in the original model of interest. The predetermined component  $w_{i,t}$  is not explicitly included in approximating model (16) since it is fixed at time  $t$ . The constants

$c_{i,t}$  and  $d_{i,t}$  are chosen such that (16) can be used to compute the posterior modal values  $\hat{\mu}$  and  $\hat{\xi}$ , respectively. The elements  $u_{i,t}$  and  $\epsilon_{j,s}$  are uncorrelated with each other, for all  $i, j = 1, \dots, N$  and  $s, t = 1, \dots, T$ . Furthermore,  $u_{i,t}$  is serially uncorrelated. It follows that

$$g(y|\mu, \xi) = \prod_{i=1}^N \prod_{t=1}^T g(y_{i,t}|\mu_i, \xi_t), \quad \text{with} \quad g(y_{i,t}|\mu_i, \xi_t) \equiv NID(c_{i,t} + \epsilon_{i,t}, d_{i,t}^2). \quad (17)$$

The maximization of  $\log p(\mu, \xi|y; x)$  with respect to  $\mu$  and  $\xi$  can be carried out via the Newton-Raphson method. The idea is to iterate between linearizing  $p(y|\mu, \xi; x)$ , by computing  $c = \{c_{i,t}\}$  and  $d = \{d_{i,t}\}$ , to obtain  $g(y|\mu, \xi)$  and updating  $\mu$  and  $\xi$  based on the linearized model given by equations (16) and (5). The following algorithm summarizes this method.

### Algorithm A

(i) Initialize the algorithm by choosing  $\mu^*$  and  $\xi^*$  as starting values, which gives  $\epsilon_{i,t}^*$  and  $z_{i,t}^*$ , for all  $i = 1, \dots, N$  and  $t = 1, \dots, T$ ;

(ii) Given the set of two equations

$$\frac{\partial \log p(y_{i,t}|z_{i,t})}{\partial z_{i,t}} = \frac{\partial \log g(y_{i,t}|\epsilon_{i,t})}{\partial \epsilon_{i,t}}, \quad \frac{\partial^2 \log p(y_{i,t}|z_{i,t})}{\partial z_{i,t} \partial z_{i,t}} = \frac{\partial^2 \log g(y_{i,t}|\epsilon_{i,t})}{\partial \epsilon_{i,t} \partial \epsilon_{i,t}},$$

for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ , where  $p(y_{i,t}|z_{i,t})$  is the observation model (2) and  $g(y_{i,t}|\epsilon_{i,t})$  is given by (17), we can deduct expressions for  $c_{i,t}$  and  $d_{i,t}$  as functions of  $z_{i,t}$ , and compute  $c_{i,t} = c_{i,t}^*$  and  $d_{i,t} = d_{i,t}^*$  for  $\epsilon_{i,t} = \epsilon_{i,t}^*$  and  $z_{i,t} = z_{i,t}^*$ ;

(iii) Compute  $\tilde{\mu} = E_g(\mu|y; \xi^*)$  from the resulting model (16) with  $\xi = \xi^*$ ,  $c_{i,t} = c_{i,t}^*$  and  $d_{i,t} = d_{i,t}^*$ ;

(iv) Replace  $\mu^*$  by  $\mu^* = \tilde{\mu}$ ;

(v) Compute  $\tilde{\xi} = E_g(\xi|y; \mu^*)$  from the resulting model (16) with  $\mu = \mu^*$ ,  $c_{i,t} = c_{i,t}^*$  and  $d_{i,t} = d_{i,t}^*$ ;

(vi) Replace  $\xi^*$  by  $\xi^* = \tilde{\xi}$

(vii) Iterate from (ii) to (vi) until convergence.

Since the mode and the mean of the approximating linear Gaussian model are set equal to the mode of the original model, it holds that  $\tilde{\mu} = \hat{\mu} = \operatorname{argmax}_{\mu} p(\mu|y; \hat{\xi}; x)$  and  $\tilde{\xi} = \hat{\xi} = \operatorname{argmax}_{\xi} p(\xi|y; \hat{\mu}; x)$ . Further, as  $\mu$  and  $\xi$  are independent, it holds that  $\{\hat{\mu}, \hat{\xi}\} = \operatorname{argmax}_{\mu, \xi} p(\mu, \xi|y; x)$ .

The performance of Algorithm A depends crucially on the efficient computation of the conditional expectations in steps (iii) and (v). With respect to step (iii), for a given value of  $\xi^*$ , the approximating model (16) is reduced to a standard random effects model, with weighted individual-specific effects and heteroskedastic error term  $u_{i,t}$ , see Baltagi (2005,

Chapters 2 and 5). This becomes clear when concatenating variables  $y_{i,t}$ , based on approximating model (16), over the time series dimension. This gives

$$\bar{y}_i = \bar{c}_i + \bar{A}_i\mu_i + \bar{\mathcal{B}}_i + \bar{u}_i, \quad \bar{u}_i \sim NID(0, \bar{D}_i), \quad i = 1, \dots, N, \quad (18)$$

where  $\bar{y}_i = (y_{i,1}, \dots, y_{i,T})'$ ,  $\bar{c}_i = (c_{i,1}, \dots, c_{i,T})'$ ,  $\bar{A}_i = (a_{i,1}, \dots, a_{i,T})'$ ,  $\bar{\mathcal{B}}_i = (b'_{i,1}\xi_1, \dots, b'_{i,T}\xi_T)'$  and  $\bar{u}_i = (u_{i,1}, \dots, u_{i,T})'$ . The  $T \times T$  variance matrix  $\bar{D}_i$  is diagonal by construction, with elements  $d_{i,1}^2, \dots, d_{i,T}^2$  on the main diagonal. Based on (18), the computation of  $E_g(\mu|y; \xi^*)$  can be performed using standard multivariate normal regression theory. The details are discussed in Appendix A.

Now consider step (v) where we need to compute  $E_g(\xi|y; \mu^*)$ . Given a value of  $\mu^*$ , approximating model (16), can be written as a linear Gaussian state space model. This can be seen by concatenating variables  $y_{i,t}$  over the cross-section dimension, which gives

$$y_t = c_t + \mathcal{A}_t + B_t\xi_t + u_t, \quad u_t \sim NID(0, D_t), \quad t = 1, \dots, T, \quad (19)$$

where  $y_t = (y_{1,t}, \dots, y_{N,t})'$ ,  $c_t = (c_{1,t}, \dots, c_{N,t})'$ ,  $\mathcal{A}_t = (a'_{1,t}\mu_1, \dots, a'_{N,t}\mu_N)'$ ,  $B_t = (b_{1,t}, \dots, b_{N,t})'$  and  $u_t = (u_{1,t}, \dots, u_{N,t})'$ . Variance matrix  $D_t$  is diagonal by construction, with elements  $d_{1,t}^2, \dots, d_{N,t}^2$  on the main diagonal. Based on (19) the computation of  $E_g(\xi|y; \mu^*)$  is carried out using the Kalman filter and smoothing methods. The details are provided in Appendix B.

## 4.2 Collapsing the approximating linear Gaussian panel data model

The evaluation of likelihood estimate  $\hat{p}(y)$  in (14), requires  $M$  samples of  $\mu$  and  $\xi$  from importance densities  $g(\mu|y; \hat{\xi})$  and  $g(\xi|y; \hat{\mu})$ , respectively. The posterior modal values  $\hat{\mu}$  and  $\hat{\xi}$  are obtained from Algorithm A. Both importance densities are based on approximating model (16). The vector representations (18) and (19), are adopted for computing the  $M$  samples by using the simulation smoother methods of Durbin & Koopman (2002). However,

both representations have large dimensions leading to simulation smoother methods that are computationally demanding. Instead, we show that more efficiency can be obtained by first performing two transformations to reduce the cross-section and time series dimensions of observed data  $y$ . In particular, the vectors series  $\bar{y}_i$  and  $y_t$  in equations (18) and (19), can be transformed into two low-dimensional vector series  $\bar{y}_i^l$  and  $y_t^l$ , for  $t = 1, \dots, T$  and  $i = 1, \dots, N$ . Based on these vector series, samples  $\xi^{(i)}$  and  $\mu^{(i)}$  can be drawn from  $g(\xi|y^l; \hat{\mu})$  and  $g(\mu|\bar{y}^l; \hat{\xi})$ , respectively, where  $\bar{y}^l = [(\bar{y}_1^l)', \dots, (\bar{y}_N^l)']'$  and  $y^l = [(y_1^l)', \dots, (y_T^l)']'$ . The resulting samples can be regarded as coming from  $g(\mu|y; \hat{\xi})$  and  $g(\xi|y; \hat{\mu})$ , respectively. In Section 5 we present the computational gains in evaluating the likelihood, for both sets of importance densities. The computational improvements resulting from the transformations are high.

#### 4.2.1 Collapsing the cross-section dimension

For the simulation of time-varying effects  $\xi^{(i)}$  from  $g(\xi|y^l; \hat{\mu})$ , we collapse  $N \times 1$  vectors  $y_t$ , based on equation (19), into low-dimensional vectors  $y_t^l$ , without losing information relevant for the extraction of  $\xi$ . This transformation has been introduced in Jungbacker & Koopman (2008) for the efficient evaluation of the likelihood for linear Gaussian dynamic factor models. Here only mild modifications of their methods are required.

Consider a linear approximating model for transformed data  $y_t^* = S_t(y_t - c_t - \hat{\mathcal{A}}_t)$ , where  $\hat{\mathcal{A}}_t = (a'_{1,t}\hat{\mu}_1, \dots, a'_{N,t}\hat{\mu}_N)'$  and  $y_t$  follows equation (19), where  $\mathcal{A}_t$  is replaced by  $\hat{\mathcal{A}}_t$ . The  $N \times N$  projection matrix  $S_t$  is assumed invertible, for all  $t = 1, \dots, T$ . The transformed observations are given by

$$y_t^* = \begin{bmatrix} y_t^l \\ y_t^h \end{bmatrix}, \quad \text{with} \quad \begin{aligned} y_t^l &= S_t^l(y_t - c_t - \hat{\mathcal{A}}_t) \\ y_t^h &= S_t^h(y_t - c_t - \hat{\mathcal{A}}_t) \end{aligned}, \quad t = 1, \dots, T, \quad (20)$$

where the projection matrices are partitioned as  $S_t = [(S_t^l)' ; (S_t^h)']'$ . For our particular model we choose matrices  $S_t^l$  and  $S_t^h$  to have dimensions  $r \times N$  and  $(N - r) \times N$ , respectively.

As a result the observation vectors  $y_t^l$  and  $y_t^h$  become of dimensions  $r \times 1$  and  $(N - r) \times 1$ . We aim to choose  $S_t^l$  and  $S_t^h$  such that  $y_t^l$  and  $y_t^h$  are uncorrelated and only  $y_t^l$  depends on  $\xi_t$ . In particular, we aim for a model of the form

$$\begin{aligned} y_t^l &= S_t^l B_t \xi_t + u_t^l, \\ y_t^h &= u_t^h, \end{aligned} \quad \begin{pmatrix} u_t^l \\ u_t^h \end{pmatrix} \sim N \left( 0, \begin{bmatrix} D_t^l & 0 \\ 0 & D_t^h \end{bmatrix} \right), \quad (21)$$

where  $D_t^l = S_t^l D_t S_t^{l'}$  and  $D_t^h = S_t^h D_t S_t^{h'}$  are  $r \times r$  and  $(N - r) \times (N - r)$  variance matrices respectively.

Suitable matrices  $S_t$ , which lead to model (21) need to satisfy the following conditions; (a) matrices  $S_t$  needs to be of full rank to prevent the loss of information, (b)  $S_t^h D_t S_t^{h'} = 0$  to ensure that observations  $y_t^l$  and  $y_t^h$  are independent, and (c)  $S_t^h B_t \xi_t = 0$  to ensure that  $y_t^h$  does not depend on  $\xi_t$ . Several of sequences of matrices  $S_t$  that fulfill these conditions can be found. A convenient choice is given by

$$S_t^l = \Delta_t' B_t' D_t^{-1}, \quad \Delta_t \Delta_t' = (B_t' D_t^{-1} B_t)^{-1}, \quad (22)$$

with  $\Delta_t$  being a lower triangular matrix. This choice for  $S_t^l$  results in

$$y_t^l = \Delta_t^{-1} \xi_t + u_t^l, \quad u_t^l \sim NID(0, I_r), \quad t = 1, \dots, T, \quad (23)$$

where  $\Delta_t^{-1}$  is a  $r \times r$  lower triangular matrix,  $\xi_t$  is defined in (6) and  $u_t^l$  is a random vector with mean zero and variance equal to the  $r$ -dimensional unit matrix  $I_r$ . Sampling time-varying effects  $\xi^{(i)}$  from  $g(\xi|y^l; \hat{\mu})$  is now performed by the applying the simulation smoother methods of Durbin & Koopman (2002) to  $r$ -dimensional vector series  $y_t^l$  and model (23), for  $t = 1, \dots, T$ . The matrices  $S_t^h$  remain of large dimensions and can be constructed from  $S_t^l$  but they are not required for any of the necessary computations. Proof of this transformation is presented in Jungbacker & Koopman (2008)

### 4.2.2 Collapsing the time series dimension

For the simulation of individual-specific effects  $\mu^{(i)}$  from  $g(\mu|\bar{y}^l; \hat{\xi})$  we collapse  $T \times 1$  vectors  $\bar{y}_i$ , for  $i = 1, \dots, N$ , based on vector representation (18), with  $\bar{\mathcal{B}}_i$  replaced by  $\hat{\mathcal{B}}_i = (b'_{i,1}\hat{\xi}_1, \dots, b'_{i,T}\hat{\xi}_T)'$ . We consider similar least squares type transformations as for the cross-section dimension above. However, because  $\mu_i$  and  $\mu_j$  are independent, the transformed observations  $\bar{y}_i^*$  become simple rescaled averages of the variables in  $\bar{y}_i$ . Let

$$\bar{y}_i^* = \begin{bmatrix} \bar{y}_i^l \\ \bar{y}_i^h \end{bmatrix}, \quad \text{with} \quad \begin{aligned} \bar{y}_i^l &= \bar{S}_i^l(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i) \\ \bar{y}_i^h &= \bar{S}_i^h(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i) \end{aligned}, \quad i = 1, \dots, N. \quad (24)$$

The motivation of the transformation is the same as above. We require to sample  $\mu_i$  based on only  $\bar{y}_i^l$  without compromising data information. We choose matrices  $\bar{S}_i^l$  and  $\bar{S}_i^h$  to have dimensions  $q \times T$  and  $(T - q) \times T$ , respectively. The model we aim to construct is given by

$$\begin{aligned} \bar{y}_i^l &= \bar{S}_i^l \bar{A}_i \mu_i + \bar{u}_i^l, \\ \bar{y}_i^h &= \bar{u}_i^h, \end{aligned} \quad \begin{pmatrix} \bar{u}_i^l \\ \bar{u}_i^h \end{pmatrix} \sim N \left( 0, \begin{bmatrix} \bar{D}_i^l & 0 \\ 0 & \bar{D}_i^h \end{bmatrix} \right), \quad (25)$$

where  $\bar{D}_i^l = \bar{S}_i^l \bar{D}_i \bar{S}_i^{l'}$  and  $\bar{D}_i^h = \bar{S}_i^h \bar{D}_i \bar{S}_i^{h'}$  are  $q \times q$  and  $(N - q) \times (N - q)$  variance matrices respectively. A convenient choice for  $\bar{S}_i^l$ , which satisfies the conditions stated above, is given by

$$\bar{S}_i^l = \bar{\Delta}_i' \bar{A}_i' \bar{D}_i^{-1}, \quad \bar{\Delta}_i \bar{\Delta}_i' = (\bar{A}_i' \bar{D}_i^{-1} \bar{A}_i)^{-1}, \quad (26)$$

with  $\bar{\Delta}_i$  being a lower triangular matrix. The resulting model for  $\bar{y}_i^l$  is given by

$$\bar{y}_i^l = \bar{\Delta}_i^{-1} \mu_i + \bar{u}_i^l, \quad \bar{u}_i^l \sim NID(0, I_q), \quad i = 1, \dots, N, \quad (27)$$

where  $\bar{\Delta}_i^{-1}$  is a lower triangular  $q \times q$  matrix,  $\mu_i$  is given in (5) and  $\bar{u}_i^l$  is a random vector with mean zero and  $q \times q$  unit variance. Again we can construct large matrices  $\bar{S}_i^h$ , but they are not required for any necessary computations. Samples  $\mu^{(i)}$  can be drawn independently from  $g(\mu_i|\bar{y}_i^l; \hat{\xi})$ , which is a Gaussian density with mean  $\Sigma_\mu \bar{\Delta}_i^{-1} (\bar{\Delta}_i^{-1'} \Sigma_\mu \bar{\Delta}_i^{-1} + I_q)^{-1} \bar{y}_i^l$  and

variance  $\Sigma_\mu - \Sigma_\mu \bar{\Delta}_i^{-1} (\bar{\Delta}_i^{-1'} \Sigma_\mu \bar{\Delta}_i^{-1} + I_q)^{-1} \bar{\Delta}_i^{-1'} \Sigma_\mu$ . Both expressions follow from the standard lemma discussed in Appendix A.

### 4.3 Constructing the Monte Carlo likelihood

Next we discuss the construction of the Monte Carlo likelihood estimate  $\hat{p}(y)$  in equation (14). The estimate relies on densities  $g(y; \hat{\mu})$  and  $g(y; \hat{\xi})$ , that are based on the approximating model (16). Density  $\log g(y; \hat{\mu})$  can be computed from the prediction error decomposition of vector representation (19), with  $\mu$  replaced by  $\hat{\mu}$ . This is obtained by a single pass through the Kalman filter, see Durbin & Koopman (2001, Chapter 7). Computational efficiency can be increased by using the lower dimensional model (23), based on vector series  $y_t^l$ . In particular, Jungbacker & Koopman (2008) show that

$$\log g(y; \hat{\mu}) = \text{constant} + \log g(y^l; \hat{\mu}) - \frac{1}{2} \sum_{t=1}^T \log |D_t| + e_t' D_t^{-1} e_t, \quad (28)$$

where  $y^l = (y_1^l, \dots, y_T^l)'$  and  $e_t = y_t - c_t - \hat{\mathcal{A}}_t - B_t(B_t' D_t^{-1} B_t)^{-1} B_t' D_t^{-1} (y_t - c_t - \hat{\mathcal{A}}_t)$  is the generalized least squares residual vector. Density  $g(y^l; \hat{\mu})$  can be computed from the prediction error decomposition of model (23), which is a  $r \times T$ -dimensional problem.

Due to the independence of the  $\mu_i$ 's logdensity  $\log g(y; \hat{\xi})$  is given by

$$\log g(y; \hat{\xi}) = \text{constant} - \frac{1}{2} \sum_{i=1}^N \log |\text{Var}_g(\bar{y}_i; \hat{\xi})| + \left[ (\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i)' \text{Var}_g(\bar{y}_i; \hat{\xi})^{-1} (\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i) \right],$$

where determinant  $|\text{Var}_g(\bar{y}_i; \hat{\xi})| = |\bar{A}_i \Sigma_\mu \bar{A}_i' + \bar{D}_i|$  can be hard to evaluate, depending on the structure of  $\bar{A}_i$ . More efficiency can be obtained by using the collapsed vector series  $\bar{y}_i^l$ , for  $i = 1, \dots, N$ . Based on model (27) we obtain

$$\log g(y; \hat{\xi}) = \text{constant} + \log g(\bar{y}^l; \hat{\xi}) - \frac{1}{2} \sum_{i=1}^N \log |\bar{D}_i| + \bar{e}_i' \bar{D}_i^{-1} \bar{e}_i, \quad (29)$$

where  $\bar{e}_i = \bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i - \bar{A}_i(\bar{A}_i' \bar{D}_i^{-1} \bar{A}_i)^{-1} \bar{A}_i' \bar{D}_i^{-1} (\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i)$ . Logdensity  $\log g(y; \hat{\xi})$  can



therefore be based on the  $N \times q$ -dimensional model (27).

The following algorithm summarizes the evaluation of the loglikelihood for balanced panels. Given parameter vector  $\psi$  we can evaluate the Monte Carlo loglikelihood estimate  $\log \hat{p}(y)$  in the following steps:

### Algorithm B

- (i) Run Algorithm A, where the posterior modal values  $\hat{\mu}$  and  $\hat{\xi}$  are calculated;
- (ii) Collapse panel  $y$  into low-dimensional vector series  $\bar{y}_i^l$  and  $y_i^l$  using Section 4.2;
- (iii) Sample  $M$  draws  $\mu^{(i)}$  and  $\xi^{(i)}$  from densities  $g(\xi|y^l; \hat{\mu})$  and  $g(\mu|\bar{y}^l; \hat{\xi})$ , which are based on transformed models (23) and (27), and compute importance weights  $w^{(i)}$ , as given in equation (15);
- (iv) Evaluate logdensities  $\log g(y; \hat{\mu})$  and  $\log g(y; \hat{\xi})$ , by computing (28) and (29), respectively;
- (v) Compute  $\log \hat{p}(y) = \log g(y; \hat{\mu}) + \log g(y; \hat{\xi}) + \log M^{-1} \sum_{i=1}^M w^{(i)}$ .

Loglikelihood estimate  $\log \hat{p}(y)$  can be optimized with respect to parameter vector  $\psi$  using an arbitrary numerical optimization method. As a practical choice we use the BFGS algorithm, see Nocedal & Wright (1999). To retain the smoothness of the likelihood in  $\psi$  we use the same random seed and the same value of  $M$  for each loglikelihood evaluation. The resulting Monte Carlo parameter estimates are denoted by  $\tilde{\psi}$ . In Section 5 we show the computational efficiency and accuracy of our methods, by providing average estimation times and summary statistics from repeated parameter estimates, for simulated data from Student's  $t$  dynamic panel data models.

## 4.4 Unbalanced or incomplete panels

In this section we provide the details for the treatment of unbalanced panels. We assume that for each individual we observe  $y_{i,t}$  and  $x_{i,t}$  for  $T_i$  consecutive time periods during a fixed

time interval. When  $y_{i,t}$  is unobserved step (ii) of Algorithm A is adjusted by removing  $x_{i,t}$  and unobservable lags  $\mathcal{Y}_{i,t-1}$  from  $z_{i,t}$ . The resulting  $z_{i,t}$  only depends on  $\mu_i$ ,  $\xi_t$  and possibly observed elements of  $\mathcal{Y}_{t-1}$ . Calculations in step (iii), conditional on  $\xi^*$ , are based on the standard random effects model, for which missing values can be handled by adopting the methods discussed by Baltagi (2005, Chapter 9). This amounts to calculating the components in Appendix A as usual with  $T$  replaced by  $T_i$ . Step (v) of Algorithm A is calculated by Kalman filter and smoothing methods, which can account for missing values, see Durbin & Koopman (2001, Section 4.8).

The transformations for panel reduction from Section 4.2 need to be adjusted for missing values as well. Jungbacker, Koopman & van der Wel (2011), show that by choosing an alternative state space representation for model (19) collapsed vectors  $y_t^l$  can be computed using similar transformations. The second transformation for the construction of  $\bar{y}_i^l$  can be computed based on the observed elements of  $\bar{y}_i$  only, as  $\mu_i$  and  $\mu_j$  are independent for all  $i, j = 1, \dots, N$ .

Likelihood estimate  $\hat{p}(y)$  is based on densities  $g(y; \hat{\mu})$  and  $g(y; \hat{\xi})$  and weights  $w^{(i)}$ . Density  $g(y; \hat{\mu})$  is based on the prediction error decomposition of lower dimensional model (23) and can be computed from the Kalman filter output. Generalized least squares residuals  $e_t$  only need to be computed for observed elements of  $y_t$ . Density  $g(y; \hat{\xi})$ , equation (29) can be computed based on lower dimensional model (27) and by adjusting the generalized least squares residual vectors  $\bar{e}_i$  to contain only the observed elements of  $\bar{y}_i$ . The weights  $w^{(i)}$  in (15) are based on elements of  $p(y|\mu^{(i)}, \xi^{(i)}; x)$  for which  $y_{i,t}$  and  $x_{i,t}$  are actually observed.

## 5 Simulation study for the Student's $t$ dynamic panel data model

We proceed by presenting a simulation study for a Student's  $t$  dynamic panel data model, which is a special case of the generalized dynamic panel data model discussed in Section 3. The simulation study is designed to evaluate the small sample properties of the estimation

procedure presented in Section 4. The focus is on determining whether a  $\sqrt{M}$  convergence rate is guaranteed for our importance sampling estimate  $\hat{p}(y)$  and whether the method is computationally feasible and accurate. The Student's  $t$  model is chosen as many data panels of interest contain certain elements  $y_{i,t}$ , that are numerically distant from the rest of the data. They can be dealt with by allowing for heavier-than-normal tail behavior, which we address by assuming a Student's  $t$  distribution with  $\nu$  degrees of freedom for density  $p(y_{i,t}|z_{i,t};\psi)$ . In particular, the data generating process for variable  $y_{i,t}$  is given by

$$y_{i,t} = z_{i,t} + \zeta_{i,t}, \quad \zeta_{i,t} \sim t(0, \sigma_\zeta, \nu), \quad i = 1, \dots, N, \quad t = 1, \dots, T_i, \quad (30)$$

where  $z_{i,t}$  is given by equation (3) and  $t(0, \sigma_\zeta, \nu)$  denotes the Student's  $t$  density with mean 0, scale  $\sigma_\zeta$  and degrees of freedom  $\nu$ .

Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures for Student's  $t$  dynamic panel data models without time-varying effects are considered by Fruhwirth-Schnatter & Kaufmann (2008), who represent the Student's  $t$  distribution as a scaled mixture of normal distributions. Juarez & Steel (2010) consider similar models, as well as skewed versions of the Student's  $t$  distribution, but also adopt MCMC.

## 5.1 Simulation Design

The main interest is in assessing the performance of the Monte Carlo estimation procedure for different signals, parameter values, panel sizes and numbers of missing values. Table 1 presents the combinations of signals and parameter values, that we investigate in our study. The signals correspond to models with; 1 individual-specific effects, 2 time-varying effects or 3 both.

For each signal we include a single covariate  $x_{i,t}$ , drawn independently from the  $N(0, 1)$  distribution and polynomial  $\gamma(B)$ , with  $p_y = 1$  and  $\gamma_1 = \gamma = 0.2$ . We choose  $\beta = 1$ . We include a univariate,  $q = 1$ , individual-specific effect  $\mu_i$  and a univariate,  $r = 1$ , time-varying effect  $\xi_t$ . Weights  $a_{i,t}$  and  $b_{i,t}$  are normalized to one. The individual-specific effect is normally

distributed with common mean  $\delta$  fixed at zero and variance  $\Sigma_\mu = \sigma_\mu^2$ . We investigate different values of standard deviation  $\Sigma_\mu = \sigma_\mu = 0.5, 1, 3$ . The time-varying effect is updated by an autoregressive process  $\alpha_t$  of order 1, where  $G = 1$ ,  $H = h$ ,  $R = 1$  and  $\Sigma_\eta = \sigma_\eta^2$ . Different degrees of persistence are investigated by taking  $h = 0.3$  or  $h = 0.9$ . The scaling of the time-varying effects is chosen as  $\sigma_\eta = 0.2$ . The initial value of the autoregressive process is given by  $N[0, \sigma_\eta^2/(1 - h^2)]$ . Parameters  $\sigma_\zeta$  and  $\nu$ , pertaining to the Student's  $t$  distribution are of less interest in our study. We fix the value of  $\sigma_\zeta$  at one and estimate degrees of freedom  $\nu$  along with the other parameters. We consider only  $\nu = 10$ , as different choices have shown not to affect our simulation results. The entire parameter vector is given by  $\psi = \{\nu, \beta, \gamma, \sigma_\mu, h, \sigma_\eta\}$ .

Each signal listed in Table 1 is analyzed for panel sizes;  $(N = 50, T = 100)$ ,  $(N = 100, T = 50)$ ,  $(N = 100, T = 100)$  and  $(N = 250, T = 250)$ . Further we consider each model with either no missing values and 40% of the values missing. The values are missing at random from the beginning and the end of each panel.

## 5.2 Diagnostic tests for the behavior of the importance sampler

A sufficient condition to guarantee a  $\sqrt{M}$  convergence rate for Monte Carlo estimate  $\hat{p}(y)$  is the existence of a variance in the importance samplings weights  $w^{(i)}$ , as given in equation (15), for  $i = 1, \dots, M$ . Koopman et al. (2009) propose test statistics for evaluating the finiteness of the variance in a sequence of importance weights. Implementation of their suggested Wald type test statistic is done by the following steps.

Simulate a panel  $y$  from the observational density (30), with a signal and parameter values given in Table 1 and panel sizes as discussed in Section 5.1. Next we estimate the parameters using the Monte Carlo maximum likelihood methods of Section 4. Note that for signals 1 and 2 from Table 1 the estimation procedure simplifies as either  $\xi_t$  or  $\mu_i$  is restricted to zero, respectively. The parameter vector  $\psi$  is then replaced by its resulting estimate  $\tilde{\psi}$  and we generate 100,000 importance sampling weights  $w^{(i)}$  using importance densities  $g(\xi|y^l; \hat{\mu})$  and  $g(\mu|\bar{y}^l; \hat{\xi})$ . For a given threshold  $w^{\min}$ , we only consider the weights

that are larger than the threshold. These, say  $s$ , exceedence values  $x_1, \dots, x_s$  are assumed to come from the generalized Pareto distribution with logdensity function  $f(a, b) = -\log b - (1 + a^{-1}) \log(1 + ab^{-1}x_i)$  for  $i = 1, \dots, s$ , where unknown parameters  $a$  and  $b$  determine the shape and scale of the density, respectively. For an appropriately chosen threshold and when  $a \leq 0.5$ , the variance of the importance sampling weights exists. We estimate  $a$  and  $b$  by maximum likelihood, denoted by  $\tilde{a}$  and  $\tilde{b}$ , respectively, and compute the t-test statistic  $t_w = \tilde{b}^{-1} \sqrt{s/3}(\tilde{a} - 0.5)$  for the null hypothesis  $H_0 : a = 0.5$ . As  $r \rightarrow \infty$  and under the null hypothesis, the distribution of the test-statistic converges to the standard normal. We reject the null hypothesis when the statistic is positive and significantly different from zero, that is, when it is larger than 1.96 with 95% confidence.

Durbin & Koopman (1997) argue that the use of antithetic variables improves the efficiency of the importance sampling weights. An antithetic variable in this context is a random draw  $\mu^{(i)}$  or  $\xi^{(i)}$  from the importance densities, that is equiprobable with  $\mu$  or  $\xi$ , respectively, and leads to smaller Monte Carlo variation. For each draw of  $\mu^{(i)}$  and  $\xi^{(i)}$  we use two antithetic variables to balance for location and scale; see Durbin & Koopman (2001, Section 11.9.3) for a detailed discussion.

Figure 1 presents the resulting test statistics for the different panel sizes and signals 1.b, 2.b, 3.b and 3.e, as listed in Table 1. The statistics are computed for different values of threshold  $w^{\min}$ . In particular, we choose  $w^{\min}$  such that the largest 1% to 50% of the weights are included. For each model the test statistics are computed with and without the use of antithetic variables.

The test statistics that are computed without using antithetic variables show mixed evidence. For the signals 1 and 2, that include only an individual-specific or a time-varying effect, the test statistics fail to reject the null-hypothesis for all thresholds. Thus providing evidence in favor of the existence of a variance in the simulated sample of importance weights. For signals including both individual-specific and time-varying effects (3.b and 3.e), the null hypothesis is often rejected for the large panels ( $N = 250$ ,  $T = 250$ ). However, all test statistics that are computed with antithetic variables fail to reject the null-hypothesis, hereby

providing evidence in favor of the existence of a variance in the importance weights. Most test statistics have highly negative values, that only increase marginally when the panel size increases. The presented results also hold for the other signals listed in Table 1. They are further discussed in a technical appendix available from the websites of the authors. The results presented in the remainder of this paper are all computed using antithetic variables.

### 5.3 Simulation Results

For each signal listed in Table 1 and panel sizes  $(N = 50, T = 100)$ ,  $(N = 50, T = 100)$ ,  $(N = 100, T = 100)$  and  $(N = 250, T = 250)$ , we simulate 100 data panels. For each simulated data panel, which is then treated as the observed panel  $y$ , we evaluate the likelihood. The evaluation procedure is implemented as discussed in Section 4 and by using  $M = 500$  draws from the importance densities. We considered implementations based on vector series  $y_t$  and  $\bar{y}_i$ , as well as on collapsed vector series  $y_t^l$  and  $\bar{y}_i^l$ , for  $i = 1, \dots, N$  and  $t = 1, \dots, T_i$ , as discussed in Section 4.2. The average evaluation times are presented in Table 2. The likelihood evaluation procedure based on the collapsed vectors is between 2 and 4 times faster compared to evaluation without collapsing the vectors.

For each simulated panel  $y$  we estimated the corresponding parameters  $\psi$  using the estimation method with collapsed vectors. From the set of estimated parameters we report the average bias and standard deviation in Table 3 for signals 1.b, 2.b, 3.b and 3.e. The results of our Monte Carlo study show that the estimation procedure is successful. All parameter estimates center around their “true” values for all different models and parameter values. Important is that individual state dependence, as captured by  $\phi(B)y_{i,t}$ , can be empirically identified and separated from stochastic components  $\mu_i$  and  $\xi_t$ .

From each simulated data panel we also removed 40% of the observations, at the beginning and end of the data set, creating unbalanced panels. The parameter estimates remained unbiased and the standard errors increase slightly. The full set of parameter estimation results, with and without missing values, are presented in a technical appendix available from the websites of the authors. Here we also provide implementation code written in the Ox

programming language version 6.10 of Doornik (2007).

## 6 Empirical evidence for economic growth

We present an empirical study for a panel of economic growth variables for countries listed in the Penn World Tables (PTW) version 6.3, see Heston, Summers & Aten (2009). A large body of literature has made use of panel data models to analyze economic growth data for different subsets of countries, see for example Islam (1995), Caselli, Esquivel & Lefort (1996), Lee, Pesaran & Smith (1997), Lee, Pesaran & Smith (1998), Hoeffler (2002), Ho (2006), Pesaran (2007), Juarez & Steel (2010) and Moral-Benito (2010). The majority of the panel data growth literature has concentrated on estimating variants of the neo-classical growth model, or Solow growth model; see Solow (1956). The formal empirical testing of this model is pioneered in Mankiw et al. (1992), for the cross-section case, and in Islam (1995) for the panel data case.

Throughout our study we consider a single sample of countries consisting of time series observations pertaining to all countries for which the per capita output variable was available in PTW 6.3<sup>2</sup>. The total sample includes 187 countries that have observations from 1952 until 2007. The resulting panel is highly unbalanced. For example, in the initial year 1952 there are observations recorded for 53 countries, whereas in 2007 there are observations recorded for 187 countries.

### 6.1 Student's $t$ dynamic panel data model

In our study we define  $y_{i,t} = 100(\log Y_{i,t} - \log Y_{i,t-1})$ , where  $Y_{i,t}$  is the per capita output of country  $i$  in year  $t$ . A model for growth rate variable  $y_{i,t}$ , that fits within the general model

---

<sup>2</sup>We used variable RGDPL and removed Serbia and Timor Leste as their output per capita was only available for 1 year.

of Section 3, is given by

$$\begin{aligned}
y_{i,t} &= z_{i,t} + \zeta_{i,t}, & \zeta_{i,t} &\sim t(0, \sigma_\zeta, \nu), \\
z_{i,t} &= y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i + \xi_t, & \mu_i &\sim NID(\delta, \sigma_\mu^2), \\
\xi_t &= h\xi_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2),
\end{aligned} \tag{31}$$

for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ , where  $z_{i,t}$  is the signal,  $\zeta_{i,t}$  is the Student's  $t$  distributed observation disturbance term, parameter  $\gamma$  measures state dependence, parameter vector  $\beta$  measures the effect of the explanatory variables  $x_{i,t}$  whereas  $\mu_i$  captures unobserved country-specific effects and  $\xi_t$  captures the common time-varying effects. Variants of this model have been considered in for example Caselli et al. (1996), Ho (2006) and Juarez & Steel (2010). The random country-specific effect  $\mu_i$  is assumed to be normally distributed with mean  $\delta$  and variance  $\sigma_\mu^2$ . The time-varying effect  $\xi_t$  is modeled by an autoregressive process of order 1. Experiments with higher order autoregressive processes did not lead to further model improvements. The disturbance  $\zeta_{i,t}$  is modeled by a Student's  $t$  density with  $\nu$  degrees of freedom and scaling  $\sigma_\zeta$ . For  $\nu \rightarrow \infty$ , the Student's  $t$  density converges to the Gaussian density.

Typically, the vector of explanatory variables  $x_{i,t}$  includes at least three elements; the investment ratio, the population growth rate and the logarithm of the level of output per capita from the previous time period ( $\log Y_{i,t-2}$ ); see Mankiw et al. (1992) and Caselli et al. (1996). These explanatory variables are likely to be correlated with the mean of the growth rate. For example, a higher mean population growth rate will generally lead to a lower mean growth rate. This type of correlation between explanatory variables  $x_{i,t}$  and country-specific effects  $\mu_i$  is unaccounted for in random effects models and leads to biased parameter estimates; see Durlauf et al. (2005) for a discussion of this issue in the context of economic growth. To overcome this issue we standardize the explanatory variables to have mean zero and unit variance for each country. The country-specific effect  $\mu_i$  can therefore be interpreted as the mean growth rate for country  $i$ . It is comparable across countries and corresponds to average covariate values; see Juarez & Steel (2010). This is a simple method for avoiding



biases from correlation between  $x_{i,t}$  and  $\mu_i$ . Other, less restrictive methods can equally well be considered for our framework; see Chamberlain (1984) for a more elaborate discussion.

The parameter estimation results are presented in Table 4. They are computed using the procedures discussed in Section 4. The left panel shows the results for the case where  $\zeta_{i,t}$  is modeled by the Student's  $t$  density. The right panel shows the results for an approximation of the Gaussian density, obtained by fixing  $\nu = 1000$  in the Student's  $t$  density. The results clearly indicate the importance of the Student's  $t$  density with a small  $\nu$ ; the degrees of freedom coefficient  $\nu$  is estimated just above the value 2. The loglikelihood values also indicate that the Student's  $t$  model is preferred over the Gaussian model. Furthermore, the inclusion of the explanatory variables seems important. The corresponding parameter estimates are significant and have their expected sign. Higher savings, lower population growth rates and lower GDP levels all lead to higher growth rates; see Solow (1956) and Mankiw et al. (1992) for a more elaborate discussion concerning the economic interpretation of these results.

Parameter  $\gamma$  measures the structural dependence on the previous growth rate variable. This parameter is estimated as highly significant between 0.1281 and 0.1608 in the different specifications. The interpretation of  $\gamma$  is interesting. If its value approaches zero the growth rate process becomes an independent disturbance process conditional on the country-specific level  $\mu_i$ , the explanatory variables  $x_{i,t}$  and the common time-varying effect  $\xi_t$ . The time trend of the corresponding country then becomes similar to the common time-varying effect. If parameter  $\gamma$  approaches unity, the country-specific growth rate process becomes a unit-root process. This indicates that the country follows its own path rather than the common path as estimated by the common time-varying effect. It may be expected that parameter  $\gamma$  will be different for different countries. This is formally investigated in the Section 6.2.

Figure 2 presents the estimated time-varying effects for the Student's  $t$  and Gaussian models including explanatory variables. Their computation is discussed in Appendix C. The time-varying effect is capable of identifying a general global business cycle as we observe in

Figure 2; the NBER recession periods of the US economy are presented as gray bars<sup>3</sup>. The difference between the Student's  $t$  model and the Gaussian model is clearly visible. The 95 % confidence bounds of the Gaussian approximation are much wider. Only in a small number of years the time-varying effect is significantly different from zero. This is in sharp contrast to the estimate of the time-varying effect for the Student's  $t$  model for which a more defined business cycle emerges.

## 6.2 Parameter heterogeneity

In this section we consider model (31) and we replace parameter  $\gamma$  in the  $i^{\text{th}}$  equation by  $\gamma_i$ , for  $i = 1, \dots, N$ . In particular, we treat  $\gamma_i$  as a random country-specific effect. The approach of randomizing parameters is discussed in the context of linear regression models by Hsiao & Pesaran (2008). A benefit of the approach is that few additional parameters are necessary to make  $\gamma$  country-specific. The parsimonious implementation is important in our modeling framework since the Monte Carlo maximum likelihood estimate of parameter vector  $\psi$  is obtained by maximizing  $\log \hat{p}(y)$  in (14) with respect to  $\psi$  using numerical methods. When the number of distinct parameters increases, these methods become numerically more challenging.

The proposed extension can be cast into our general framework by setting

$$y_{i,t} = z_{i,t} + \zeta_{i,t}, \quad z_{i,t} = x'_{i,t}\beta + a'_{i,t}\mu_i + \xi_t, \quad a_{i,t} = (y_{i,t-1}, 1)', \quad \mu_i = (\gamma_i, \dot{\mu}_i)', \quad (32)$$

where  $\mu_i$  is here a vector of two country-specific effects, which are normally distributed with mean  $\delta$  and diagonal variance matrix  $\Sigma_\mu$ . The country-specific effect  $\dot{\mu}_i$  is the mean component allowing for level differences in the growth rate variables, which was the role of the univariate  $\mu_i$  in model (31). The vector  $a_{i,t}$  is known at time  $t$  and can be treated as fixed. The other components in model (32) are treated in a similar way as in our previous analysis. Based on our initial results of Section 6.1 we model disturbances  $\zeta_{i,t}$  in model (32)

---

<sup>3</sup>The recession dates are obtained from <http://www.nber.org/cycles.html>

by the Student's  $t$  density.

The estimates of the country-specific effects are visually displayed in Figure 3. They are estimated as discussed in Appendix C. We show the estimated state dependence parameters and the estimated mean growth rates. The borders of countries that are not included in our sample, such as Burma and Western Sahara, are not displayed on the maps. For 146 countries (out of 187) the estimate for the mean growth rate is significantly different from zero (with a 95% confidence level). For the state dependence parameter this result is obtained for 112 countries. The estimated time-varying effect (not shown) is identical as in Section 6.1; see Figure 2. Further details for individual countries are presented in the technical appendix.

Several interesting facts emerge from Figure 3. Most of which are in agreement with the stylized facts discussed in Durlauf et al. (2005). Between 1952 and 2007 most countries are found to have positive mean growth rates. However large differences in magnitude are displayed. High growth rates are predominantly found for countries located in East and Southeast Asia, whereas low growth rates (sometimes negative) are found for countries from the sub-Saharan African region. The “miracle” growth countries such as Taiwan, Republic of Korea, Botswana, Cyprus, Ireland and Portugal, are also nicely identified by our model.

Interesting variation is found between the estimates for state dependence, for model (32). The countries that are associated with former communist or dictatorial regimes show high levels of state dependence. The typical examples are China, Russia, Vietnam and Cambodia. Also, countries with unique export products tend to show high levels of state dependence as their economies are less affected by global trends. A prominent example is Botswana, where most of the economy relies on the export of minerals such as diamonds.

Western European countries show mixed results. All mean growth rates are positive and typically above 2%. The exception is Germany, which is likely due to its reunification. Eastern European countries also show high growth rates. These countries appear in our sample after they have become independent from the former Soviet-Union. The high growth rates are likely based on the last few years in which these countries have received economic support from the United States and Western European countries. The estimates for  $\gamma_i$  for

these countries are also quite high indicating that their economies have not fully integrated to the global economy.

For Africa, we have obtained a high variety of mean growth rates and of levels of state dependence. The African countries can be divided into roughly two groups. The first group has a significant positive mean growth rate. These are the countries for which we suspect that they are catching up. Typical examples are Cape Verde, Ghana and Sri Lanka. The second group consists of countries with very low and negative mean growth rates. Typical examples are Djibouti, Eritrea and Somalia.

The countries that have close political, economic and trade links with the United States show estimates similar to the U.S. The key examples are Australia, United Kingdom and The Netherlands. Countries such as Mexico, Canada and many other Central-American countries also have similar estimates compared to the United States. For South-America we have obtained lower growth rates, except for Chili.

### 6.3 Time-varying effects

Next we extend the model to include additional time-varying effects. So far we assumed that all countries are equally affected by a univariate common time-varying effect  $\xi_t$ . The estimated effect was able to identify the general global (US) business cycle; see Figure 2. Here we investigate whether different time-varying effects can be detected for different groups of countries.

On the basis of the estimated country-specific effects of Section 6.2, we identify nine groups of countries: Western Europe, Eastern Europe, Middle East, Africa high, Africa low, US and related, South-America, East Asia and the rest of the world. The groups and their countries are listed in Appendix D.

The extended model is summarized in our general framework as follows

$$y_{i,t} = z_{i,t} + \zeta_{i,t}, \quad z_{i,t} = x'_{i,t}\beta + a'_{i,t}\mu_i + b'_{i,t}\xi_t, \quad a_{i,t} = (y_{i,t-1}, 1)', \quad \mu_i = (\gamma_i, \dot{\mu}_i)', \quad (33)$$

where  $\xi_t$  is the  $9 \times 1$  vector of time-varying effects and  $b_{i,t}$  is set equal to the  $j$ th column of the  $9 \times 9$  identity matrix if country  $i$  is in group  $j$ . The time-varying effects are modeled as nine independent autoregressive processes of order one with persistence parameter  $h^j$  and scaling parameter  $\sigma_\eta^j$ .

Figure 4 presents the estimated time-varying effects for the nine country groups. A large variety of different growth patterns is found. The Western European and United States groups show similar cyclical patterns indicating the presence of a business cycle. The African groups (high and low) show that within one continent at least two different groups of countries are emerging. First, the high growth group consist of countries that started with very low negative growth but are catching up. Second, the low growth group consist of countries that are experiencing a declining growth pattern.

The Eastern European group illustrates convincingly how missing values are handled within our framework. Before 1990, the confidence bounds are relatively wide, because data for only a few countries is available. After 1990 the growth rates increase and the confidence bound become more narrow as more countries are included in the panel. The persistence parameters and scaling parameters of the estimated time-varying effects are presented in Table 5. The Western European and US based groups show lower levels of persistence.

## 7 Conclusion

We have developed a simulation-based methodology for the estimation of parameters in a general class of dynamic panel data models with cross-section and time-varying random effects. The new estimation method for this class of models is developed in this paper. The use of importance sampling and related methods provides the means for a feasible analysis. The computational efficiency of our methods is due to the ability to separate the cross-section effects from the time-varying effects and to collapse high-dimensional vectors to low-dimensional vectors that contain the sufficient statistics relevant for the analysis. Further, the use of the Kalman filter allows for the efficient sampling of the time-varying

effects. In a Monte Carlo study we have given clear evidence of the validity of our estimation methods for finite samples.

Given the generality of the generalized dynamic panel data model, many different models can be designed for many different purposes. We have limited ourselves to illustrate the methodology for a large panel of time series with observations from a Student's  $t$  density. Other possible applications of our model are also hinted and they include dynamic panel data models for binary, count and categorical observations. Such models can be relevant in economic studies but also in environmental or educational policy studies which need to be based on high-dimensional panel data models. Although our illustration for the Penn World Tables have used a time dimension of  $T = 56$ , the methods can also be used for data sets with smaller or larger time spans.

In our current modeling framework we let the signal be dependent on cross-section and time effects in a linear way (effects are additive). Further flexibility can be introduced by having a signal that depends on the two effects in a nonlinear way (effects are multiplicative). This extension leads to an even more general class of dynamic panel data models. It requires further amendments in our methodology of estimation. A motivation to pursue this plan for further research is the ability to estimate time-varying effects (or dynamic factors) and their associated (cross-sectional or factor) loadings simultaneously. The generalized dynamic factor model is an example of a model in this class and it can be analyzed by using the importance sampling methods developed in this paper. We expect that the computational efficiency is not affected by this modification.

## Appendix A

Based on vector representation (18), with  $\bar{\mathcal{B}}_i$  replaced by  $\bar{\mathcal{B}}_i^* = (b'_{i,1}\xi_1^*, \dots, b'_{i,T}\xi_T^*)'$ , we calculate conditional expectation  $E_g(\mu|y; \xi^*)$ , as needed in step (iii) of Algorithm A, by using a

standard lemma from multivariate normal regression. There holds that

$$\mathbb{E}_g(\mu|y; \xi^*) = \mathbb{E}_g(\mu; \xi^*) + \text{Cov}_g(\mu, y; \xi^*) \text{Var}_g(y; \xi^*)^{-1} [y - \mathbb{E}_g(y; \xi^*)],$$

which can be solved separately for each element  $\mathbb{E}_g(\mu_i|\bar{y}_i; \xi^*)$ , as given  $\xi^*$ ,  $\mu_i$  only depends on  $y$  by means of  $\bar{y}_i$ . Some simple manipulations give

- $\mathbb{E}_g(\mu_i; \xi^*) = 0$ ;
- $\text{Cov}_g(\mu_i, \bar{y}_i; \xi^*) = \Sigma_\mu \bar{A}_i'$ ;
- $\text{Var}_g(\bar{y}_i; \xi^*)^{-1} = \bar{D}_i^{-1} - \bar{D}_i^{-1} \bar{L}_i (\bar{L}_i' \bar{D}_i^{-1} \bar{L}_i + I_q)^{-1} \bar{L}_i' \bar{D}_i^{-1}$ , where  $\bar{L}_i = \bar{A}_i \cdot \text{choleski}(\Sigma_\mu)$ , see Roy & Sarhan (1956) and Roy (1958);
- $\mathbb{E}_g(\bar{y}_i; \xi^*) = \bar{c}_i + \bar{\mathcal{B}}_i^*$ .

Efficient implementation of the calculated can be accomplished without storing variance matrices  $\text{Var}_g(\bar{y}_i; \xi^*)$  or its inverses.

## Appendix B

Based on vector representation (19), with  $\mathcal{A}_t$  replaced by  $\mathcal{A}_t^*(a'_{1,t}\mu_1^*, \dots, a'_{N,t}\mu_N^*)'$ , the calculation of expected value  $\mathbb{E}_g(\xi|y; \mu^*)$  in step (v) of Algorithm A is carried out using the Kalman filter and smoothing methods; see Anderson & Moore (1979) and Durbin & Koopman (2001, Chapter 4). Moreover, since  $D_t$  is diagonal the fast Kalman filter and smoothing methods from Koopman & Durbin (2003) can be used.

## Appendix C

Given the estimated parameter vector  $\tilde{\psi}$  we calculate Monte Carlo estimates of the individual-specific and time-varying effects. A more detailed discussion of this approach is given in

Durbin & Koopman (2001, Chapter 11). Let  $f(\mu, \xi)$  denote a general function of  $\mu$  and  $\xi$  that is of interest. It holds that

$$E_p[f(\mu, \xi)|y] = \int_{\xi} \int_{\mu} f(\mu, \xi) p(\mu, \xi|y; x) \, d\mu \, d\xi,$$

where  $E_p[\cdot|y]$  refers to the expectation with respect to the density  $p(\mu, \xi|y; x)$ . For given modal values  $\hat{\mu}$  and  $\hat{\xi}$ , the accompanying importance sampling representation is given by

$$E_p[f(\mu, \xi)|y] = p(y)^{-1} \int_{\xi} \int_{\mu} f(\mu, \xi) \frac{p(y|\mu, \xi; x)p(\mu)p(\xi)}{g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu})} g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi.$$

When applying Bayes rule twice to the right hand side we obtain

$$E_p[f(\mu, \xi)|y] = \frac{g(y; \hat{\xi})g(y; \hat{\mu})}{p(y)} \int_{\xi} \int_{\mu} f(\mu, \xi) w(y, \mu, \xi; \hat{\mu}, \hat{\xi}) g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi,$$

where

$$w(y, \mu, \xi; \hat{\mu}, \hat{\xi}) = \frac{p(y|\mu, \xi; x)}{g(y|\mu; \hat{\xi})g(y|\xi; \hat{\mu})}.$$

Now, when setting  $f(\mu, \xi) = 1$  we obtain

$$1 = \frac{g(y; \hat{\xi})g(y; \hat{\mu})}{p(y)} \int_{\xi} \int_{\mu} w(y, \mu, \xi; \hat{\mu}, \hat{\xi}) g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi.$$

And when dividing the two equations above we get

$$E_p[f(\mu, \xi)|y] = \frac{\int_{\xi} \int_{\mu} f(\mu, \xi) w(y, \mu, \xi; \hat{\mu}, \hat{\xi}) g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi}{\int_{\xi} \int_{\mu} w(y, \mu, \xi; \hat{\mu}, \hat{\xi}) g(\mu|y; \hat{\xi})g(\xi|y; \hat{\mu}) \, d\mu \, d\xi},$$

for which a Monte Carlo estimate  $\tilde{f}(\mu, \xi)$  is given by

$$\tilde{f}(\mu, \xi) = \frac{\sum_{i=1}^M f(\mu^{(i)}, \xi^{(i)}) w^{(i)}}{\sum_{i=1}^M w^{(i)}},$$



where  $w^{(i)}$  is defined in equation (15).

## Appendix D

Decomposition of the world:

### **Western Europe**

Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom;

### **Eastern Europa**

Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, Latvia, Lithuania, Macedonia, Moldova, Montenegro, Poland, Romania, Russia, Slovak Republic, Slovenia, Turkey, Ukraine;

### **Middle East**

Bahrain, Brunei, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, United Arabic Emirates, Yemen;

### **Africa high positive growth $\mu_i > 0.5$**

Algeria, Angola, Benin, Burkina Faso, Cape Verde, Republic of the Congo, Egypt, Equatorial Guinea, Ethiopia, Gabon, Botswana, Ghana, Guinea Bissau, Kenya, Kiribati, Lesotho, Liberia, Libya, Malawi, Mali, Mauritania, Morocco, Mozambique, Nigeria, Rwanda, Sierra Leone, South Africa, Sri Lanka, Sudan, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zimbabwe;

### **Africa low and negative growth $\mu_i < 0.5$**

Burundi, Cameroon, Central African Republic, Chad, Democratic Republic of the Congo, Ivory Coast, Djibouti, Eritrea, Gambia, Guinea, Madagascar, Namibia, Nicaragua, Niger, Senegal, Somalia, Zambia;

### **US and related countries**

Antigua and Barbuda, Australia, Canada, Costa Rica, Dominica, Dominican Republic, Guatemala, Honduras, Mexico, New Zealand, Panama, Puerto Rico, St. Kitts and Nevis, St. Lucia, St. Vincent and Grenadines, Trinidad and Tobago, United States;

### **South America**

Argentina, Belize, Bolivia, Chile, Colombia, Ecuador, Brazil, Paraguay, Peru, Suriname, Uruguay, Venezuela;

### **East Asia**

Cambodia, China, Hong Kong, India, Indonesia, Japan, South Korea, Laos, Macao, Malaysia, Philippines, Singapore, Taiwan, Thailand, Vietnam;

### **Rest**

Afghanistan, Bahamas, Bangladesh, Barbados, Bermuda, Bhutan, Comoros, Cuba, El Salvador, Fiji, Grenada, Guyana, Haiti, Jamaica, Kyrgyzstan, Maldives, Marshall Islands, Mauritius, Micronesia, Fed. Sts. ,Mongolia, Nepal, Pakistan, Palau, Papua New Guinea, Samoa, Sao Tome and Principe, Seychelles, Solomon Islands, Tajikistan, Tonga, Turkmenistan, Uzbekistan, Vanuatu.

## **References**

Anderson, B. D. O. & Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.

- Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, John Wiley & Sons, New York.
- Caselli, F., Esquivel, G. & Lefort, F. (1996), ‘Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics’, *Journal of Economic Growth* **1**, 363–389.
- Chamberlain, G. (1984), Panel Data, in Z. Griliches & M. D. Intriligator, eds, ‘Handbook of Econometrics, vol. 2’, Elsevier, Amsterdam, North-Holland, pp. 1247–1318.
- De Long, J. B. & Summers, L. H. (1991), ‘Equipment Investment and Economic Growth’, *The Quarterly Journal of Economics* **106**, 445–502.
- Doornik, J. A. (2007), *Object-Oriented Matrix Programming Using Ox*, Timberlake Consultants Press, London.
- Duffie, D., Saita, L. & Wang, K. (2007), ‘Multi-period corporate default prediction with stochastic covariates’, *Journal of Financial Economics* **83**, 635–665.
- Durbin, J. & Koopman, S. J. (1997), ‘Monte Carlo maximum likelihood estimation of non-Gaussian state space models’, *Biometrika* **84**, 669–684.
- Durbin, J. & Koopman, S. J. (2000), ‘Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives’, *Journal of the Royal Statistical Society, Series B* **62**, 3–56.
- Durbin, J. & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.
- Durbin, J. & Koopman, S. J. (2002), ‘A simple and efficient simulation smoother for state space time series analysis’, *Biometrika* **89**, 603–616.
- Durlauf, S. N., Johnson, P. & Temple, J. (2005), Growth Econometrics, in P. Aghion & S. N. Durlauf, eds, ‘Handbook of Economic Growth, vol. 1, Part A,’ Elsevier, Amsterdam, North-Holland, pp. 555–677.

- Fruhwirth-Schnatter, S. & Kaufmann, S. (2008), ‘Model-Based Clustering of Multiple Time Series’, *Journal of Business and Economic Statistics* **26**, 78–89.
- Geweke, J. F. (1989), ‘Bayesian Inference in Econometric Models Using Monte Carlo Integration’, *Econometrica* **57**, 1317–1339.
- Geweke, J. F. (1991), ‘Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints’, *Computer Science and Statistics. Proceedings of the 23rd Symposium on the Interface. Seattle Washington, April 21-24, 1991* pp. 571–578.
- Geweke, J. & Keane, M. (2001), Computationally Intensive Methods for Integration in Econometrics, *in* J. Heckman & E. Leamer, eds, ‘Handbook of Econometrics,V’, pp. 3465–3568.
- Hajivassiliou, V. (1990), ‘Smooth Simulation Estimation of Panel Data LDV Models’. Mimeo, Yale University.
- Heckman, J. J. (1981*a*), Heterogeneity and State Dependence, *in* S. Rosen, ed., ‘Studies of Labor Markets’, The University of Chicago Press, Chicago: The National Bureau of Economic Research, pp. 91–140.
- Heckman, J. J. (1981*b*), Statistical Models for Discrete Panel Data, *in* C. F. Manski & D. L. McFadden, eds, ‘Structural Analysis of Discrete Data and Econometric Applications’, The MIT Press, Cambridge, pp. 113–177.
- Heston, A., Summers, R. & Aten, B. (2009), ‘Penn world table version 6.3’. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- Ho, T. W. (2006), ‘Income Thresholds abd Growth Convergence: A Panel Data Approach’, *The Manchester School* **74**, 170–189.

- Hoeffler, A. E. (2002), ‘The Augmented Solow Model and the African Growth Debate’, *Oxford Bulletin of Economics and Statistics* **64**, 135–158.
- Hsiao, C. (2003), *Analysis of Panel Data*, Cambridge University Press, Cambridge.
- Hsiao, C. & Pesaran, M. H. (2008), Random Coefficients Models, *in* L. Mátyás & P. Sevestre, eds, ‘The Econometrics of Panel Data’, Springer Verlag, Berlin Heidelberg, pp. 187–216.
- Hyslop, D. R. (1999), ‘State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women’, *Econometrica* **67**, 1255–1294.
- Islam, N. (1995), ‘Growth Empirics: A Panel Data Approach’, *Journal of Economic Surveys* **17**, 1127–1170.
- Juarez, M. A. & Steel, M. F. J. (2010), ‘Model-Based Clustering of Non-Gaussian Panel Data Based on Skew- $t$  Distributions’, *Journal of Business and Economic Statistics* **28**, 52–66.
- Jungbacker, B. & Koopman, S. J. (2007), ‘Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models’, *Biometrika* **94**, 827–839.
- Jungbacker, B. & Koopman, S. J. (2008), ‘Likelihood-based Analysis for Dynamic Factor Models’. Tinbergen Institute Discussion Paper: 08-007/4.
- Jungbacker, B., Koopman, S. J. & van der Wel, M. (2011), ‘Maximum likelihood estimation for dynamic factor models with missing data’, *Journal of Economic Dynamics and Control* **35**, 1358–1368.
- Keane, M. P. (1994), ‘A Computationally Practical Simulation Estimator for Panel Data’, *Econometrica* **62**, 95–116.
- Keane, M. P. & Sauer, R. M. (2009), ‘Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior’, *Econometrica* **77**, 975–991.
- Koopman, S. J. & Durbin, J. (2003), ‘Filtering and smoothing of state vector for diffuse state space models’, *Journal of Time Series Analysis* **24**, 85–98.

- Koopman, S. J. & Lucas, A. (2008), ‘A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk’, *Journal of Business and Economic Statistics* **26**, 510–525.
- Koopman, S. J., Lucas, A. & Scharth, M. (2011), ‘Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models’. Tinbergen Institute Discussion Paper: 11-057/4.
- Koopman, S. J., Shephard, N. & Creal, D. D. (2009), ‘Testing the assumptions behind importance sampling’, *Journal of Econometrics* **149**, 2–11.
- Lee, K., Pesaran, H. M. & Smith, R. (1997), ‘Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model’, *Journal of Applied Econometrics* **1997**, 357–392.
- Lee, K., Pesaran, H. M. & Smith, R. (1998), ‘Growth Empirics: A Panel Data Approach – A Comment’, *The Quarterly Journal of Economics* **1998**, 319–323.
- Lee, L. F. (1997), ‘Simulated maximum likelihood estimation of dynamic discrete choice models, Some Monte Carlo results’, *Journal of Econometrics* **82**, 1–35.
- Liesenfeld, R. & Richard, J. (2008), Simulation Techniques for Panels: Efficient Sampling, in L. Mátyás & P. Sevestre, eds, ‘The Econometrics of Panel Data’, Springer Verlag, Berlin Heidelberg, pp. 215–246.
- Mankiw, N. G., Romer, D. & Weil, D. N. (1992), ‘A Contribution to the Empirics of Economic Growth’, *The Quarterly Journal of Economics* **107**, 407–437.
- Monahan, J. F. (2001), *Numerical Methods of Statistics*, Cambridge University Press, Cambridge.
- Moral-Benito, E. (2010), ‘Determinants of Economic Growth: A Bayesian Panel Data Approach’, *Review of Economics and Statistics* . Forthcoming.
- Nocedal, J. & Wright, S. J. (1999), *Numerical Optimization*, Springer Verlag, New York.

- Pesaran, H. M. (2007), ‘A Pair-Wise Approach To Testing For Output And Growth Convergence’, *Journal of Econometrics* **138**, 312–355.
- Richard, J. F. & Zhang, W. (2007), ‘Efficient High-Dimensional Importance Sampling’, *Journal of Econometrics* **141**, 1385–1411.
- Ripley, B. D. (1987), *Stochastic Simulation*, John Wiley & Sons, New York.
- Roy, S. N. (1958), ‘A Note On Inverting a Certain Type of Patterned Matrix’. United States. Air Force Office of Scientific Research Technical Notes No. 191.
- Roy, S. N. & Sarhan, A. E. (1956), ‘On Inverting a Class of Patterned Matrices’, *Biometrika* **43**, 227–231.
- Shephard, N. & Pitt, M. K. (1997), ‘Likelihood analysis of non-Gaussian measurement time series’, *Biometrika* **84**, 653–667.
- So, M. K. P. (2003), ‘Posterior mode estimation for nonlinear and non-Gaussian state space models’, *Statistica Sinica* **13**, 255–274.
- Solow, R. (1956), ‘A contribution to the theory of economic growth’, *The Quarterly Journal of Economics* **70**, 65–94.
- Stock, J. H. & Watson, M. W. (2002), ‘Macroeconomic Forecasting Using Diffusion Indexes’, *Journal of Business and Economic Statistics* **220**, 147–162.
- Temple, J. (1999), ‘The New Growth Evidence’, *Journal of Economic Literature* **37**, 112–156.
- Woolridge, J. M. (2005), ‘Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity’, *Journal of Applied Econometrics* **20**, 39–54.

Signal			Parameters					
			$\nu$	$\beta$	$\gamma$	$\sigma_\mu$	$h$	$\sigma_\eta$
1;	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i$	a;	10	1	0.2	0.5	-	-
		b;	10	1	0.2	1	-	-
		c;	10	1	0.2	3	-	-
2;	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \xi_t$	a;	10	1	0.2	-	0.3	0.2
		b;	10	1	0.2	-	0.9	0.2
3;	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i + \xi_t$	a;	10	1	0.2	0.5	0.3	0.2
		b;	10	1	0.2	0.5	0.9	0.2
		c;	10	1	0.2	1	0.3	0.2
		d;	10	1	0.2	1	0.9	0.2
		e;	10	1	0.2	3	0.3	0.2
		f;	10	1	0.2	3	0.9	0.2

Table 1: Signal specifications and parameter values for simulating observations in the Monte Carlo study. The DGP is further given by  $y_{i,t} \stackrel{i.i.d.}{\sim} t(z_{i,t}, 1, \nu)$ ,  $x_{i,t} \sim NID(0, 1)$ ,  $\mu_i \sim NID(0, \sigma_\mu^2)$ ,  $\xi_t = \alpha_t$ ,  $\alpha_{t+1} = h\alpha_t + \eta_t$  and  $\eta_t \sim NID(0, \sigma_\eta^2)$ . The initial time varying effect is taken  $N(0, \sigma_\eta^2/(1 - h^2))$ .



Signal	$(N = 100, T = 50)$		$(N = 50, T = 100)$		$(N = 100, T = 100)$		$(N = 250, T = 250)$	
	red.	not red.	red.	not red.	red.	not red.	red.	not red.
1.a	11.99	21.47	14.11	24.02	24.78	50.65	132.68	250.70
1.b	14.51	24.69	13.81	26.95	25.86	50.46	135.36	250.42
1.c	14.20	25.08	13.60	27.08	24.41	51.70	132.60	250.76
2.a	18.44	68.61	18.69	62.05	36.98	129.79	195.52	791.37
2.b	19.83	69.91	19.70	60.22	38.21	124.46	185.38	794.19
3.a	33.72	107.06	37.53	99.21	71.57	203.90	480.65	1377.32
3.b	45.07	120.97	46.60	116.64	75.04	243.38	482.59	1596.09
3.c	40.78	120.29	35.56	110.86	78.48	230.26	510.27	1401.38
3.d	43.42	137.35	42.63	132.88	79.09	269.05	500.52	2059.23
3.e	44.30	118.61	35.01	105.16	75.85	216.41	520.26	1423.87
3.f	41.92	144.78	37.79	155.44	86.28	283.84	584.46	1861.93

Table 2: Average likelihood evaluation times for the Student’s  $t$  panel data model, given in one hundreds of a second. The signals are taken as in Table 1. For each model the likelihood is evaluated as discussed in Section 4 and by using  $M = 500$  samples from the importance densities. The columns labeled with red. present the evaluation times of the estimation procedure where samples  $\mu^{(i)}$  and  $\xi^{(i)}$  are drawn from  $g(\xi|y^t; \hat{\mu})$  and  $g(\mu|\bar{y}^t; \hat{\xi})$ , respectively. The columns with labels no red. give the evaluation times where samples  $\mu^{(i)}$  and  $\xi^{(i)}$  are drawn from  $g(\xi|y; \hat{\mu})$  and  $g(\mu|y; \hat{\xi})$ , respectively.

$N$	$T$	$\nu$	$\beta$	$\gamma_y$	$\sigma_\mu$	$h$	$\sigma_\eta$
1.b							
100	50	0.467 <sub>1.644</sub>	0.000 <sub>0.013</sub>	0.000 <sub>0.010</sub>	0.000 <sub>0.070</sub>	-	-
50	100	0.442 <sub>1.640</sub>	-0.001 <sub>0.014</sub>	0.001 <sub>0.010</sub>	-0.019 <sub>0.107</sub>	-	-
100	100	0.251 <sub>0.980</sub>	0.000 <sub>0.010</sub>	0.001 <sub>0.007</sub>	-0.015 <sub>0.074</sub>	-	-
250	250	0.097 <sub>0.350</sub>	0.000 <sub>0.004</sub>	0.000 <sub>0.003</sub>	-0.001 <sub>0.045</sub>	-	-
2.b							
100	50	0.305 <sub>1.532</sub>	0.000 <sub>0.015</sub>	0.001 <sub>0.009</sub>	-	-0.023 <sub>0.076</sub>	0.007 <sub>0.027</sub>
50	100	0.271 <sub>1.533</sub>	0.000 <sub>0.013</sub>	0.000 <sub>0.009</sub>	-	-0.011 <sub>0.052</sub>	-0.002 <sub>0.023</sub>
100	100	0.165 <sub>0.927</sub>	0.000 <sub>0.010</sub>	0.001 <sub>0.007</sub>	-	-0.018 <sub>0.055</sub>	-0.005 <sub>0.020</sub>
250	250	0.099 <sub>0.368</sub>	0.000 <sub>0.004</sub>	0.000 <sub>0.003</sub>	-	0.002 <sub>0.031</sub>	-0.001 <sub>0.010</sub>
3.b							
100	50	0.550 <sub>1.704</sub>	-0.001 <sub>0.013</sub>	0.001 <sub>0.010</sub>	-0.002 <sub>0.041</sub>	-0.025 <sub>0.077</sub>	-0.006 <sub>0.027</sub>
50	100	0.436 <sub>1.660</sub>	-0.002 <sub>0.013</sub>	0.001 <sub>0.010</sub>	-0.010 <sub>0.046</sub>	-0.013 <sub>0.055</sub>	-0.001 <sub>0.023</sub>
100	100	0.301 <sub>0.987</sub>	0.000 <sub>0.009</sub>	0.001 <sub>0.006</sub>	-0.005 <sub>0.039</sub>	-0.009 <sub>0.055</sub>	-0.005 <sub>0.019</sub>
250	250	0.130 <sub>0.373</sub>	0.000 <sub>0.004</sub>	0.000 <sub>0.003</sub>	0.001 <sub>0.024</sub>	0.002 <sub>0.031</sub>	0.001 <sub>0.010</sub>
3.e							
100	50	0.566 <sub>1.725</sub>	-0.002 <sub>0.013</sub>	0.001 <sub>0.010</sub>	0.007 <sub>0.226</sub>	-0.019 <sub>0.181</sub>	-0.011 <sub>0.025</sub>
50	100	0.471 <sub>1.669</sub>	-0.001 <sub>0.012</sub>	0.000 <sub>0.010</sub>	-0.019 <sub>0.225</sub>	-0.014 <sub>0.141</sub>	-0.005 <sub>0.022</sub>
100	100	0.323 <sub>0.991</sub>	0.000 <sub>0.009</sub>	0.001 <sub>0.006</sub>	-0.018 <sub>0.224</sub>	-0.020 <sub>0.112</sub>	-0.007 <sub>0.018</sub>
250	250	0.127 <sub>0.370</sub>	0.000 <sub>0.004</sub>	0.000 <sub>0.003</sub>	-0.005 <sub>0.142</sub>	0.003 <sub>0.068</sub>	0.001 <sub>0.010</sub>

Table 3: Simulation results for the Student's  $t$  dynamic panel data models. We present the average bias and in lower case the standard deviation of the parameter estimates resulting from 100 repetitive estimates from different simulated data panels. Signal specifications 1.b, 2.b, 3.b and 3.e from Table 1 together with observation model (30) are used for simulation. All parameters are estimated by procedures outlined in Section 4, with  $M = 500$  draws from importance densities  $g(\xi|y^l; \hat{\mu})$  and  $g(\mu|\bar{y}^l; \hat{\xi})$ , respectively.

	$y_{i,t} \sim t(z_{i,t}, \sigma_\zeta, \nu)$		$y_{i,t} \sim t(z_{i,t}, \sigma_\zeta, 1000)$	
	No Regressors	Regressors	No Regressors	Regressors
$\nu$	2.0079 <sub>0.0113</sub>	2.0120 <sub>0.031</sub>	1000*	1000*
$\sigma_\zeta$	49.379 <sub>35.114</sub>	35.580 <sub>18.482</sub>	7.0814 <sub>0.0577</sub>	6.9203 <sub>0.0564</sub>
$\gamma$	0.1496 <sub>0.0099</sub>	0.1608 <sub>0.0097</sub>	0.1281 <sub>0.0113</sub>	0.1565 <sub>0.0112</sub>
$\delta$	2.0344 <sub>0.2336</sub>	1.8442 <sub>0.2446</sub>	1.7763 <sub>0.3630</sub>	1.3862 <sub>0.9524</sub>
$\sigma_\mu$	1.4177 <sub>0.0986</sub>	1.4510 <sub>0.0989</sub>	1.1416 <sub>0.1252</sub>	1.1674 <sub>0.1208</sub>
$h$	0.5187 <sub>0.1353</sub>	0.5377 <sub>0.1424</sub>	0.7069 <sub>0.1216</sub>	0.9053 <sub>0.0814</sub>
$\sigma_\eta$	0.7293 <sub>0.0914</sub>	0.7536 <sub>0.0937</sub>	0.7750 <sub>0.1314</sub>	0.7710 <sub>0.1367</sub>
$\beta^1$	0*	0.2724 <sub>0.0496</sub>	0*	0.4526 <sub>0.0799</sub>
$\beta^2$	0*	-0.3211 <sub>0.0528</sub>	0*	-0.2304 <sub>0.0842</sub>
$\beta^3$	0*	-0.9075 <sub>0.0669</sub>	0*	-1.8623 <sub>0.1021</sub>
$\log \hat{p}(y)$	-39615	-39504	-41832	-41655
AIC	79244	79028	83676	83328
BIC	79267	79060	83695	83357

Table 4: Parameter estimation results (and standard errors in lower case) for economic growth rate application. The model for growth rate variables  $y_{i,t}$  is determined by signal  $z_{i,t}$ , which is given by  $z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i + \xi_t$ ,  $\mu_i \sim NID(\delta, \sigma_\mu^2)$ ,  $\xi_t = \alpha_t$ ,  $\alpha_t = h\alpha_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, \sigma_\eta)$  and  $\alpha_1 \sim N(0, \sigma_\eta^2/(1-h^2))$ . The parameters are estimated for an unbalanced panel consisting of 188 countries and a maximum of  $T = 56$  time periods from 1952 until 2007. The estimation method is implemented as discussed in Section 4 and by using  $M = 500$  draws from the importance samplers. Parameters  $\beta^1$ ,  $\beta^2$  and  $\beta^3$  measure the effects of the investment ratio, the population growth rate and the level of GDP, respectively. The AIC and BIC criteria are computed as;  $AIC = 2\mathcal{P} - 2\log \hat{p}(y)$  and  $BIC = \mathcal{P} \log N - 2\log \hat{p}(y)$ , where  $\mathcal{P}$  denotes the number of distinct parameters.

Group	$h^j$	$\sigma_\eta^j$
Western Europe	0.3017 <sub>0.1661</sub>	1.2960 <sub>0.17112</sub>
Eastern Europe	0.8343 <sub>0.1060</sub>	1.0664 <sub>0.23715</sub>
Middle East	0.8809 <sub>0.0689</sub>	0.9895 <sub>0.24986</sub>
Africa high	0.9464 <sub>0.0591</sub>	0.5102 <sub>0.14910</sub>
Africa low	0.9705 <sub>0.0310</sub>	0.5471 <sub>0.22568</sub>
US and related	0.5670 <sub>0.2322</sub>	0.9123 <sub>0.20921</sub>
South-America	0.7844 <sub>0.1262</sub>	0.9671 <sub>0.22091</sub>
East Asia	0.7518 <sub>0.2684</sub>	0.9801 <sub>0.37081</sub>
Rest of the world	0.9617 <sub>0.0454</sub>	0.4931 <sub>0.15495</sub>

Table 5: Estimated persistence and scaling parameters for the time-varying effects presented in Figure 4.

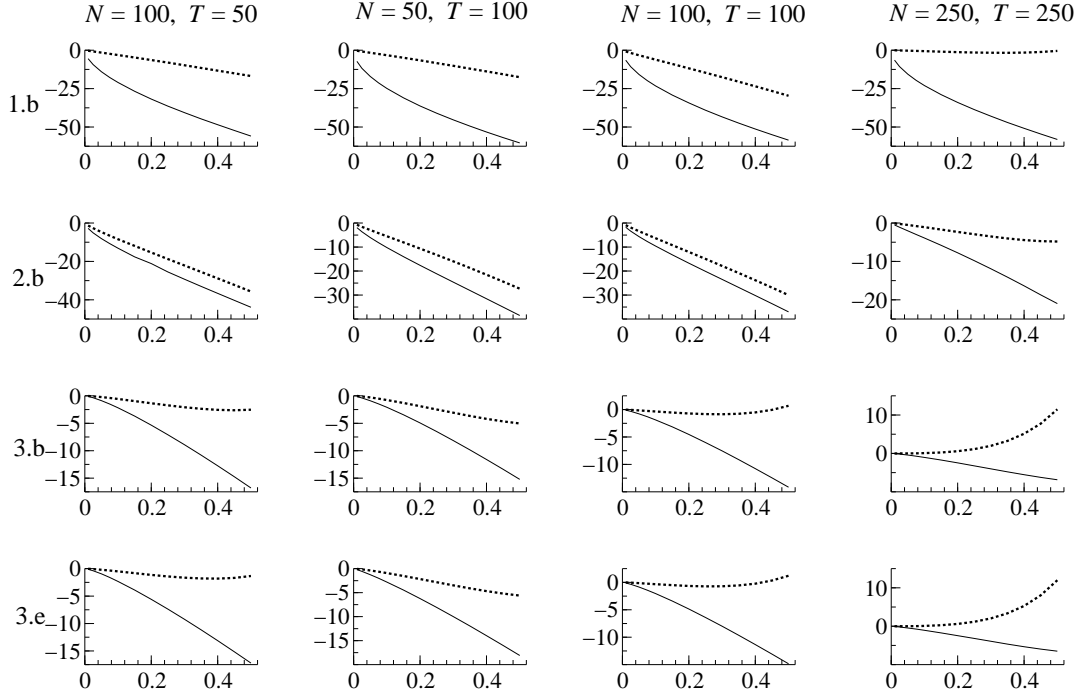


Figure 1: Importance sampling diagnostics for Student's  $t$  dynamic panel data models, based on 100,000 simulations of weights  $w^{(i)}$  defined in equation (15). The test statistic are computed with (solid line) and without (dotted line) the use of antithetic variables. The test statistics are presented for signals 1.b, 2.b, 3.b and 3.e from Table 1 and for different panel sizes. For each combination we computed test statistics for different thresholds  $w^{\min}$ , by procedures explained in Section 5.2. Thresholds are based on the number of exceedence values  $x_1, \dots, x_s$  included. We have taken  $0.01 = s/100000$ ,  $0.025 = s/100000$ ,  $0.05 = s/100000$ , until  $0.5 = s/100000$ .

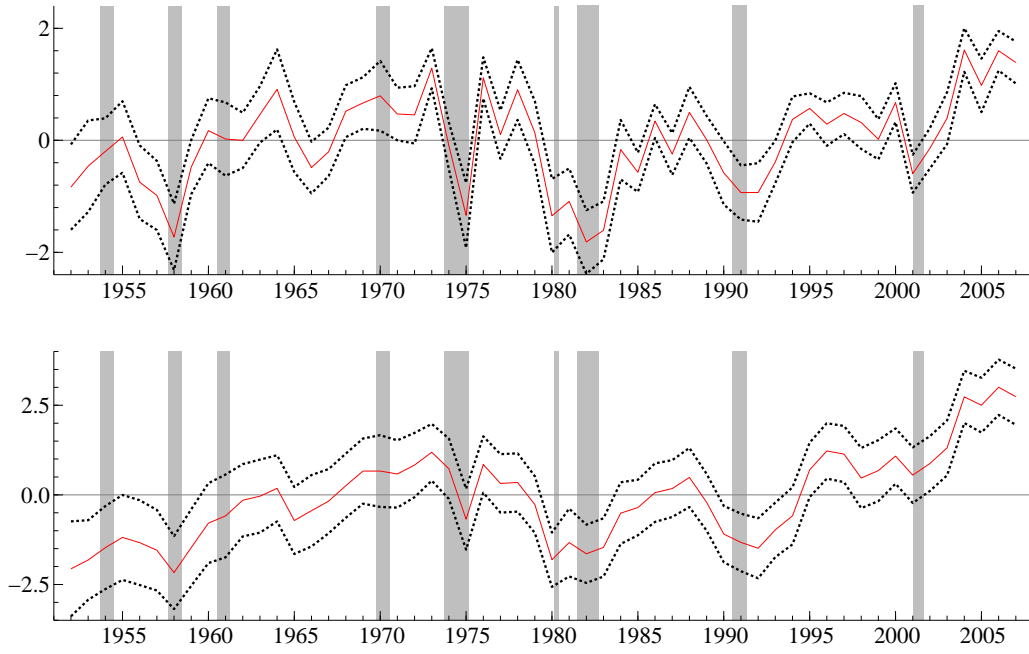


Figure 2: Estimated time-varying effects and 95 % confidence bounds (dotted lines) for economic growth data for the Student's  $t$  (top panel) and Gaussian (lower panel) dynamic panel data models. The effects are computed based on estimated parameters  $\tilde{\psi}$  (given in Table 4) for an unbalanced panel of 187 countries and a total of  $T = 56$  time periods from 1952 until 2007. Further computational details are presented in Appendix C.

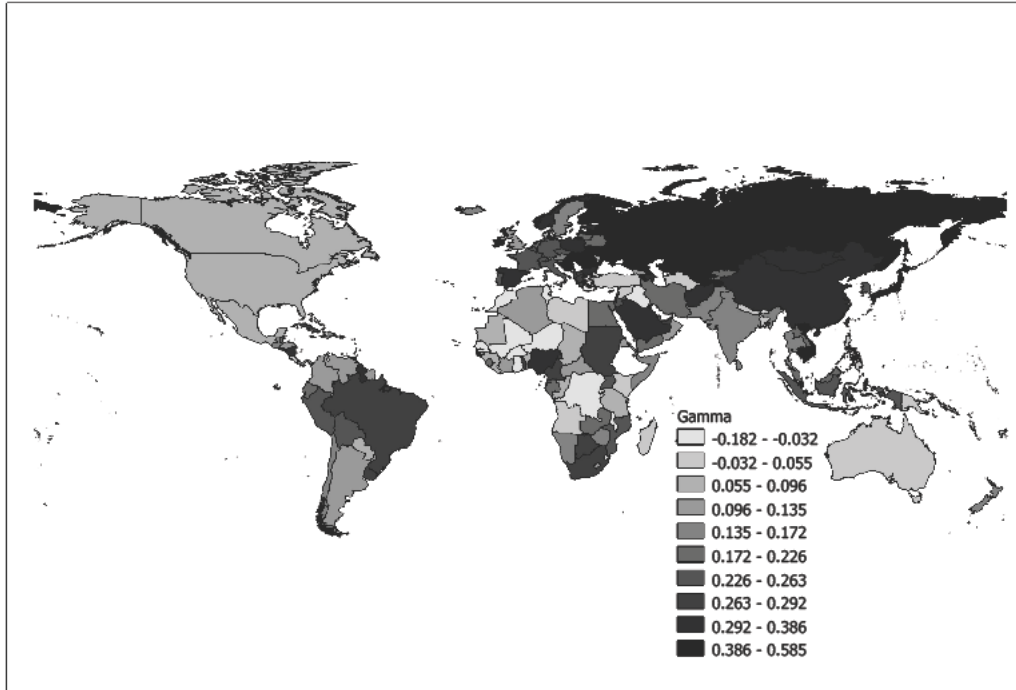
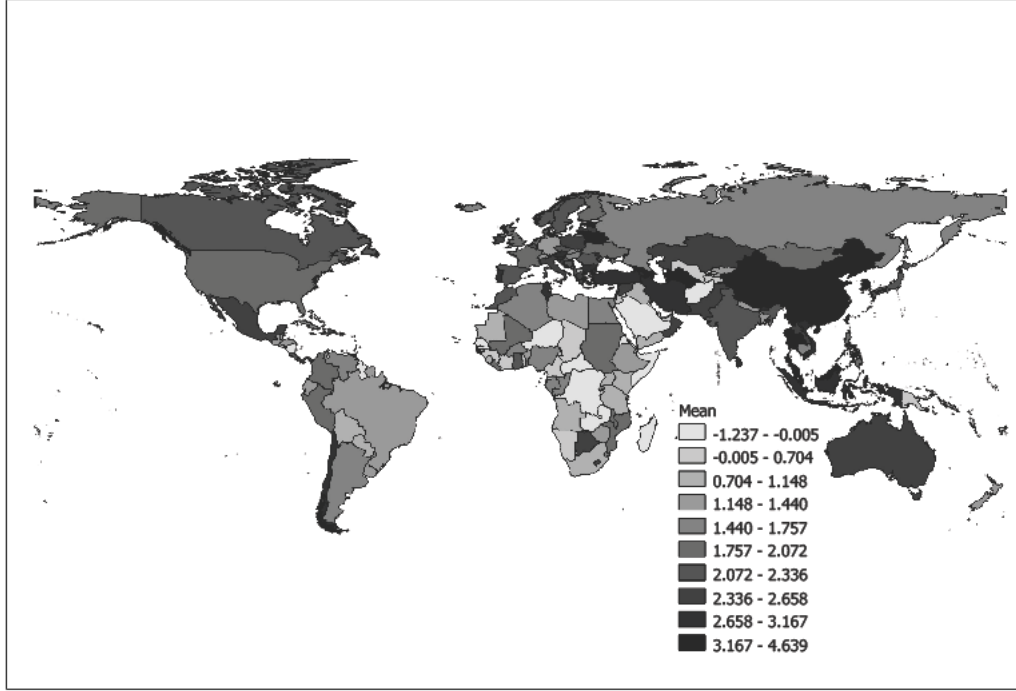


Figure 3: Estimated country-specific effects for economic growth data for the Student's  $t$  dynamic panel data model (32). The top panel shows the estimated mean growth rates and the bottom panel shows the state dependence variables. The random effects are computed based on estimated parameters  $\tilde{\psi}$  for an unbalanced panel of 187 countries and a total of  $T = 56$  time periods from 1952 until 2007. Further computational details are presented in Appendix C.

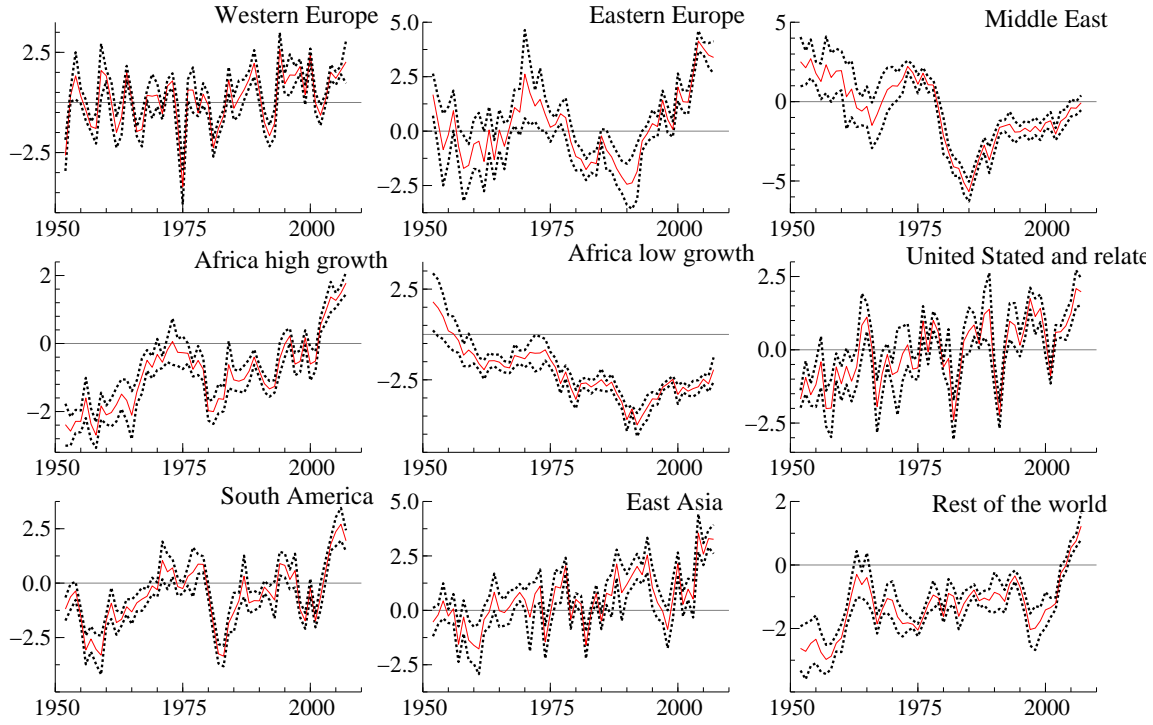


Figure 4: Estimated time-varying effects and 95 % confidence bounds (dotted lines) for economic growth data for the Student's  $t$  dynamic panel data models. The effects are computed for an unbalanced panel of 187 countries and a total of  $T = 56$  time periods from 1952 until 2007. Further computational details are presented in Appendix C. The effects are computed for each group discussed in Appendix D.