Generalized dynamic panel data models with random effects for cross-section and time

CrossMark

G. Mesters [a,b,c,*,1], S.J. Koopman [b,c]

[a] Netherlands Institute for the Study of Crime and Law Enforcement, The Netherlands
[b] Department of Econometrics, VU University, Amsterdam, The Netherlands
[c] Tinbergen Institute, Amsterdam, The Netherlands

ARTICLE INFO

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. In particular, a sequence of conditional importance densities is derived which integrates out all random effects from the joint distribution of endogenous variables. We disentangle the integration over both the cross-section and the time series dimensions. The estimation method facilitates the modeling of large panels in both dimensions. We evaluate the method in an extended Monte Carlo study for dynamic panel data models with observations from different non-Gaussian distributions. We finally present three empirical illustrations for (i) union choice of young males using a Binary panel, (ii) crime rates of families using a Binomial panel and (iii) economic growth modeling using a Student's t panel.

© 2014 Elsevier B.V. All rights reserved.

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. The 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 not trivial for the generalized dynamic panel data model because the likelihood does not exist in closed form. The functional form of the observation density together with the stochastic component of the latent signal prohibit closed form solutions. For a simultaneous analysis of random individual-specific and time-varying effects we extend the methods of Shephard and Pitt (1997) and Durbin and Koopman (1997) which 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 of latent signals. We extend their method for the treatment of 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 for the simulation of latent signals.

We further show that the panel of time series can be collapsed into two low-dimensional vector series, which are used to sample the random individual-specific and time-varying effects from the importance densities. In particular, we collapse the cross-sectional

* Correspondence to: NSCR, PO box 71304, 1008 BH Amsterdam, The Netherlands. Tel.: +31 20 59 89405.
E-mail address: gmesters@nscr.nl (G. Mesters).
1 Technical Appendix: http://personal.vu.nl/s.j.koopman and http://www.geertmesters.nl.

dimension of y without compromising any information that is needed to sample the time-varying effects and visa versa. The transformations are based on modifications of the methods introduced in Jungbacker and Koopman (2008) and lead to large computational savings when evaluating the Monte Carlo likelihood.

For linear dynamic panel data models, transformations can be adopted to eliminate the individual-specific effects, whereas time-varying effects are typically modeled using deterministic functions. Then instrumental variables can be found for the implementation in a generalized method of moments framework, see for example Arellano and Bond (1991), Blundell and Bond (1998). For most non-Gaussian dynamic panel data models no known transformations exist to eliminate the individual-specific effects. Therefore, it is common to assume stochastic specifications for these effects, see the discussion in Wooldridge (2005). In order to integrate the random effects from the complete likelihood, several other Monte Carlo estimation methods have been proposed in the literature. Examples for models without time-varying effects, are simulated maximum likelihood approaches based on the Geweke–Hajivassiliou–Keane (GHK) sampler, see Geweke (1991); Hajivassiliou (1990); Keane (1994), and the more general Markov Chain Monte Carlo methods including Gibbs sampling and the Metropolis–Hastings algorithm, see Geweke and Keane (2001) and Chib (2008). Arellano and Bonhomme (2011) discuss further advances in nonlinear panel data analysis for models without time-varying random effects.

Richard and Zhang (2007) and Liesenfeld and 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 Richard and Zhang (2007) approach in several 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. This allows us to separate the treatment of the different random effects. Second, by transforming the panel into two low-dimensional vector series prior to sampling from the importance densities increasing panel dimensions only marginally impact the overall computational efficiency of our proposed estimation method. Third, the construction of the importance densities differs highly from each other.

Our proposed estimation method for the general model provides a number of 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). Similar improvements, using the EIS methodology for probit panel data models, have been proposed by Liesenfeld et al. (2010) and Liesenfeld and Richard (2010). Second, our framework allows for the simultaneous analysis of unobserved heterogeneity, state dependence and correlated error terms. This is useful as in many empirical panel data studies there are multiple underlying sources of error. Third, parameter heterogeneity can be imposed for our non-Gaussian framework with respect to the time periods as well as the individuals by following the implementations described in Harvey (1989) and Hsiao and Pesaran (2008). They discuss heterogeneous parameters in the context of linear models. Fourth, 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 and Watson (2002).

The remainder of the paper is organized as follows. Section 2 formally describes the generalized dynamic panel data model in detail. In Section 3 we develop our Monte Carlo maximum likelihood method for the general model. Section 4 evaluates the performance of our estimation method in a simulation study. We consider dynamic panel data models with Binary, Binomial and Student's t densities. In Section 5 we present empirical studies for the union choice of young males, the crime rates of families and economic growth rates of countries. The empirical studies highlight the flexibility of our framework with respect to handling missing values, state dependence and parameter heterogeneity. Section 6 summarizes our findings.

2. Statistical model formulation

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 $i = 1, \ldots, N$. 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 general model for variable $y_{i,t}$ is given by

$$y_{i,t} \overset{i.i.d.}{\sim} p(y_{i,t}|z_{i,t}; \psi), \tag{1}$$

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 ψ. We assume that $p(y_{i,t}|z_{i,t}; \psi)$ is possibly non-Gaussian and is correctly specified. In Section 5 we consider the Binary, Binomial and Student's t densities as examples for $p(y_{i,t}|z_{i,t}; \psi)$ and illustrate their empirical relevance. 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

$$z_{i,t} = w_{i,t} + \epsilon_{i,t}, \tag{2}$$

where $w_{i,t}$ is a fixed component and $\epsilon_{i,t}$ is a stochastic component. The components are given by

$$w_{i,t} = x'_{i,t}\beta + \gamma_1 y_{i,t-1} + \cdots + \gamma_{p_y} y_{t-p_y}, \qquad \text{and}$$

$$\epsilon_{i,t} = a'_{i,t}\mu_i + b'_{i,t}\xi_t, \tag{3}$$

where $x_{i,t}$ is a $k \times 1$ vector of observable explanatory variables, β is a $k \times 1$ parameter vector, γ_j is an unknown parameter for $j = 1, \ldots, p_y$ and for some non-negative integer p_y, the $q \times 1$ vector of individual-specific effects μ_i for individual i is weighted at time period t by the $q \times 1$ vector $a_{i,t}$ and the $r \times 1$ vector of time-varying effects ξ_t is weighted for individual i at time period t by $r \times 1$ vector $b_{i,t}$. Both weight vectors, $a_{i,t}$ and $b_{i,t}$, are assumed fixed and may depend on the parameter vector ψ.

The individual-specific effects μ_i and the time-varying effects ξ_t are given by

$$\mu_i \sim \text{NID}(\delta, \Sigma_\mu), \qquad \xi_t = G\alpha_t,$$
$$\alpha_{t+1} = H\alpha_t + R\eta_t, \quad \eta_t \sim \text{NID}(0, \Sigma_\eta), \tag{4}$$

where the individual effects μ_i are assumed normally and independently distributed, with $q \times 1$ common mean vector δ and $q \times q$ variance matrix Σ_μ, which are both considered fixed, the $r \times p$ dimensional matrix G relates the generating linear autoregressive process α_t to the time-varying effects ξ_t, H is a $p \times p$ transition matrix, R is a $p \times l$ disturbance selection matrix, and η_t is an $l \times 1$ vector of disturbances with variance matrix Σ_η. The system matrices G, H, R and Σ_η are fixed and known, although some elements may depend on parameter vector ψ. The initial state vector α_1 is assumed normally distributed with mean zero and variance matrix P. Individual-specific effects μ_i and μ_j are assumed mutually uncorrelated and independent from the time-varying effects, ξ_t, for all $i, j = 1, \ldots, N$ and $t = 1, \ldots, T$.

For the initial signal $z_{i,1}$ we assume that $y_{i,t}$, for $t < 1$, are fixed and known constants for all $i = 1, \ldots, N$. Also, we will assume that parameters β and γ 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 $i, j = i, \ldots, N$ and common time periods t. Section 5 discusses options for relaxing these

assumptions. In particular, there we adopt the methods developed in Chamberlain (1980), Wooldridge (2005), Hsiao and Pesaran (2008) and Harvey (1989).

The specification of the signal allows us to distinguish between two sources capable of explaining the dynamics between particular outcomes, see Heckman (1981b; 1981a). 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 the terms $\gamma_1 y_{i,t-1} + \cdots + \gamma_{p_y} y_{i,t-p_y}$. The second source, referred to by Heckman (1981a) 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 μ_i and time-varying effects ξ_t.

The general model contains many parameters. To identify these parameters in the model different strategies can be considered of which the appropriateness needs to be evaluated on a case by case basis. In general we need to restrict either the distribution of μ_i, ξ_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.

The generalized dynamic panel data model of this paper is fully defined by Eqs. (1)–(4). All parameters are collected in vector ψ 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

$$p(y|z; \psi) = \prod_{i=1}^{N} \prod_{t=1}^{T_i} p(y_{i,t}|z_{i,t}; \psi)$$

$$= \prod_{i=1}^{N} p(y_{i,1}|z_{i,1}; \psi) \prod_{t=2}^{T_i} p(y_{i,t}|\mu_i, \xi_t; x_{i,t}, \mathcal{Y}_{i,t-1}, \psi), \quad (5)$$

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

3. Likelihood evaluation by Monte Carlo integration

This section discusses the method of Monte Carlo maximum likelihood for the estimation of the parameter vector ψ. We first consider the generalized dynamic panel data model for balanced panels, $T_i = T$ for all $i = 1, \ldots, N$. In Section 3.4 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 ψ. In the remainder of this section we drop the dependence on parameter vector ψ for notational convenience. For example, we have $\log p(y) \equiv \log p(y; \psi)$.

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

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

$$= \int_\xi \int_\mu p(y|\mu, \xi; x) p(\mu, \xi) \, d\mu \, d\xi, \quad (6)$$

where the second equality holds as $x = \{x_{i,t}\}$ is deterministic and where $p(\mu, \xi) = p(\mu)p(\xi)$, since μ and ξ are independent. The evaluation of the high dimensional integral (6) 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 using the importance sampling technique, see Ripley (1987).

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

$$p(y) = \int_\xi \int_\mu \frac{p(y|\mu, \xi; x) p(\mu, \xi)}{g(\mu, \xi|y)} g(\mu, \xi|y) \, d\mu \, d\xi, \quad (7)$$

where $g(\mu, \xi|y)$ denotes the importance density. Integral (7) can be solved by Monte Carlo integration for which we sample $\mu^{(i)}$ and $\xi^{(i)}$ from the importance density $g(\mu, \xi|y)$ 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 $g(\mu, \xi|y)$, which accounts for the complicated a-posteriori covariance structure of μ and ξ, sampling from it is likely to be complicated. More specific, as y depends on both μ and ξ, there exists correlation between all individuals (due to ξ) and time periods (due to μ). This makes sampling from densities $g(\mu, \xi|y)$ that account for these correlations infeasible even for moderate panel sizes. As a consequence, the importance sampling based methods of Durbin and Koopman (1997) and Shephard and Pitt (1997), which construct $g(\mu, \xi|y)$ based on an approximating linear Gaussian model, need to be modified in order to obtain a feasible importance sampler.

We propose to integrate out μ by keeping ξ fixed at its posterior modal value and we propose to integrate out ξ by keeping μ fixed at its posterior modal value. When either μ or ξ is fixed the covariance structure of y greatly simplifies. The posterior modal values are chosen for computational convenience. For density $p(y)$ from the generalized dynamic panel data model we propose the following importance sampling representation

$$p(y) = \int_\xi \int_\mu \frac{p(y|\mu, \xi; x) p(\mu) p(\xi)}{g(\mu|y; \hat{\xi}) g(\xi|y; \hat{\mu})} g(\mu|y; \hat{\xi}) g(\xi|y; \hat{\mu}) \, d\mu \, d\xi, \quad (8)$$

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 $p(\mu, \xi|y; x)$, which can be found iteratively as we discuss in Section 3.1. When applying Bayes' rule twice to the right hand side of Eq. (8) we obtain

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

$$\times g(\mu|y; \hat{\xi}) g(\xi|y; \hat{\mu}) \, d\mu \, d\xi, \quad (9)$$

where we have retained the marginal properties of μ and ξ by imposing $g(\xi) = p(\xi)$ and $g(\mu) = p(\mu)$. It holds that $g(y; \hat{\xi})$ and $g(y; \hat{\mu})$ are the conditional likelihoods, which are implicitly defined by the choice for the importance densities. 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 (9) and given by

$$\hat{p}(y) = g(y; \hat{\xi}) g(y; \hat{\mu}) \frac{1}{M} \sum_{i=1}^{M} \frac{p(y|\mu^{(i)}, \xi^{(i)}; x)}{g(y|\mu^{(i)}; \hat{\xi}) g(y|\xi^{(i)}; \hat{\mu})}, \quad (10)$$

where samples $\{\mu^{(1)}, \ldots, \mu^{(M)}\}$ are drawn independently from importance density $g(\mu|y; \hat{\xi})$ and samples $\{\xi^{(1)}, \ldots, \xi^{(M)}\}$ from $g(\xi|y; \hat{\mu})$. The density $p(y|\mu^{(i)}, \xi^{(i)}; x)$ is evaluated using Eq. (5).

The quality of the estimate in Eq. (10) 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. The Lindeberg–Levy central limit theorem implies a $\sqrt{M}$ convergence rate for $\hat{p}(y) \to p(y)$ as $M \to \infty$ if draws from the importance sampler are independent and if importance weights

$$w^{(i)} = p(y|\mu^{(i)}, \xi^{(i)}; x)/\left[g(y|\xi^{(i)}; \hat{\mu}) g(y|\mu^{(i)}; \hat{\xi})\right], \quad (11)$$

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 et al. (2009). In the simulation study of Section 4 we consider diagnostic test statistics for the existence of a variance in a sequence of importance weights drawn from Binary, Binomial and Student's t dynamic panel data models.

3.1. 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 Eq. (10). 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 and Pitt (1997) and Durbin and Koopman (1997, 2000). So (2003) and Jungbacker and 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 μ and ξ.

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

$$y_{i,t} = c_{i,t} + \epsilon_{i,t} + u_{i,t}, \quad u_{i,t} \sim \text{NID}(0, d_{i,t}^2), \quad (12)$$

where $c_{i,t}$ is a fixed constant, stochastic component $\epsilon_{i,t}$ is given by Eq. (3) and $u_{i,t}$ is a random variable with mean zero and fixed variance $d_{i,t}^2$. The predetermined component $w_{i,t}$ is not explicitly included in approximating model (12) since it is fixed at time t. The constants $c_{i,t}$ and $d_{i,t}$ are chosen such that (12) 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 $i, j = 1, \ldots, N$ and $s, t = 1, \ldots, T$. Furthermore, $u_{i,t}$ is serially uncorrelated. It follows that

$$g(y|\mu, \xi) = \prod_{i=1}^{N} \prod_{t=1}^{T} g(y_{i,t}|\mu_i, \xi_t), \quad \text{with}$$

$$g(y_{i,t}|\mu_i, \xi_t) \equiv \text{NID}(c_{i,t} + \epsilon_{i,t}, d_{i,t}^2). \quad (13)$$

The maximization of $\log p(\mu, \xi|y; x)$ with respect to μ and ξ 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 μ and ξ based on the linearized model given by equations (12) and (3). The following algorithm summarizes this method.

Algorithm A. (i) Initialize the algorithm by choosing μ^* and ξ^* as starting values, which gives $\epsilon_{i,t}^*$ and $z_{i,t}^*$, for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$;
(ii) Given the set of two equations

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

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

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where $p(y_{i,t}|z_{i,t})$ is the observation model (1) and $g(y_{i,t}|\epsilon_{i,t})$ is given by (13), 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 new $\mu^* = \text{E}_g(\mu|y; \xi^*)$ from the resulting model (12) with $\xi = \xi^*$, $c_{i,t} = c_{i,t}^*$ and $d_{i,t} = d_{i,t}^*$;
(iv) Compute new $\xi^* = \text{E}_g(\xi|y; \mu^*)$ from the resulting model (12) with $\mu = \mu^*$, $c_{i,t} = c_{i,t}^*$ and $d_{i,t} = d_{i,t}^*$;
(v) Iterate from (ii) to (iv) 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 after convergence $\mu^* = \hat{\mu} = \text{argmax}_\mu p(\mu|y; \hat{\xi}; x)$ and $\xi^* = \hat{\xi} = \text{argmax}_\mu p(\xi|y; \hat{\mu}; x)$. Further, it holds that $\{\hat{\mu}, \hat{\xi}\} = \text{argmax}_{\mu, \xi} p(\mu, \xi|y; x)$.

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 ξ, the approximating model (12) is reduced to a standard random effects model as given by

$$\bar{y}_i = \bar{c}_i + \bar{A}_i \mu_i + \bar{\mathcal{B}}_i + \bar{u}_i, \quad \bar{u}_i \sim \text{NID}(0, \bar{D}_i), \ i = 1, \ldots, N, \quad (14)$$

where $\bar{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\bar{c}_i = (c_{i,1}, \ldots, c_{i,T})'$, $\bar{A}_i = (a_{i,1}, \ldots, a_{i,T})'$, $\bar{\mathcal{B}}_i = (b_{i,1}'\xi_1, \ldots, b_{i,T}'\xi_T)'$ and $\bar{u}_i = (u_{i,1}, \ldots, u_{i,T})'$, see Baltagi (2005, Chapters 2 and 5). The $T \times T$ variance matrix $\bar{D}_i$ is diagonal by construction, with elements $d_{i,1}^2, \ldots, d_{i,T}^2$ on the main diagonal. Based on (14), the computation of $\text{E}_g(\mu|y; \xi^*)$ can be performed using standard multivariate normal regression theory.

In step (v) we need to compute $\text{E}_g(\xi|y; \mu^*)$. Given a value of μ, the approximating model (12) can be written as a linear Gaussian state space model as given by

$$y_t = c_t + \mathcal{A}_t + B_t \xi_t + u_t, \quad u_t \sim \text{NID}(0, D_t), \ t = 1, \ldots, T, \quad (15)$$

where $y_t = (y_{1,t}, \ldots, y_{N,t})'$, $c_t = (c_{1,t}, \ldots, c_{N,t})'$, $\mathcal{A}_t = (a_{1,t}'\mu_1, \ldots, a_{N,t}'\mu_N)'$, $B_t = (b_{1,t}, \ldots, b_{N,t})'$ and $u_t = (u_{1,t}, \ldots, u_{N,t})'$. Variance matrix D_t is diagonal by construction, with elements $d_{1,t}^2, \ldots, d_{N,t}^2$ on the main diagonal. Based on (15) the computation of $\text{E}_g(\xi|y; \mu^*)$ is carried out using the Kalman filter and smoothing methods. Additional computational details are provided in the Technical Appendix.

3.2. Collapsing the approximate linear Gaussian panel data model

The evaluation of likelihood estimate $\hat{p}(y)$ in (10) requires M samples of μ and ξ from importance densities $g(\mu|y; \hat{\xi})$ and $g(\xi|y; \hat{\mu})$, respectively. Both importance densities are based on approximating models given in (14) and (15). However, both models 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, we show that samples $\xi^{(i)}$ and $\mu^{(i)}$ can be drawn from $g(\xi|\bar{y}^l; \hat{\mu})$ and $g(\mu|\bar{y}^l; \hat{\xi})$, respectively, where $\bar{y}^l = [(\bar{y}_1^l)', \ldots, (\bar{y}_N^l)']'$ and $y^l = [(y_1^l)', \ldots, (y_T^l)']'$ are low-dimensional vector series. The resulting samples can be regarded as coming from $g(\mu|y; \hat{\xi})$ and $g(\xi|y; \hat{\mu})$, respectively.

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 Eq. (15), with $\mathcal{A}_t$ replaced by $\hat{\mathcal{A}}_t = (a_{1,t}'\hat{\mu}_1, \ldots, a_{N,t}'\hat{\mu}_N)'$, into low-dimensional $r \times 1$ vectors y_t^l, without losing information relevant for the extraction of ξ. A $r \times N$ transformation matrix S_t^l that satisfies the conditions to prevent information loss is given by

$$S_t^l = \Delta_t' B_t' D_t^{-1}, \qquad \Delta_t \Delta_t' = (B_t' D_t^{-1} B_t)^{-1}, \quad (16)$$

with Δ_t being a lower triangular matrix, see Jungbacker and Koopman (2008). This choice for $y_t^l = S_t^l y_t$ gives

$$y_t^l = \Delta_t^{-1} \xi_t + u_t^l, \quad u_t^l \sim \text{NID}(0, I_r), \ t = 1, \ldots, T, \quad (17)$$

where Δ_t^{-1} is a $r \times r$ lower triangular matrix, ξ_t is defined in (4) 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 performed by applying the simulation smoother methods of Durbin and Koopman (2002) to r-dimensional vector series y_t^l and model (17), for $t = 1, \ldots, T$. The high dimensional orthogonal complement of S_t^l can be constructed from S_t^l but is not required for any of the necessary computations.

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 $i = 1, \ldots, N$, based on vector representation (14), with $\bar{\mathcal{B}}_i$ replaced by $\hat{\mathcal{B}}_i = (b_{i,1}'\hat{\xi}_1, \ldots, b_{i,T}'\hat{\xi}_T)'$, into $q \times 1$ vectors $\bar{y}_i^l$. We consider similar least squares type transformations as for the cross-section dimension above. However, because μ_i and μ_j are independent, the transformed

observations $\bar{y}_i^l$ become simple rescaled averages of the variables in $\bar{y}_i$. A suitable $T \times q$ transformation matrix $\bar{S}_i^l$ is given by

$$\bar{S}_i^l = \bar{\Delta}_i' \bar{A}_i' \bar{D}_i^{-1}, \qquad \bar{\Delta}_i \bar{\Delta}_i' = (\bar{A}_i' \bar{D}_i^{-1} \bar{A}_i)^{-1}, \tag{18}$$

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

$$\bar{y}_i^l = \bar{\Delta}_i^{-1} \mu_i + \bar{u}_i^l, \quad \bar{u}_i^l \sim \text{NID}(0, I_q), \ i = 1, \dots, N, \tag{19}$$

where $\bar{\Delta}_i^{-1}$ is a lower triangular $q \times q$ matrix, μ_i is given in (4) and $\bar{u}_i^l$ is a random vector with mean zero and $q \times q$ unit variance. 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 $\Sigma_\mu - \Sigma_\mu \bar{\Delta}_i^{-1} (\bar{\Delta}_i^{-1'} \Sigma_\mu \bar{\Delta}_i^{-1} + I_q)^{-1} \bar{\Delta}_i^{-1'} \Sigma_\mu$. In the Technical Appendix we present the computational details for the transformations and we discuss the computational gains. On average the likelihood evaluation times reduce between 2 and 10 times depending on the different random effects included in the model.

3.3. Constructing the Monte Carlo likelihood

Next we discuss the construction of the Monte Carlo likelihood estimate $\hat{p}(y)$ in Eq. (10). The estimate relies on densities $g(y; \hat{\mu})$ and $g(y; \hat{\xi})$, that are based on the approximating model (12). Density $\log g(y; \hat{\mu})$ can be computed from the prediction error decomposition of vector representation (15), with μ replaced by $\hat{\mu}$. This is obtained by a single pass through the Kalman filter, see Durbin and Koopman (2012, Chapter 7). Computational efficiency can increased by using the lower dimensional model (17), based on vector series y_t^l. In particular, Jungbacker and Koopman (2008) show that

$$\log g(y; \hat{\mu}) = \text{constant} + \log g(y^l; \hat{\mu})$$
$$- \frac{1}{2} \sum_{t=1}^{T} \log |D_t| + e_t' D_t^{-1} e_t, \tag{20}$$

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 (17), which is a $r \times T$-dimensional problem.

The logdensity $\log g(y; \hat{\xi})$ is also evaluated using the collapsed vector series $\bar{y}_i^l$, for $i = 1, \dots, N$. Based on model (19) we obtain

$$\log g(y; \hat{\xi}) = \text{constant} + \log g(\bar{y}^l; \hat{\xi})$$
$$- \frac{1}{2} \sum_{i=1}^{N} \log |\bar{D}_i| + \bar{e}_i' \bar{D}_i^{-1} \bar{e}_i, \tag{21}$$

where $\bar{y}^l = (\bar{y}_1^{l'}, \dots, \bar{y}_N^{l'})'$ and $\bar{e}_i = \bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i - \bar{A}_i(\bar{A}_i' \bar{D}_i^{-1} \bar{A}_i)^{-1} \bar{A}_i' \bar{D}_i^{-1}(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i)$.

The following algorithm summarizes the evaluation of the loglikelihood for balanced panels. Given parameter vector ψ 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_t^l using Section 3.2;
(iii) Sample M draws $\mu^{(i)}$ and $\xi^{(i)}$ from densities $g(\xi \mid y^l; \hat{\mu})$ and $g(\mu \mid \bar{y}^l; \hat{\xi})$, which are based on transformed models (17) and (19), and compute importance weights $w^{(i)}$, as given in equation (11);

(iv) Evaluate logdensities $\log g(y; \hat{\mu})$ as in (20) and $\log g(y; \hat{\xi})$ as in (21);
(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 ψ using an arbitrary numerical optimization method. As a practical choice we use the BFGS algorithm, see Nocedal and Wright (1999). To retain the smoothness of the likelihood in ψ 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 $\tilde{\psi}$. Durbin and Koopman (1997) advocate the use of antithetic variables to improve the efficiency of the importance sampling weights. An antithetic variable in our context is constructed for each random draw $\mu^{(i)}$ or $\xi^{(i)}$ from the importance densities such that it is equiprobable with μ or ξ, respectively, and it leads to smaller Monte Carlo variation. For each draw of $\mu^{(i)}$ and $\xi^{(i)}$ we manufacture antithetic variables that balance for location and for scale.

3.4. 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 μ_i, ξ_t and possibly observed elements of $\mathcal{Y}_{t-1}$. Calculations in step (iii), conditional on ξ^*, 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). Step (v) of Algorithm A is calculated by Kalman filter and smoothing methods, which can account for missing values, see Durbin and Koopman (2012, Section 4.10).

The transformations for panel reduction from Section 3.2 need to be adjusted for missing values as well. Jungbacker et al. (2011), show that by choosing an alternative state space representation for model (15) 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 μ_i and μ_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 (17) 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 $g(y; \hat{\xi})$, Eq. (21) can be computed based on lower dimensional model (19) 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 (11) are based on elements of $p(y | \mu^{(i)}, \xi^{(i)}; x)$ for which $y_{i,t}$ and $x_{i,t}$ are actually observed.

4. Monte Carlo study

We present and discuss our results from an extended Monte Carlo study for the generalized dynamic panel data model with the Binary, Binomial and Student's t observation densities. These three models are also considered in our empirical illustrations discussed in Section 5. The purpose of the Monte Carlo study is to evaluate the small sample properties of the estimation procedure presented in Section 3. In particular, we examine whether a $\sqrt{M}$ convergence rate is likely to exist for our simulated likelihood estimate from Algorithm B. Additionally, we study the magnitude of the Monte Carlo variance of the simulated likelihood estimate. Finally, the computational feasibility and accuracy of the estimation methodology is studied.

Table 1

Monte Carlo design with our signal specifications, parameter values and panel dimensions for simulating the observations. The data generation process is further given by $x_{i,t} \sim \text{NID}(0,1)$, $\mu_i \sim \text{NID}(0,\sigma_\mu^2)$, $\xi_t = \alpha_t$, $\alpha_{t+1} = h\alpha_t + \eta_t$ and $\eta_t \sim \text{NID}(0,\sigma_\eta^2)$. The initial time varying effect is taken as $N(0, \sigma_\eta^2/(1-h^2))$. For the Student's t density it holds that $a(\nu) = \Gamma(\nu/2 + 1/2)/\Gamma(\nu/2)$ and $\lambda^{-1} = (\nu-2)\sigma^2$.

Observation density		
(A)	Binary	$\log p(y_{i,t}|z_{i,t}) = y_{i,t}z_{i,t} - \log(1 + \exp z_{i,t})$
(B)	Binomial	$\log p(y_{i,t}|z_{i,t}) = y_{i,t}z_{i,t} - n_{i,t}(1 + \exp z_{i,t}) - \log\binom{n_{i,t}}{y_{i,t}}$
(C)	Student's t	$\log p(y_{i,t}|z_{i,t}) = \log a(\nu) + \frac{1}{2}\log\lambda - \frac{\nu+1}{2}\log(1 + \lambda(y_{i,t} - z_{i,t})^2)$

Signal			Parameters					
			γ	β	σ_μ	h	σ_η	ν
(1)	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i$	(a)	0.2	1	0.5	–	–	(3, 5, 10)
		(b)	0.2	1	1	–	–	(3, 5, 10)
		(c)	0.2	1	3	–	–	(3, 5, 10)
(2)	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \xi_t$	(a)	0.2	1	–	0.3	0.2	(3, 5, 10)
		(b)	0.2	1	–	0.9	0.2	(3, 5, 10)
(3)	$z_{i,t} = y_{i,t-1}\gamma + x'_{i,t}\beta + \mu_i + \xi_t$	(a)	0.2	1	0.5	0.3	0.2	(3, 5, 10)
		(b)	0.2	1	0.5	0.9	0.2	(3, 5, 10)
		(c)	0.2	1	1	0.3	0.2	(3, 5, 10)
		(d)	0.2	1	1	0.9	0.2	(3, 5, 10)
		(e)	0.2	1	3	0.3	0.2	(3, 5, 10)
		(f)	0.2	1	3	0.9	0.2	(3, 5, 10)

4.1. Monte Carlo design

Our aim is to assess the performance of our simulation-based estimation procedure for a set of different densities, signals, parameter values, panel sizes and numbers of missing values. Table 1 presents the combinations of densities, signals and parameter values, that we investigate in our study. Three different observation densities are considered for (1) and correspond to (A) Binary, (B) Binomial or (C) Student's t distributions. The signals correspond to models with 1. individual-specific effects or 2. time-varying effects or 3 both 1 and 2.

The signal is generated as explained in Eqs. (2)–(4) of Section 2. The fixed component $w_{i,t}$ is based on a single covariate $x_{i,t}$, that is drawn independently from the $N(0,1)$ distribution, with $\beta = 1$ and on the lagged dependent variable with $p_y = 1$ and $\gamma_1 = \gamma = 0.2$. The stochastic component is based on the univariate, $q = 1$, individual-specific effect μ_i and the univariate, $r = 1$, time-varying effect ξ_t with weights $a_{i,t} = b_{i,t} = 1$. The individual-specific effect μ_i is normally distributed with common mean $\delta = 0$ and variance $\Sigma_\mu = \sigma_\mu^2$, that is $\mu_i \sim N(0,\sigma_\mu^2)$. We investigate the results for different values of standard deviation $\sigma_\mu = 0.5, 1, 3$. The time-varying effect is updated by an autoregressive process α_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 $\sigma_\eta = 0.2$.

For the Student's t density we have additional parameters ν and σ, where ν is the degrees of freedom and σ is the scaling. We fix the value of σ at one and estimate degrees of freedom ν along with the other parameters. We consider $\nu = 3$, $\nu = 5$ and $\nu = 10$. The entire parameter vector is given by $\psi = \{\gamma, \beta, \sigma_\mu, h, \sigma_\eta, \nu\}$.

We consider five different combinations of the panel dimensions $N, T = 10, 50, 100, 250$. In the case of $N = T = 10$, the dimensions are sufficiently small that we can also consider the standard implementation of Shephard and Pitt (1997) and Durbin and Koopman (1997), see Section 4.3. In all other cases, it is computationally only feasible to implement our newly proposed method presented in Section 3.

4.2. Diagnostic tests for the importance sampler

A sufficient condition to guarantee a $\sqrt{M}$ convergence rate for the Monte Carlo estimate $\hat{p}(y)$ in (10) is the existence of the variance in the importance samplings weights $w^{(i)}$, as given in Eq. (11), for $i = 1, \ldots, M$, see Geweke (1989). To test whether the variance is finite Koopman et al. (2009) propose test statistics for evaluating variance in a sequence of importance weights. Implementation of their suggested Wald type test statistic is done by the following steps: (1) Simulate a panel y using a combination of observation density, signal, parameter values and panel dimensions as given in Table 1. (2) Estimate the parameters using the Monte Carlo maximum likelihood methods of Section 3 and replace the parameter vector ψ by the resulting estimate $\tilde{\psi}$. (3) Generate 100,000 importance sampling weights $w^{(i)}$ using the importance densities $g(\xi|y^l; \hat{\mu})$ and $g(\mu|\bar{y}^l; \hat{\xi})$. (4) Consider s exceedance sampling weights, denoted by $x_1, \ldots, x_s$, which are larger than some threshold $w^{\min}$ and 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 $i = 1, \ldots, s$, where unknown parameters a and b determine the shape and scale of the density, respectively. When $a \leq 0.5$, the variance of the importance sampling weights exists and a $\sqrt{M}$ convergence rate can be assumed. (5) Estimate a and b by maximum likelihood and denote the estimates by $\tilde{a}$ and $\tilde{b}$, respectively. (6) Compute the t-test statistic $t_w = \tilde{b}^{-1}\sqrt{s/3}(\tilde{a} - 0.5)$ to test the null hypothesis $H_0: a = 0.5$. 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.

Figs. 1(a) and (b) present the diagnostic test statistics for the importance samplers applied to three different densities (A, B and C), and four signals (1.b, 2.b, 3.b and 3.e) as listed in Table 1. Furthermore, the test statistics are computed for different values of the threshold $w^{\min}$. We choose $w^{\min}$ values such that a range of the largest 1%–50% of the weights are included. The test statistics are computed with the use of antithetic variables in all cases.

The test statistics in Fig. 1(a) for the Binary and Binomial densities are negative for all combinations of signals, parameter values and panel dimensions. They provide strong evidence for the existence of a variance for the importance density function. The test statistics in Fig. 1(b) for the Student's t density present a similar picture. However, when the degrees of freedom is small, $\nu = 3$, and the panel dimensions are large $N = 250$ and $T = 250$, the test is rejected for threshold values that lead to exceedance samples that include more than 35% of all weights. When the degrees of freedom becomes even smaller similar results are obtained. We
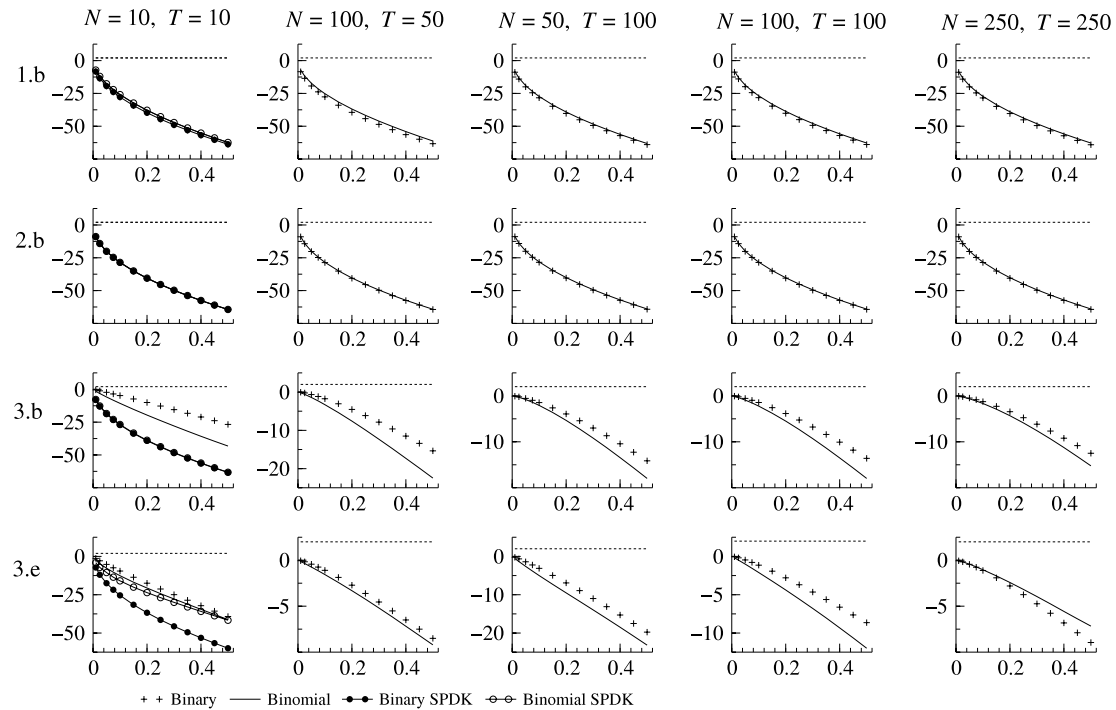
(a) Binary and Binomial.
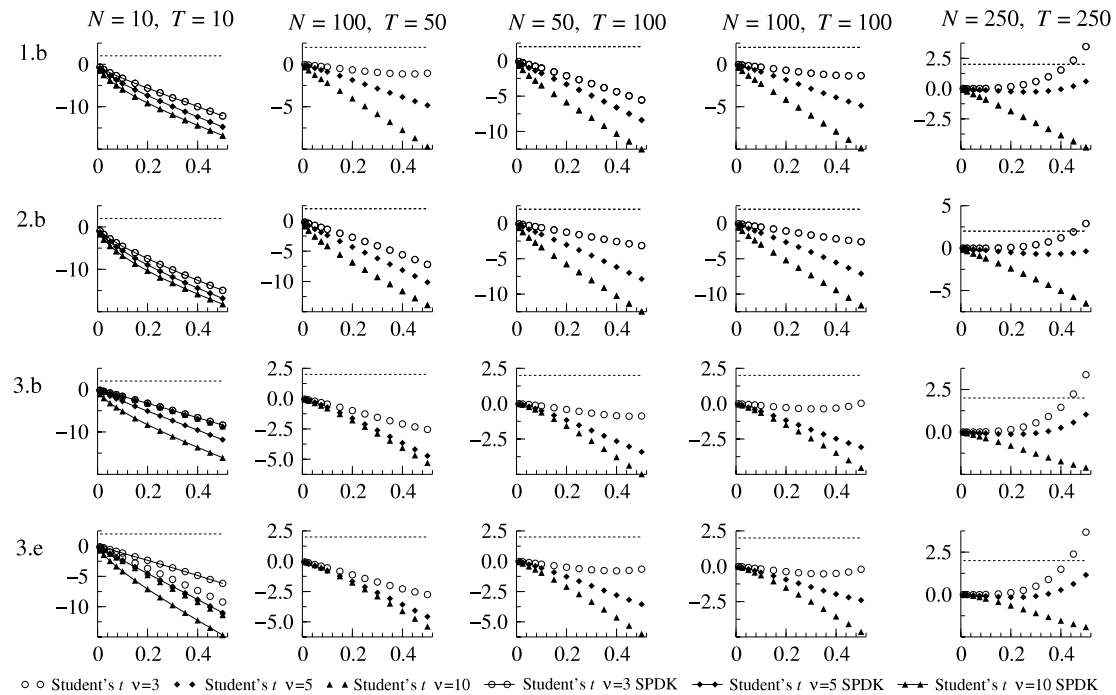
(b) Student's t.

Fig. 1. Importance sampling diagnostics for dynamic panel data models, based on 100 000 simulations of weights $w^{(i)}$ defined in Eq. (11). The test statistics are presented for densities A Binary and B Binomial, and 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 4.2. Thresholds are based on the number of exceedance values $x_1, \ldots, x_s$ included. We have taken $0.01 = s/100\,000$, $0.025 = s/100\,000$, $0.05 = s/100\,000$, until $0.5 = s/100\,000$. The area above the dotted straight line at 2 indicates the rejection area.

notice that only for large exceedance samples, we move away from the left-tail of the distribution of the test statistic.

We may conclude that there is substantial evidence that a $\sqrt{M}$ convergence rate exists for our proposed likelihood estimate for the generalized dynamic panel data model. However we do recommend to check for each application whether the importance weights are finite. The presented results also hold for the other signals listed in Table 1. They are further discussed in the Technical Appendix.

4.3. Measuring the efficiency loss of Algorithm A

We discussed in Section 3 that our proposed implementation of importance sampling for the generalized dynamic panel data model is different from the standard implementation of Shephard

Table 2
Simulation results for the non-Gaussian 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. Specifications 3.b and 3.e from Table 1 together with observation models A, B and C are used for simulation. The third panel corresponds to the Student's t density with $\nu = 3$ degrees of freedom whereas the fourth panel corresponds to the Student's t density with $\nu = 5$ degrees of freedom. All parameters are estimated by procedures outlined in Section 3, with $M = 1000$ draws from importance densities $g(\xi|y^l; \hat{\mu})$ and $g(\mu|\bar{y}^l; \hat{\xi})$, respectively.

	N	T	γ	β	σ_μ	h	σ_η	ν
A.3.b	100	50	$-0.004_{0.072}$	$0.001_{0.036}$	$-0.002_{0.057}$	$-0.057_{0.127}$	$0.001_{0.042}$	–
	50	100	$-0.006_{0.065}$	$-0.004_{0.041}$	$-0.012_{0.058}$	$-0.032_{0.085}$	$-0.001_{0.038}$	–
	100	100	$0.004_{0.045}$	$0.004_{0.032}$	$-0.004_{0.043}$	$-0.034_{0.081}$	$0.002_{0.037}$	–
	250	250	$-0.001_{0.019}$	$-0.000_{0.010}$	$-0.004_{0.024}$	$-0.007_{0.035}$	$-0.001_{0.016}$	–
A.3.e	100	50	$0.004_{0.100}$	$0.004_{0.046}$	$-0.028_{0.271}$	$-0.027_{0.379}$	$-0.033_{0.096}$	–
	50	100	$-0.019_{0.100}$	$-0.001_{0.052}$	$-0.055_{0.303}$	$-0.059_{0.357}$	$-0.015_{0.073}$	–
	100	100	$-0.003_{0.061}$	$0.003_{0.035}$	$-0.021_{0.234}$	$-0.040_{0.325}$	$-0.019_{0.067}$	–
	250	250	$0.002_{0.029}$	$0.001_{0.013}$	$-0.020_{0.138}$	$-0.019_{0.112}$	$-0.000_{0.018}$	–
B.3.b	100	50	$0.003_{0.012}$	$-0.001_{0.023}$	$-0.011_{0.037}$	$-0.042_{0.092}$	$-0.000_{0.030}$	–
	50	100	$0.002_{0.013}$	$0.002_{0.021}$	$-0.009_{0.047}$	$-0.019_{0.056}$	$0.002_{0.027}$	–
	100	100	$0.000_{0.009}$	$0.002_{0.014}$	$0.000_{0.037}$	$-0.026_{0.064}$	$-0.001_{0.021}$	–
	250	250	$0.000_{0.003}$	$0.001_{0.005}$	$-0.001_{0.024}$	$-0.006_{0.031}$	$-0.001_{0.010}$	–
B.3.e	100	50	$0.002_{0.018}$	$0.005_{0.028}$	$-0.013_{0.224}$	$-0.034_{0.207}$	$-0.005_{0.037}$	–
	50	100	$0.005_{0.020}$	$-0.001_{0.027}$	$-0.021_{0.300}$	$-0.045_{0.200}$	$-0.005_{0.037}$	–
	100	100	$0.002_{0.013}$	$0.000_{0.019}$	$0.002_{0.225}$	$-0.035_{0.160}$	$-0.005_{0.025}$	–
	250	250	$0.000_{0.004}$	$0.000_{0.007}$	$-0.008_{0.142}$	$0.000_{0.073}$	$-0.001_{0.010}$	–
C.3.b	100	50	$-0.001_{0.008}$	$-0.002_{0.009}$	$-0.011_{0.030}$	$-0.018_{0.040}$	$0.002_{0.022}$	$0.010_{0.051}$
	50	100	$-0.001_{0.009}$	$-0.003_{0.009}$	$-0.012_{0.040}$	$-0.017_{0.054}$	$-0.003_{0.023}$	$0.008_{0.050}$
	100	100	$0.000_{0.005}$	$-0.000_{0.007}$	$-0.000_{0.025}$	$-0.012_{0.035}$	$0.002_{0.015}$	$-0.002_{0.038}$
	250	250	$0.000_{0.002}$	$0.000_{0.003}$	$0.000_{0.011}$	$-0.000_{0.025}$	$0.000_{0.010}$	$0.002_{0.015}$
C.3.e	100	50	$-0.001_{0.008}$	$-0.002_{0.009}$	$-0.037_{0.168}$	$-0.029_{0.124}$	$-0.000_{0.020}$	$0.010_{0.051}$
	50	100	$-0.001_{0.008}$	$-0.003_{0.009}$	$-0.049_{0.226}$	$-0.001_{0.107}$	$-0.004_{0.020}$	$0.010_{0.050}$
	100	100	$0.000_{0.005}$	$-0.000_{0.007}$	$0.015_{0.261}$	$-0.012_{0.097}$	$0.000_{0.021}$	$-0.002_{0.039}$
	250	250	$0.000_{0.002}$	$0.001_{0.003}$	$-0.009_{0.057}$	$-0.006_{0.041}$	$0.000_{0.007}$	$0.001_{0.012}$
C.3.b	100	50	$-0.001_{0.009}$	$0.005_{0.013}$	$-0.014_{0.090}$	$-0.067_{0.090}$	$-0.001_{0.027}$	$0.053_{0.244}$
	50	100	$-0.001_{0.008}$	$0.005_{0.014}$	$-0.009_{0.111}$	$-0.038_{0.072}$	$-0.001_{0.018}$	$0.050_{0.222}$
	100	100	$0.001_{0.005}$	$-0.002_{0.010}$	$-0.034_{0.083}$	$-0.017_{0.055}$	$-0.003_{0.018}$	$0.043_{0.154}$
	250	250	$0.000_{0.002}$	$0.001_{0.002}$	$-0.017_{0.037}$	$-0.020_{0.030}$	$0.003_{0.009}$	$0.029_{0.040}$
C.3.e	100	50	$-0.002_{0.008}$	$0.004_{0.013}$	$-0.032_{0.266}$	$-0.081_{0.177}$	$-0.003_{0.022}$	$0.052_{0.239}$
	50	100	$-0.000_{0.007}$	$0.005_{0.014}$	$0.018_{0.341}$	$-0.024_{0.130}$	$-0.005_{0.016}$	$0.055_{0.222}$
	100	100	$0.000_{0.005}$	$-0.002_{0.010}$	$-0.089_{0.247}$	$-0.008_{0.110}$	$-0.004_{0.018}$	$0.046_{0.155}$
	250	250	$0.000_{0.003}$	$0.001_{0.002}$	$-0.037_{0.117}$	$-0.019_{0.053}$	$0.002_{0.009}$	$0.028_{0.042}$

and Pitt (1997) and Durbin and Koopman (1997), hereafter SPDK. The standard SPDK method would sample $\mu^{(i)}$ and $\xi^{(i)}$ from $g(\mu, \xi|y)$, where $g(\mu, \xi|y)$ is based on the linear Gaussian model (12). The corresponding likelihood is then estimated by

$$\hat{p}(y) = g(y)M^{-1} \sum_{i=1}^{M} w^{(i)}, \quad w^{(i)} = \frac{p(y|\mu^{(i)}, \xi^{(i)}; x)}{g(y|\mu^{(i)}, \xi^{(i)})}, \qquad (22)$$

where $\mu^{(i)}$ and $\xi^{(i)}$ are drawn from $g(\mu, \xi|y)$. This implementation is not feasible for even moderately large panels, see the discussion in Section 3.

It is anticipated that the variance of the importance weight function increases when the SPDK method in (22) is replaced by our method in (10) because Algorithm A does not account for the dependence between μ and ξ in $g(\mu, \xi|y)$. To study the extent of the increase in the variance, we compare the t-test statistics for the importance weights in (11) with importance weights in (22). We notice that sampling from the SPDK importance density is only computationally feasible for small panel sizes and that the collapsing method of Section 3.2 is not applicable to $g(\mu, \xi|y)$.

For the small panel $N = 10$ with $T = 10$, we have implemented the standard SPDK method for our models to facilitate its comparison with our proposed method. Hence Figs. 1(a) and (b) also present the diagnostic test statistics of Section 4.2 for the SPDK method and for the smallest panel only. The differences between the test statistics for SPDK and Algorithm A are small. For all three models, the test statistics for the SPDK importance weights are slightly lower, but both are very negative. We notice that SPDK is clearly not feasible for even modest panel dimensions.

In the Technical Appendix we also provide estimates for the loss of efficiency in the likelihood that results from the increase in variance. We document that the loss can be substantial for the Binary and Binomial densities, but low for the Student's t density. However, the increase of variance in the likelihood estimate can be offset by increasing the number of draws M. Any degree of accuracy can be achieved and given the large computational improvements that are also documented in the Technical Appendix, this is a minor difficulty. In the remainder of this paper we always take $M = 1000$ draws from the importance density of Algorithm A.

4.4. Finite sample properties of Algorithm B

We investigate the accuracy of the likelihood estimates by simulating 100 data panels for each possible design given in Table 1. For each simulated data panel, we estimate the parameter vector ψ using Algorithm B with collapsed vectors. After collecting all estimated parameter vectors, we report the average bias and standard deviation in Table 2 for signals 3.b and 3.e. The results of the 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. We notice that individual state dependence as captured by $\gamma(B)y_{i,t}$ can be empirically identified and separated from stochastic components μ_i and ξ_t.

For each simulated data panel we also create unbalanced panels by removing 40% of the observations, at the beginning and end of the dataset. The parameter estimates remained unbiased and the standard errors increase slightly due to the loss of data. The full set of parameter estimation results, with and without missing values, are presented in our Technical Appendix. For all computations in this study, we have written the code in the Ox programming language version 6.10 of Doornik (2007).

5. Empirical illustrations

5.1. Union participation of young males

We first analyze the union membership decision of young males, see Vella and Verbeek (1998), Wooldridge (2005) and Liesenfeld and Richard (2008). We use the data from Vella and Verbeek (1998) to estimate a logistic model for union membership. The binary outcome of the union membership decision for individual i in year t is denoted by $y_{i,t} \in \{0, 1\}$. The data consists of $N = 545$ males and their possible membership is yearly observed between 1981 and 1987, $T = 7$. The initial time period $t = 0$ corresponds to 1980 and the panel is balanced.

The model for the union membership decision is given by density (A) and signal 3 in Table 1. For this signal the coefficient γ captures the effect whether the individual is a current union member or not. The explanatory variable vector $x_{i,t}$ includes variables capturing, schooling, experience, marriage, ethnicity, marital status and location effects, see Table 1 in Vella and Verbeek (1998) for more detailed descriptions of these variables. We depart from the study of Vella and Verbeek (1998) by not including the industry and occupational dummy variables in $x_{i,t}$ and by replacing their time dummies with our stochastically time-varying effect ξ_t.

We conduct three analysis. First, we aim to assess whether the persistence is a result from either individual satisfaction of last year's decision as captured by $\gamma y_{i,t-1}$, referred to as true state dependence, or from underlying aspects that cause the observed persistence due to individual-specific μ_i and time-varying ξ_t effects, referred to as spurious state dependence. Second, we recognize that the initial observations and the time-varying explanatory variables tend to be correlated with the individual-specific effect μ_i. For example, the marriage decision partly explains the union choice, but may also influence other unobserved characteristics that determine μ_i. To address this issue we allow the individual effect to be correlated with the initial observations $y_{i,0}$ and the time-varying explanatory variables $x_{i,t}$. In particular, we implement the method of Chamberlain (1980) which has been extended for nonlinear dynamic panel data models by Wooldridge (2005) and specify μ_i as

$$\mu_i = y_{i,0}\lambda_0 + \sum_{t=1}^{T} x'_{i,t}\lambda_t + v_i \quad v_i \sim \text{NID}(\delta, \sigma_\mu^2), \qquad (23)$$

where $\lambda_0, \lambda_1, \ldots, \lambda_T$, with $T = 7$, are fixed parameters that measure the correlation between μ_i and the predetermined variables, and where v_i represents the remainder individual-specific effect with fixed mean δ and variance σ_μ^2. Third and finally, we consider a different strategy for capturing spurious state dependence that is discussed in Heiss (2008). Here individual-specific time-varying stochastic trends are considered to capture the serial correlation in $\epsilon_{i,t}$. It makes the use of time-invariant individual-specific effects μ_i redundant. In our setting we can implement it as

$$\epsilon_{i,t} = b'_i\xi_t + y_{i,0}\lambda_0 + \sum_{t=1}^{T} x'_{i,t}\lambda_t, \qquad (24)$$

where ξ_t is the $(N + 1) \times 1$ vector of time varying effects and b_i is the $(N + 1) \times 1$ selection vector with the ith and the $(N + 1)$th elements set equal to one and zero otherwise, with $N = 545$. The first N time-varying effects are specific for each time series while the $(N + 1)$th element is common for all time series. The estimation method suggested in Heiss (2008) does not allow for the common time-varying effect. Following Heiss (2008), we model the first N elements by $\xi_{i,t+1} = g\xi_{i,t} + \eta_{i,t}$, with $|g| \leq 1 \eta_{i,t} \sim \text{NID}(\delta(1 - g), \sigma_\lambda^2(1 - g^2))$. The $(N + 1)$th element is modeled by

Variable	Model 1	Model 2	Model 3
$y_{i,t-1}$	2.344 $_{0.167}$	1.770 $_{0.153}$	1.782 $_{0.157}$
log experience	−0.201 $_{0.236}$	−0.250 $_{0.200}$	−0.252 $_{0.480}$
Schooling	−0.020 $_{0.054}$	−0.030 $_{0.058}$	−0.028 $_{0.065}$
Married	0.382 $_{0.148}$	0.291 $_{0.196}$	0.294 $_{0.198}$
Black	1.124 $_{0.268}$	0.990 $_{0.286}$	1.008 $_{0.290}$
Hispanic	0.518 $_{0.257}$	0.326 $_{0.278}$	0.336 $_{0.281}$
Rural	0.023 $_{0.196}$	0.021 $_{0.211}$	0.019 $_{0.212}$
Health	−0.865 $_{0.528}$	−0.730 $_{0.541}$	−0.724 $_{0.547}$
North-East	0.364 $_{0.273}$	0.259 $_{0.295}$	0.263 $_{0.300}$
South	0.023 $_{0.248}$	−0.018 $_{0.269}$	−0.004 $_{0.282}$
North Central	0.490 $_{0.263}$	0.411 $_{0.285}$	0.424 $_{0.213}$
$y_{i,0}$	–	2.137 $_{0.242}$	2.166 $_{0.231}$
$\text{Married}_{i,1}$	–	0.115 $_{0.326}$	0.113 $_{0.331}$
$\text{Married}_{i,2}$	–	−0.145 $_{0.387}$	−0.135 $_{0.317}$
$\text{Married}_{i,3}$	–	−0.104 $_{0.390}$	−0.102 $_{0.399}$
$\text{Married}_{i,4}$	–	0.045 $_{0.422}$	0.033 $_{0.415}$
$\text{Married}_{i,5}$	–	0.579 $_{0.397}$	0.591 $_{0.422}$
$\text{Married}_{i,6}$	–	0.248 $_{0.397}$	0.247 $_{0.400}$
$\text{Married}_{i,7}$	–	−0.653 $_{0.311}$	−0.653 $_{0.316}$
h	−0.418 $_{0.566}$	−0.271 $_{0.606}$	−0.514 $_{0.610}$
σ_η	0.126 $_{0.069}$	0.003 $_{0.059}$	0.117 $_{0.064}$
g	–	–	0.989 $_{0.012}$
$\sigma_{N,\eta}$	–	–	0.040 $_{0.003}$
δ	−2.440 $_{0.923}$	−2.614 $_{0.910}$	−2.681 $_{1.603}$
σ_μ	1.373 $_{0.131}$	1.486 $_{0.113}$	1.514 $_{0.182}$
$\log \hat{p}(y)$	−4863.1	−4808.8	−4798.9
AIC	9756.3	9663.6	9645.9

$\xi_{N+1,t+1} = h\xi_{N+1,t} + \eta_{N+1,t}$, with $|h| < 1$ and $\eta_{N+1,t} \sim \text{NID}(0, \sigma_\eta^2)$, which is similar as in the original model. When $g = 1$ it holds that $\xi_{i,1} \sim N(\delta, \sigma_\lambda^2)$ and $\xi_{i,t} = \xi_{i,t-1}$ for $t > 1$, and the model collapses to the original model but with $\xi_{i,1}$ replacing v_i in (23). We can now formally establish whether time-invariant individual-specific effects are appropriate.

The results for the three model specification are presented in Table 3. The parameter estimates are similar compared to the estimates reported in Vella and Verbeek (1998) and Liesenfeld and Richard (2008). The state dependence parameter γ is strongly significant. The estimated effects of marriage and being Black/Hispanic are positive and significant while those for other explanatory variables are not significant. The distribution of the individual-specific effects indicates that large differences exist between individuals.

When we allow for correlated random effects via (23) we find that the variable for marriage in 1987 is significantly correlated with the individual-specific effect. The state dependence parameter is estimated smaller for the extended model. However a large part of the dependence is attributed to the correlation between the individual-specific effect and the initial conditions $y_{i,0}$ as measured by λ_0. Similar changes for the state dependence and marriage parameters are documented in Wooldridge (2005).

For the third model that allows for time-varying individual-specific effects (24) we find that the parameter g is estimated very close to 1 while the estimate of $\sigma_{N,\eta}$ is relatively small. This may partly be due to the relatively small time series dimension $T = 7$. For this application we may conclude that including a time-invariant individual-specific effect μ_i is sufficient. For a larger T we may find more evidence of time-varying effects. The other parameter values have similar estimates as for Models 1 and 2.

5.2. Crime rates between 1930 and 2005

The general credit (or default) risk modeling frameworks of Duffie et al. (2007) and Koopman and Lucas (2008) can be applied
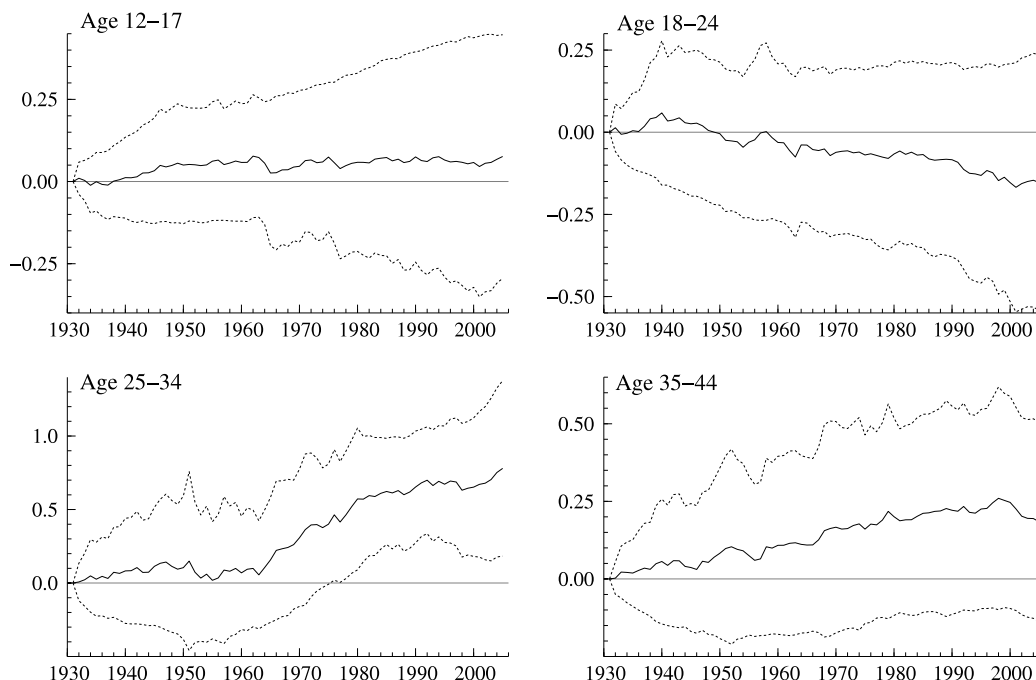
Fig. 2. Time-varying effects of the age coefficients for the age crime curve.

to other areas of empirical research where the prediction of unattractive events is of key interest. In this empirical illustration we consider modeling criminal behavior, see Koopman et al. (2008) for an earlier example. We analyze crime panel data from the so-called TRANS-5 study, see Bijleveld and Wijkman (2009) for a detailed description. The data panel consists of observations for $N = 188$ high-risk families from The Netherlands that reaches over five generations between the years 1930 and 2005, $T = 75$. For all individuals from these families, 6039 in total, their committed crimes are registered as well as age and gender information. The size of the families fluctuates over time.

In this illustration, we aggregate the data at the family level and let $y_{i,t}$ denote the number of crimes committed by family i in year t. The observation $y_{i,t}$ is modeled by the Binomial density as given by density (B) in Table 1, with $n_{i,t}$ being the number of family members, and signal 3. The family-specific effect $\mu_i = y_{i,0}\lambda_0 + v_i$, with $v_i \sim \text{NID}(\delta, \sigma_\mu^2)$, and the common time-varying effect ξ_t follows a random walk ($h = 1$). The common time-varying effect possibly accounts for the general climate in the justice and police systems. The signal $z_{i,t}$ includes explanatory variables that represent the age and gender composition of the family. The current model extends the non-Gaussian state space framework of Koopman and Lucas (2008) and Koopman et al. (2008) by including the individual-specific means μ_i. This is an important extension as typically large differences in propensity exist between families.

The parameter estimates are presented in Table 4. The state dependence coefficient γ is significantly estimated and has a positive value, indicating strong positive correlation between past and future criminal behavior within families. The proportion of males in the family has a strongly significant and positive effect on the crime intensity in the family. The estimated coefficients for the age variables, which are constructed as the fraction of individual family members within each age group, reach their peak in the late teenage years after which they slowly decline.

We extend our study by investigating whether the age effects are changing over time. Since crime reducing factors such as marriage and employment have taken place much earlier in life in the earlier years of our sample, we may expect shifts in the

Table 4
Estimation results (and standard errors in lower case) for the crime application. The panel consists of the number of crimes committed by $N = 188$ families between 1930 and 2005, $T = 75$. Model 1 is the basic model specification. Model 2 extends Model 1 by also considering time-varying effects for the age coefficients.

Variable	Model 1	Model 2
$y_{i,t-1}$	0.195 $_{0.021}$	0.193 $_{0.027}$
% males	1.468 $_{0.244}$	0.853 $_{0.051}$
Age 12–17	1.770 $_{0.233}$	1.370 $_{0.037}$
Age 18–24	2.214 $_{0.189}$	2.194 $_{0.041}$
Age 25–34	1.351 $_{0.213}$	0.619 $_{0.031}$
Age 35–44	0.564 $_{0.253}$	0.451 $_{0.034}$
$y_{i,0}$	0.068 $_{0.044}$	0.052 $_{0.029}$
σ_η	0.058 $_{0.020}$	0.019 $_{0.028}$
δ	−6.885 $_{0.093}$	−6.734 $_{0.041}$
σ_v	0.960 $_{0.084}$	0.915 $_{0.026}$
$\log \hat{p}(y)$	−28 033	−24 856
AIC	56 084	49 738

age effects. We can investigate this feature by considering time-varying effects for the age regression coefficients. We replace the common time-varying effect by

$$b'_{i,t}\xi_t, \quad \text{where } b_{i,t} = (1, \text{Age 12–17}_{i,t},$$
$$\text{Age 18–24}_{i,t}, \text{Age 25–34}_{i,t}, \text{Age 35–44}_{i,t}),$$

where ξ_t is here a 5×1 vector that captures the common time-varying factor and the time-varying coefficients for age. Each element of ξ_t is modeled by a random walk with different standard deviations. The last four initial elements in ξ_1 are fixed at zero since we also include the age variables in $x_{i,t}$. It allows us to separate the mean age effects from the time-varying effects, see Hsiao and Pesaran (2008). The estimates for the age effects are shown in Fig. 2. We find that the variable for the age group 25–34 increases significantly from say 1980 onwards. This corresponds to an age crime curve that is decaying more slowly after its peak. Hence we may conclude that during the more recent years, the crime reducing factors (marriage and employment) are considered later in life when compared to say the 1930–1950 period. Some evidence of this effect may also be observed for the age group 35–44 but the increase is more steadily and strictly not significant.
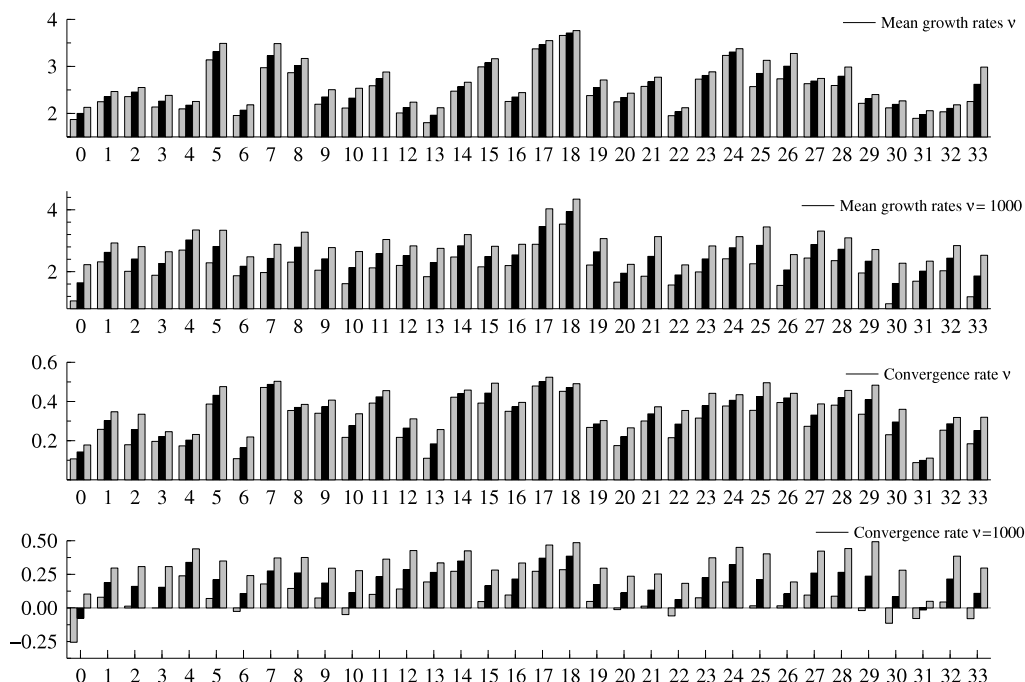
Fig. 3. Country-specific effects for the economic growth model. The black bars indicate the estimated posterior means, whereas the lighter bars provide the confidence intervals.

The developed methodology in this paper has provided the means to extract such interesting features from the data.

5.3. Economic growth

Many studies in the econometric growth literature have the aim to identify determinants of long run economic growth and to develop various measures and tests of growth convergence, see Durlauf et al. (2005) for an extensive review of the literature. We revisit the stochastic Solow growth model that is derived in Binder and Pesaran (1999). The stochastic Solow growth model differs from its deterministic counterpart by explicitly allowing for random technological progress and random labor input in the otherwise deterministic model. The model has been empirically examined in, among others, Lee et al. (1997) and Pesaran (2007). The parameters of the log linearized solution for the stochastic Solow growth model can be estimated by a variety of panel data estimators discussed in Pesaran and Smith (1995) and Pesaran (2006). The method proposed in Section 3 can also be adopted for this purpose. More standard empirical estimators, such as fixed effects, difference and system GMM estimators impose common slope parameters and can therefore not be considered for this model. They have been adopted for estimating variants of the deterministic Solow growth model with common slopes, see for example Islam (1995) and Caselli et al. (1996).

We present an empirical study for a panel of economic growth rates for $N = 34$ OECD countries for the years 1950–2010 that are listed in the Penn World Tables (PWT) version 7.1, see Heston et al. (2012) and Appendix. The resulting panel of time series is highly unbalanced with many missing entries. We generalize the model considered by Lee et al. (1997) but replace the Gaussian density with the Student's t density for the measurement errors. We remain to have heterogeneous stochastic growth trends and heterogeneous slope parameters in the model. The Student's t density is included to capture large outliers. Further, it is likely that different countries have similar access to available technology. To capture these common paths, we model the stochastic trends by a multi-factor structure, see also Pesaran (2007) and Phillips and Sul (2009).

Let $y_{i,t} = 100(\log Y_{i,t} - \log Y_{i,t-1})$, where $Y_{i,t}$ is the per capita output of country i in year t. The model specification of Pesaran (2007) for the growth rates $y_{i,t}$ is given by

$$y_{i,t} = (1 - \gamma_i)g_i + \gamma_i y_{i,t-1} + (1 - \alpha)\Delta\zeta_{i,t}$$
$$+ (\alpha - \gamma_i)\Delta\zeta_{i,t-1} - \alpha\Delta^2 v_{i,t}, \tag{25}$$

where Δ is the difference operator, g_i is the mean growth rate, γ_i is the within-country convergence parameter, α is the elasticity of output with respect to capital and $\zeta_{i,t}$ and $v_{i,t}$ are the stochastic components of technology and employment, respectively. The model for $y_{i,t}$ is stationary and ergodic, regardless whether $\zeta_{i,t}$ or $v_{i,t}$ contains a unit root. The stochastic components of technology and employment cannot be empirically separated and are typically modeled by a country-specific autoregressive process. However, since common factors are likely to exist we can capture these terms by a multi-factor structure.

Model (25) can be expressed as a generalized dynamic panel data model, where the observations are modeled by the Student's t density (C) in Table 1. We consider the signal $z_{i,t}$ in (2) with $w_{i,t} = 0$ and $\epsilon_{i,t}$ as in (3). Further, $a_{i,t} = (1, y_{i,t-1})'$ such that the 2×1 vector μ_i captures the mean growth rate and the slope, $b_{i,t} = b_i$ is the $r \times 1$ vector of factor loadings and ξ_t is the $r \times 1$ vector of common factors. The common factors follow independent autoregressive processes of order 2 and capture technology and employment shocks which may contain business cycle effects.

We estimate the coefficients of model specifications with Student's t and Gaussian (fix $v = 1000$) densities. The most important parameter estimates for the Gaussian and Student's t models are presented in Table 5 for the models with $r = 2$ common factors. For $r = 1$, $r = 3$ and a variety of other model specifications similar results are obtained. When comparing the Student's t with the Gaussian specification, we find that the Student's t model yields a much higher loglikelihood value, both evaluated at their maximum likelihood estimates. The degrees of freedom v is estimated as 2.41 which clearly indicates that many large and outlying shocks are present in the OECD PWT dataset.

The estimated mean growth rates g_i and convergence parameters γ_i are plotted per country in Fig. 3. We find that the estimates for the Student's t model are extracted with much more
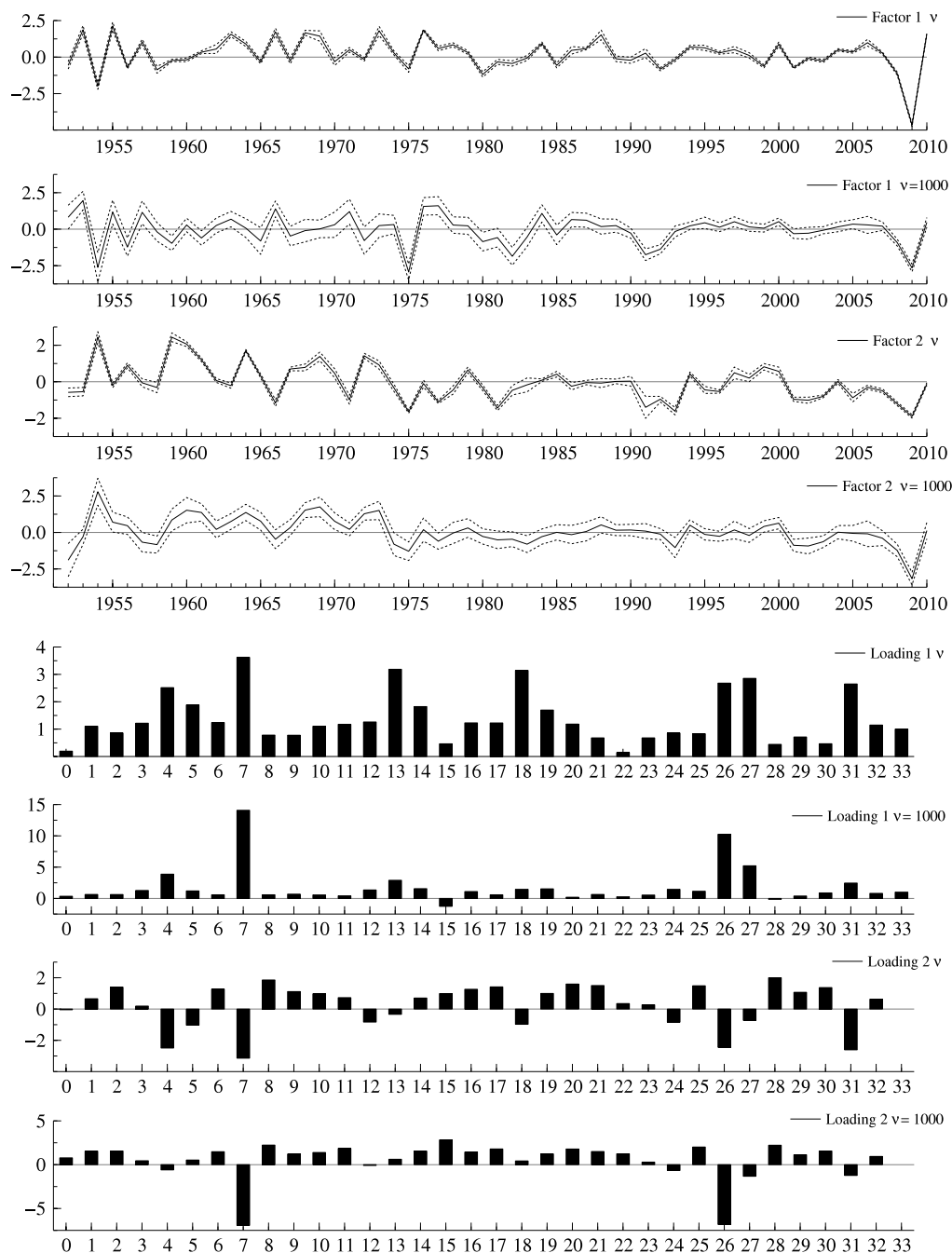
Fig. 4. Top panel: Common time-varying factor estimates and 95 % confidence bounds (dotted lines) for the Student's t and Gaussian dynamic panel data models. Bottom panel: Factor loadings b_i for the Student's t and Gaussian dynamic panel data models. The effects are computed based on estimated parameters $\tilde{\psi}$ (given in Table 5).

precision. The estimates for the Gaussian model correspond to the estimates reported by Lee et al. (1997) where the stochastic growth model is also considered for the levels of per capita output. The mean growth rates g_i are positive and on average 2.47 for the Student's t model and 2.44 for the Gaussian model. The implied speed of convergence γ_i has mean around 0.3 for the Student's t model and around 0.2 for the Gaussian model. Large differences in convergence rates exist for different countries.

In Fig. 4 we present the estimated common trends ξ_t and loadings b_i for the Student's t and Gaussian models. For the Gaussian model, the 95% confidence bounds for the estimates are much wider. For the Student's t model a more pronounced business cycle emerges from the factor estimates. The first factor is normalized to the US growth rate. Each country loads positively on this factor. The second factor corresponds to the western countries

where economic growth has slowed down in the last decade. The large values found for the loadings for the Gaussian model can be explained by the large outliers in the growth rates. Examples are Estonia, Slovenia and Slovakia.

For all models considered in Sections 5.1–5.3 we calculated the Wald test statistics for the importance sampling weights, see the discussion in Section 4.2. All test statistics are sufficiently negative, which indicates that a $\sqrt{M}$ convergence rate exists for all models. The individual results for the weight tests are presented in the Technical Appendix.

6. Conclusion

We have developed simulation-based methodology for the estimation of parameters in a general class of dynamic panel data

Table 5
Summary of the parameter estimation results (and standard errors in lower case) for the economic growth rate application. The model considered is given by observation density (C) in Table 1 with the signal specification $z_{i,t}$ as discussed in Section 5.3. The data panel consists of the growth rates for 34 countries between 1950 and 2010 listed in the Appendix.

	Model 1: $y_{i,t} \sim t(z_{i,t}, \sigma_\zeta, \nu)$	Model 2: $y_{i,t} \sim t(z_{i,t}, \sigma_\zeta, 1000)$
ν	2.413 $_{0.227}$	1000
σ	4.741 $_{2.025}$	2.788 $_{0.049}$
δ_1	1.680 $_{0.177}$	1.955 $_{0.235}$
δ_2	0.321 $_{0.030}$	0.200 $_{0.033}$
$\sigma_{\mu,1}$	0.291 $_{0.143}$	0.311 $_{0.130}$
$\sigma_{\mu,2}$	0.086 $_{0.020}$	0.118 $_{0.024}$
$h_{1,1}$	0.102 $_{0.075}$	0.222 $_{0.165}$
$h_{1,2}$	-0.005 $_{0.034}$	-0.012 $_{0.015}$
$h_{2,1}$	0.236 $_{0.198}$	0.359 $_{0.149}$
$h_{2,2}$	-0.052 $_{0.012}$	-0.042 $_{0.014}$
$\log \hat{p}(y)$	-7581.66	-7807.39
AIC	15317.06	15768.52

models with cross-section and time-varying random effects. The new estimation method felicitates the modeling of high-dimensional non-Gaussian panel data. 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 the Monte Carlo study we have given clear evidence of the validity of our estimation methods for finite samples. The empirical illustrations highlight the relevance and flexibility of our methods.

Acknowledgments

Appendix

Current OECD countries: 0. Australia (AUS), 1. Austria (AUT), 2. Belgium (BEL), 3. Canada (CAN), 4. Chile (CHL), 5. Czech Republic (CZE), 6. Denmark (DNK), 7. Estonia (EST), 8. Finland (FIN), 9. France (FRA), 10. Germany (DUE), 11. Greece (GRC), 12. Hungary (HUN), 13. Iceland (ISL), 14. Ireland (IRL), 15. Israel (ISR), 16. Italy (ITA), 17. Japan (JPN), 18. Korea (KOR), 19. Luxembourg (LUX), 20. Mexico (MEX), 21. Netherlands (NLD), 22. New Zealand (NLZ), 23. Norway (NOR), 24. Poland (POL), 25. Portugal (PRT), 26. Slovakia (SVK), 27. Slovenia (SVN), 28. Spain (ESP), 29. Sweden (SWE), 30. Switzerland (CHE), 31. Turkey (TUR), 32. United Kingdom (GBR), 33. United States (USA).

References

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econom. Stud. 58, 277–297.

Arellano, M., Bonhomme, S., 2011. Nonlinear panel data analysis. Annu. Rev. Econ. 3, 395–424.

Baltagi, B.H., 2005. Econometric Analysis of Panel Data. John Wiley & Sons, New York.

Bijleveld, C.C.J.H., Wijkman, M., 2009. Intergenerational continuity in convictions: a five-generation study. Criminal Behav. Men. Health 19, 142–155.

Binder, M., Pesaran, M.H., 1999. Stochastic growth models and their econometric implications. J. Econ. Growth 4, 139–183.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econometrics 87, 115–143.

Caselli, F., Esquivel, G., Lefort, F., 1996. Reopening the convergence debate: a new look at cross-country growth empirics. J. Econ. Growth 1, 363–389.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econ. Stud. 47, 225–238.

Chib, S., 2008. Panel data modeling and inference: a bayesian primer. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Springer Verlag, Berlin Heidelberg, pp. 215–246.

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. J. Financ. Econ. 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. J. R. Stat. Soc., Ser. B 62, 3–56.

Durbin, J., Koopman, S.J., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89, 603–616.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods, second ed.. Oxford University Press, Oxford.

Durlauf, S.N., Johnson, P., Temple, J., 2005. Growth econometrics. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth, vol. 1. Elsevier, Amsterdam, North-Holland, pp. 555–677. Part A.

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. In: Computer Science and Statistics. Proceedings of the 23rd Symposium on the Interface. Seattle, Washington, April 21–24. pp. 571–578.

Geweke, J., Keane, M., 2001. Computationally intensive methods for integration in econometrics. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, V. Elsevier, Amsterdam, North-Holland, pp. 3465–3568.

Hajivassiliou, V., 1990. Smooth Simulation Estimation of Panel Data LDV Models. Mimeo, Yale University.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Heckman, J.J., 1981a. Heterogeneity and state dependence. In: Rosen, S. (Ed.), Studies of Labor Markets. The University of Chicago Press, Chicago, pp. 91–140. The National Bureau of Economic Research.

Heckman, J.J., 1981b. Statistical models for discrete panel data. In: Manski, C.F., McFadden, D.L. (Eds.), Structural Analysis of Discrete Data and Econometric Applications. The MIT Press, Cambridge, pp. 113–177.

Heiss, F., 2008. Sequential numerical integration in nonlinear state space models for microeconometric panel data. J. Appl. Econometrics 23, 373–389.

Heston, A., Summers, R., Aten, B., 2012. Penn world table version 7.1 Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, Cambridge.

Hsiao, C., Pesaran, M.H., 2008. Random coefficients models. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Springer Verlag, Berlin, Heidelberg, pp. 187–216.

Islam, N., 1995. Growth empirics: a panel data approach. Q. J. Econ. 110, 1127–1170.

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., van der Wel, M., 2011. Maximum likelihood estimation for dynamic factor models with missing data. J. Economic Dynamics Control 35, 1358–1368.

Jungbacker, B., Koopman, S.J., 2008, 'Likelihood-based Analysis for Dynamic Factor Models'. Tinbergen Institute Discussion Paper: 08-007/4.

Keane, M.P., 1994. A computationally practical simulation estimator for panel data. Econometrica 62, 95–116.

Koopman, S.J., Lucas, A., 2008. A non-Gaussian panel time series model for estimating and decomposing default risk. J. Bus. Econom. Statist. 26, 510–525.

Koopman, S.J., Ooms, M., Lucas, A., Montfort, K., Van der Geest, V., 2008. Estimating systematic continuous-time trends in recidivism using a non-Gaussian panel data model. Stat. Neerl. 62, 104–130.

Koopman, S.J., Shephard, N., Creal, D.D., 2009. Testing the assumptions behind importance sampling. J. Econometrics 149, 2–11.

Lee, L.F., 1997. Simulated maximum likelihood estimation of dynamic discrete choice models, Some Monte Carlo results. J. Econometrics 82, 1–35.

Lee, K., Pesaran, H.M., Smith, R., 1997. Growth and convergence in a multi-country empirical stochastic solow model. J. Appl. Econometrics 12, 357–392.

Liesenfeld, R., Moura, G.V., Richard, J.F., 2010. Determinants and dynamics of current account reversals: an empirical analysis. Oxford Bull. Econ. Statist. 72, 486–517.

Liesenfeld, R., Richard, J., 2008. Simulation techniques for panels: efficient sampling. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Springer Verlag, Berlin, Heidelberg, pp. 215–246.

Liesenfeld, R., Richard, J.F., 2010. Efficient estimation of probit modeld with correlated errors. J. Econometrics 2010, 367–376.

Monahan, J.F., 2001. Numerical Methods of Statistics. Cambridge University Press, Cambridge.

Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer Verlag, New York.

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

Pesaran, H.M., 2007. A pair-wise approach to testing for output and growth convergence. J. Econometrics 138, 312–355.

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

Phillips, P.C.B., Sul, D., 2009. Economic transition and growth. J. Appl. Econometrics 24, 1153–1185.

Richard, J.F., Zhang, W., 2007. Efficient high-dimensional importance sampling. J. Econometrics 141, 1385–1411.

Ripley, B.D., 1987. Stochastic Simulation. John Wiley & Sons, New York.

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. Statist. Sinica 13, 255–274.

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 220, 147–162.

Vella, F., Verbeek, M., 1998. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. J. Appl. Econometrics 13, 163–183.

Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. J. Appl. Econometrics 20, 39–54.