Technical Appendix for: Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time

G. Mesters[(a,b)*] *and S.J. Koopman*[(b,c)]

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

January 28, 2014

Abstract

In this technical appendix we present the details for the likelihood evaluation procedure of Mesters & Koopman (2014, Section 3) and the estimation of the posterior random effects. Further, we present the complete set of simulation and empirical results that are discussed in Mesters & Koopman (2014).

1 Details Likelihood Evaluation

The likelihood estimate $\hat{p}(y)$ for the generalized dynamic panel data model (Mesters & Koopman (2014, Section 2)) is given in Mesters & Koopman (2014, equation 10) by

$$\hat{p}(y) = g(y;\hat{\xi})g(y;\hat{\mu})\frac{1}{M}\sum_{i=1}^{M} w^{(i)}, \tag{1}$$

where the weights are

$$w^{(i)} = p(y|\mu^{(i)},\xi^{(i)};x)/\left[g(y|\xi^{(i)};\hat{\mu})g(y|\mu^{(i)};\hat{\xi})\right]. \tag{2}$$

*Corresponding author: G. Mesters, PhD candidate Department of Econometrics, VU University Amsterdam, The Netherlands. Contact address: NSCR, PO box 71304, 1008 BH Amsterdam, The Netherlands, tel. nr. +31 20 59 89405, email: gmesters@nscr.nl, Technical Appendix: http://www.geertmesters.nl and http://personal.vu.nl/s.j.koopman

The 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 $g(\xi|y;\hat{\mu})$. We now discuss in detail how to construct these importance densities.

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 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

$$y_{i,t} = c_{i,t} + \epsilon_{i,t} + u_{i,t}, \qquad u_{i,t} \sim \text{NID}(0, d_{i,t}^2), \tag{3}$$

where $c_{i,t}$ is a fixed constant, stochastic component $\epsilon_{i,t}$ is given by Mesters & Koopman (2014, equation 3) 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 (3) since it is fixed at time t. The constants $c_{i,t}$ and $d_{i,t}$ are chosen such that (3) 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

$$g(y|\mu, \xi) = \prod_{i=1}^{N}\prod_{t=1}^{T} g(y_{i,t}|\mu_i, \xi_t), \qquad \text{with} \qquad g(y_{i,t}|\mu_i, \xi_t) \equiv \text{NID}(c_{i,t} + \epsilon_{i,t}, d_{i,t}^2). \tag{4}$$

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 (3) and Mesters & Koopman (2014, equation 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 $i = 1, \dots, N$ and $t = 1, \dots, T$;

(ii) Given the set of two equations

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

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where $p(y_{i,t}|z_{i,t})$ is the observation model Mesters & Koopman (2014, equation 5) and $g(y_{i,t}|\epsilon_{i,t})$ is given by (4), 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 $\tilde{\mu} = \mathrm{E}_g(\mu|y;\xi^*)$ from the resulting model (3) with $\xi = \xi^*$, $c_{i,t} = c^*_{i,t}$ and $d_{i,t} = d^*_{i,t}$;

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

(v) Compute $\tilde{\xi} = \mathrm{E}_g(\xi|y;\mu^*)$ from the resulting model (3) with $\mu = \mu^*$, $c_{i,t} = c^*_{i,t}$ and $d_{i,t} = d^*_{i,t}$;

(vi) Replace ξ^* 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 $\tilde{\mu} = \hat{\mu} = \mathrm{argmax}_\mu\, p(\mu|y;\hat{\xi};x)$ and $\tilde{\xi} = \hat{\xi} = \mathrm{argmax}_\mu\, p(\xi|y;\hat{\mu};x)$. Further, it holds that $\{\hat{\mu},\hat{\xi}\} = \mathrm{argmax}_{\mu,\xi}\, p(\mu,\xi|y;x)$.

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 (3) 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 simplification becomes clear when we concatenate the observations $y_{i,t}$ over the time index t and when we consider the approximating model (3). We then obtain

$$\bar{y}_i = \bar{c}_i + \bar{A}_i\mu_i + \bar{\mathcal{B}}_i + \bar{u}_i, \qquad \bar{u}_i \sim \mathrm{NID}(0,\bar{D}_i), \qquad i = 1,\ldots,N, \tag{5}$$

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 $\bar{u}_i = (u_{i,1},\ldots,u_{i,T})'$. The $T\times T$ variance matrix $\bar{D}_i$ is diagonal by construction, with elements $d^2_{i,1},\ldots,d^2_{i,T}$ on the main diagonal. Based on (5), the computation of $\mathrm{E}_g(\mu|y;\xi^*)$ can be performed using standard multivariate normal regression theory. In particular, it holds that

$$\mathrm{E}_g(\mu|y;\xi^*) = \mathrm{E}_g(\mu;\xi^*) + \mathrm{Cov}_g(\mu,y;\xi^*)\mathrm{Var}_g(y;\xi^*)^{-1}\left[y - \mathrm{E}_g(y;\xi^*)\right], \tag{6}$$

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

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

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

Now consider step (v) where we need to compute $\text{E}_g(\xi|y;\mu^*)$. Given a value of μ^*, approximating model (3), 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

$$y_t = c_t + \mathcal{A}_t + B_t\xi_t + u_t, \qquad u_t \sim \text{NID}(0, D_t), \qquad t = 1,\dots,T, \tag{7}$$

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 $d^2_{1,t},\dots,d^2_{N,t}$ on the main diagonal. Based on (7) the computation of $\text{E}_g(\xi|y;\mu^*)$ is carried out using the Kalman filter and smoothing methods; see Anderson & Moore (1979) and Durbin & Koopman (2012, Chapter 4). Moreover, since D_t is diagonal the fast Kalman filter and smoothing methods from Koopman & Durbin (2003) can be used.

The evaluation of likelihood estimate $\hat{p}(y)$ in (1), requires M samples of μ and ξ 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 (3). The vector representations (5) and (7), 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 (5) and (7), 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 $\bar{y}^l = \left[(\bar{y}_1^l)',\dots,(\bar{y}_N^l)'\right]'$ and $y^l = \left[(y_1^l)',\dots,(y_T^l)'\right]'$. The resulting samples can be regarded as coming from $g(\mu|y;\hat{\xi})$ and $g(\xi|y;\hat{\mu})$, respectively. In Section 2 we present the computational gains in evaluating the likelihood, for both sets of importance densities. The computational improvements resulting from the transformations are high. Apart from the computational gains, we also need to use less common random numbers for sampling the same number of draws $\mu^{(i)}$ and $\xi^{(i)}$, regardless of the simulation smoother used.

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 (7), into low-dimensional vectors y_t^l, without losing information relevant for the extraction of ξ. 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 S_t is an $N\times N$ nonsingular projection matrix and where y_t, c_t and $\mathcal{A}_t$ are as given by (7) with $\mathcal{A}_t$ replaced by $\hat{\mathcal{A}}_t = (a'_{1,t}\hat{\mu}_1,\dots,a'_{N,t}\hat{\mu}_N)'$ and $\hat{\mu}_i$ is the vector of posterior modal individual-specific effects for time series i, for $i = 1,\dots,N$ and $t = 1,\dots,T$. The transformed observations are given by

$$y_t^* = \begin{bmatrix} y_t^l \\ y_t^h \end{bmatrix}, \quad \text{with} \quad \begin{matrix} y_t^l = S_t^l(y_t - c_t - \hat{\mathcal{A}}_t) \\ y_t^h = S_t^h(y_t - c_t - \hat{\mathcal{A}}_t) \end{matrix}, \quad S_t = \begin{bmatrix} S_t^l \\ S_t^h \end{bmatrix}, \quad t = 1, \ldots, T, \tag{8}$$

where the partioned projection matrices S_t^l and S_t^h 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 ξ_t. In particular, we aim for a model of the form

$$\begin{matrix} y_t^l = S_t^l B_t \xi_t + u_t^l, \\ y_t^h = u_t^h, \end{matrix} \qquad \begin{pmatrix} u_t^l \\ u_t^h \end{pmatrix} \sim N\left(0, \begin{bmatrix} D_t^l & 0 \\ 0 & D_t^h \end{bmatrix}\right), \tag{9}$$

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

Suitable matrices S_t, which lead to model (9) need to satisfy the following conditions; (a) matrices S_t needs to be of full rank to prevent the loss of information, (b) $S_t^h D_t S_t^{l\prime} = 0$ 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 ξ_t. Many matrix series S_t can be found that fulfill these conditions. A convenient choice is given by

$$S_t^l = \Delta_t' B_t' D_t^{-1}, \qquad \Delta_t \Delta_t' = (B_t' D_t^{-1} B_t)^{-1}, \tag{10}$$

with Δ_t being a lower triangular matrix. This choice for S_t^l results in

$$y_t^l = \Delta_t^{-1} \xi_t + u_t^l, \qquad u_t^l \sim \text{NID}(0, I_r), \qquad t = 1, \ldots, T, \tag{11}$$

where Δ_t^{-1} is a $r \times r$ lower triangular matrix, ξ_t is defined in Mesters & Koopman (2014, equation 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 & Koopman (2002) to r-dimensional vector series y_t^l and model (11), for $t = 1, \ldots, T$. 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. Further discussions of this transformation method are given in Jungbacker & Koopman (2008)

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 (5), with $\bar{\mathcal{B}}_i$ replaced by $\hat{\mathcal{B}}_i = (b_{i,1}' \hat{\xi}_1, \ldots, b_{i,T}' \hat{\xi}_T)'$. 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 $\bar{y}_i^*$ become simple rescaled averages of the variables in $\bar{y}_i$. Let

$$\bar{y}_i^* = \begin{bmatrix} \bar{y}_i^l \\ \bar{y}_i^h \end{bmatrix}, \qquad \text{with} \qquad \begin{matrix} \bar{y}_i^l = \bar{S}_i^l(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i) \\ \bar{y}_i^h = \bar{S}_i^h(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i) \end{matrix}, \qquad i = 1, \ldots, N. \tag{12}$$

The motivation of the transformation is the same as above. We require to sample μ_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

$$\begin{aligned} \bar{y}_i^l &= \bar{S}_i^l \bar{A}_i \mu_i + \bar{u}_i^l, \\ \bar{y}_i^h &= \bar{u}_i^h, \end{aligned} \qquad \left(\begin{array}{c} \bar{u}_i^l \\ \bar{u}_i^h \end{array} \right) \sim N\left(0, \left[\begin{array}{cc} \bar{D}_i^l & 0 \\ 0 & \bar{D}_i^h \end{array} \right]\right), \tag{13}$$

where $\bar{D}_i^l = \bar{S}_i^l \bar{D}_i \bar{S}_i^{l\prime}$ and $\bar{D}_i^h = \bar{S}_i^h \bar{D}_i \bar{S}_i^{h\prime}$ 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

$$\bar{S}_i^l = \bar{\Delta}_i' \bar{A}_i' \bar{D}_i^{-1}, \qquad \bar{\Delta}_i \bar{\Delta}_i' = (\bar{A}_i' \bar{D}_i^{-1} \bar{A}_i)^{-1}, \qquad . \tag{14}$$

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

$$\bar{y}_i^l = \bar{\Delta}_i^{-1} \mu_i + \bar{u}_i^l, \qquad \bar{u}_t^l \sim \text{NID}(0, I_q), \qquad i = 1, \ldots, N, \tag{15}$$

where $\bar{\Delta}_i^{-1}$ is a lower triangular $q \times q$ matrix, μ_i is given in Mesters & Koopman (2014, equation 4) 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 $\Sigma_\mu \bar{\Delta}_i^{-1}(\bar{\Delta}_i^{-1\prime} \Sigma_\mu \bar{\Delta}_i^{-1} + I_q)^{-1} \bar{y}_i^l$ and variance $\Sigma_\mu - \Sigma_\mu \bar{\Delta}_i^{-1}(\bar{\Delta}_i^{-1\prime} \Sigma_\mu \bar{\Delta}_i^{-1} + I_q)^{-1} \bar{\Delta}_i^{-1\prime} \Sigma_\mu$. Both expressions follow from the standard lemma (6).

Next we discuss the construction of the Monte Carlo likelihood estimate $\hat{p}(y)$ in equation (1). The estimate relies on densities $g(y; \hat{\mu})$ and $g(y; \hat{\xi})$, that are based on the approximating model (3). Density $\log g(y; \hat{\mu})$ can be computed from the prediction error decomposition of vector representation (7), with μ replaced by $\hat{\mu}$. This is obtained by a single pass through the Kalman filter, see Durbin & Koopman (2012, Chapter 7). Computational efficiency can increased by using the lower dimensional model (11), based on vector series y_t^l. In particular, Jungbacker & Koopman (2008) show that

$$\log g(y; \hat{\mu}) = \text{constant} + \log g(y^l; \hat{\mu}) - \frac{1}{2} \sum_{t=1}^{T} \log |D_t| + e_t' D_t^{-1} e_t, \tag{16}$$

where $y^l = (y_1^{l\prime}, \ldots, y_T^{l\prime})'$ 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 (11), which is a $r \times T$-dimensional problem.

Due to the independence of the μ_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

$i = 1, \ldots, N$. Based on model (15) we obtain

$$\log g(y;\hat{\xi}) = \text{constant} + \log g(\bar{y}^l;\hat{\xi}) - \frac{1}{2}\sum_{i=1}^{N} \log|\bar{D}_i| + \bar{e}_i' D_i^{-1}\bar{e}_i, \tag{17}$$

where $\bar{e}_i = \mathcal{M}_i(\bar{y}_i - \bar{c}_i - \hat{\mathcal{B}}_i)$ with $\mathcal{M}_i = I - \bar{A}_i(\bar{A}_i'\bar{D}_i^{-1}\bar{A}_i)^{-1}\bar{A}_i'\bar{D}_i^{-1}$. Logdensity $\log g(y;\hat{\xi})$ can therefore be cased on the $N \times q$-dimensional model (15).

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;

(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 (11) and (15), and compute importance weights $w^{(i)}$, as given in equation (2);

(iv) Evaluate logdensities $\log g(y;\hat{\mu})$ as in (16) and $\log g(y;\hat{\xi})$ as in (17);

(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 & 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 $\tilde{\psi}$. In Section 2 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 different dynamic panel data models. Durbin & 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 manifacture antithetic variables that balance for location and for scale. see Durbin & Koopman (2012, Section 11.4.3) for a detailed discussion.

1.1 Estimation of the posterior random effects

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 (2012, Chapter 11). Let $f(\mu,\xi)$ denote a general function of μ and ξ that is of interest. It holds that

$$\mathrm{E}_p\left[f(\mu,\xi)|y\right] = \int_\xi \int_\mu f(\mu,\xi) p(\mu,\xi|y;x) \; \mathrm{d}\mu \; \mathrm{d}\xi,$$

where $\mathrm{E}_p[\cdot|y]$ 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

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

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

$$\mathrm{E}_p\left[f(\mu,\xi)|y\right] = \frac{g(y;\hat{\xi})g(y;\hat{\mu})}{p(y)} \int_\xi \int_\mu f(\mu,\xi) w(y,\mu,\xi;\hat{\mu},\hat{\xi}) g(\mu|y;\hat{\xi})g(\xi|y;\hat{\mu}) \; \mathrm{d}\mu \; \mathrm{d}\xi,$$

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

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

And when dividing the two equations above we get

$$\mathrm{E}_p\left[f(\mu,\xi)|y\right] = \frac{\int_\xi \int_\mu f(\mu,\xi) w(y,\mu,\xi;\hat{\mu},\hat{\xi}) g(\mu|y;\hat{\xi})g(\xi|y;\hat{\mu}) \; \mathrm{d}\mu \; \mathrm{d}\xi}{\int_\xi \int_\mu w(y,\mu,\xi;\hat{\mu},\hat{\xi}) g(\mu|y;\hat{\xi})g(\xi|y;\hat{\mu}) \; \mathrm{d}\mu \; \mathrm{d}\xi},$$

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 (2).

2 Simulation Results

2.1 Diagnostic tests importance sampling weights

In Figures 1 - 9 we present the importance sampling diagnostics for all density, signal and parameters combinations given in Table 1. They are computed as discussed in Mesters & Koopman (2014, Section 4.2).

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)

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 $\lambda^{-1} = (\nu-2)\sigma^2$.

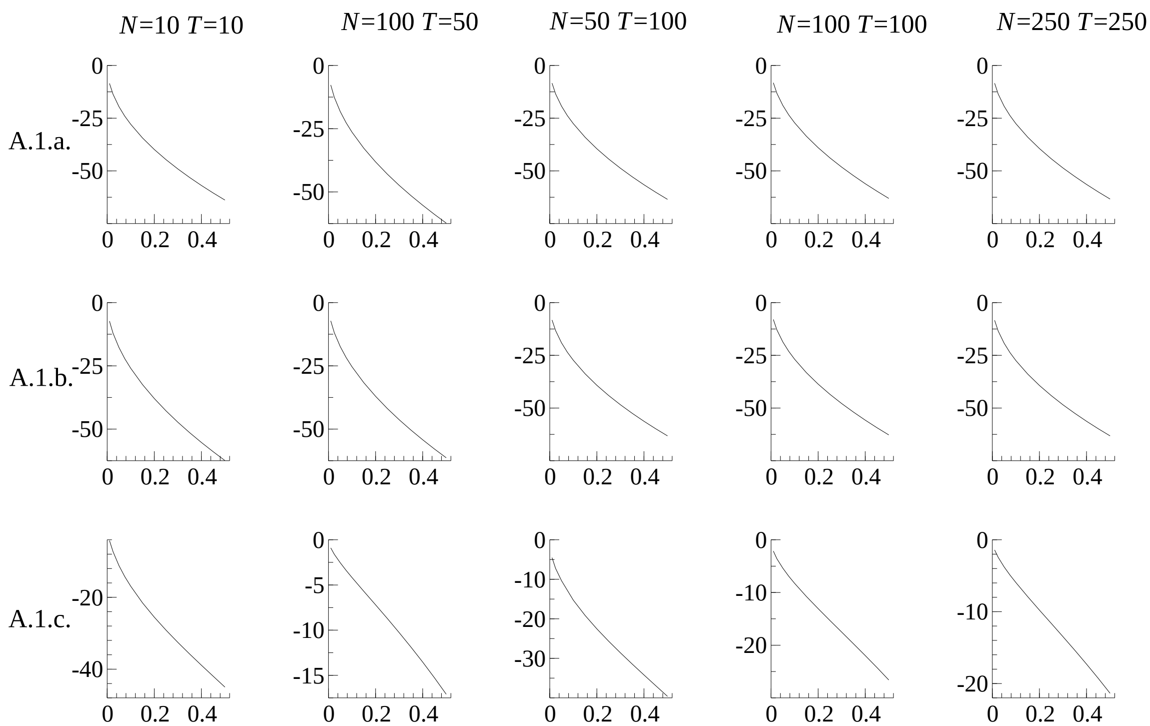

Figure 1: Diagnostic tests Binary signal 1

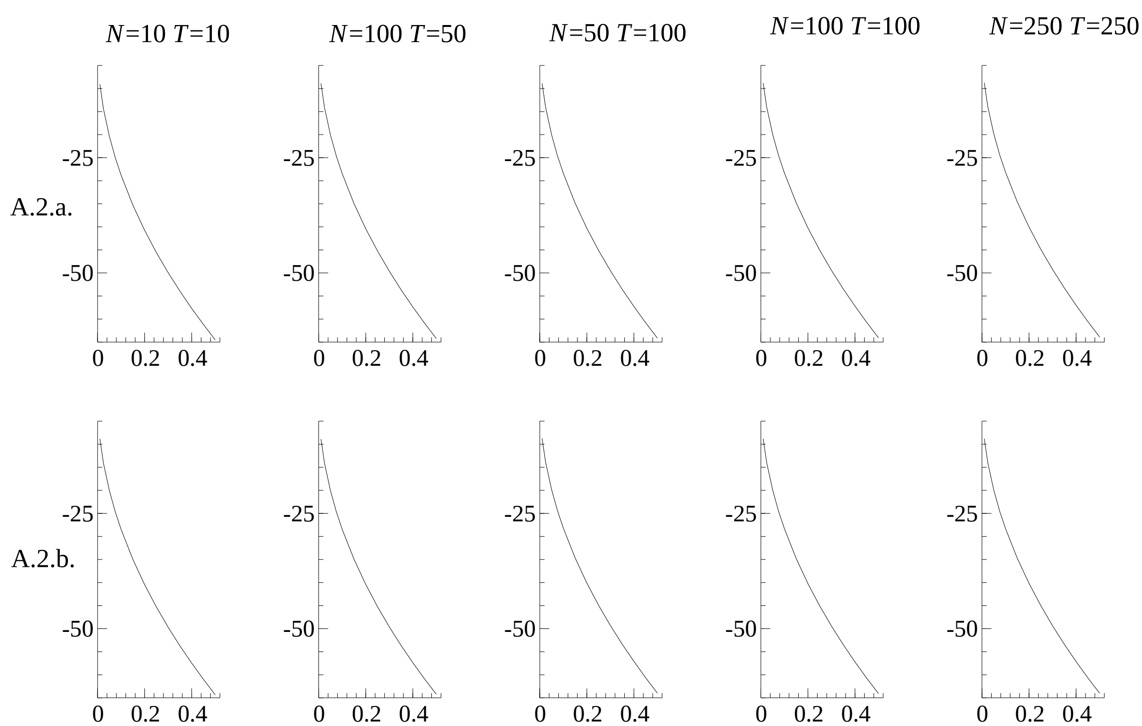

Figure 2: Diagnostic tests Binary signal 2

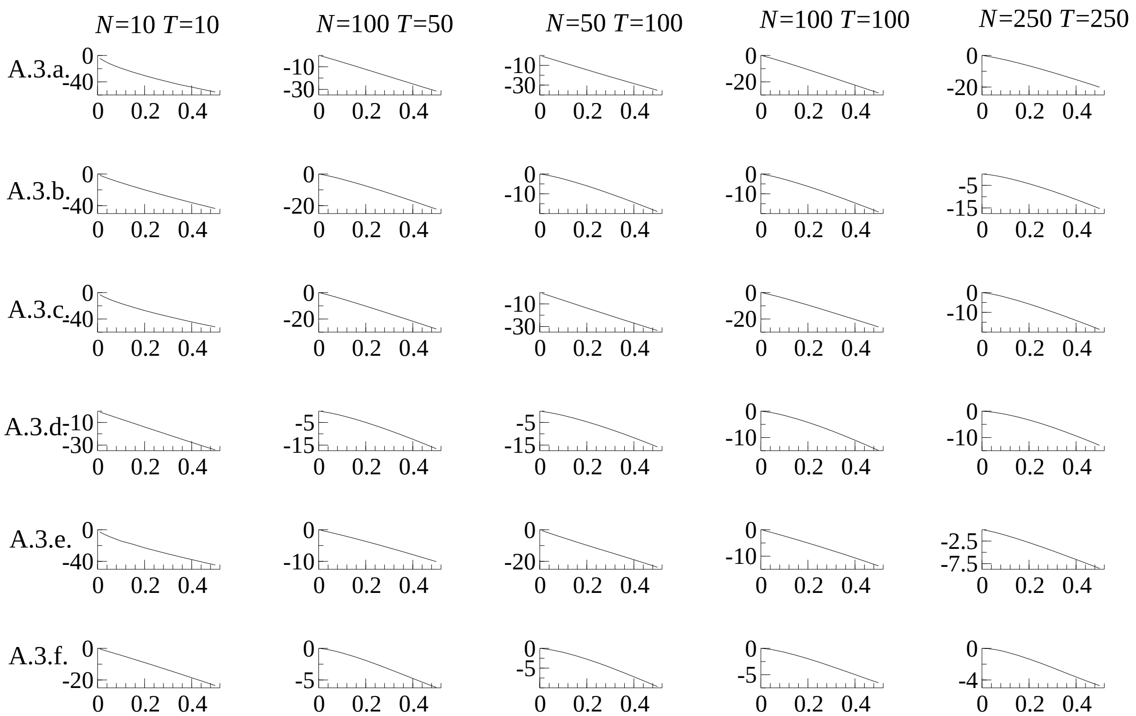

Figure 3: Diagnostic tests Binary signal 3

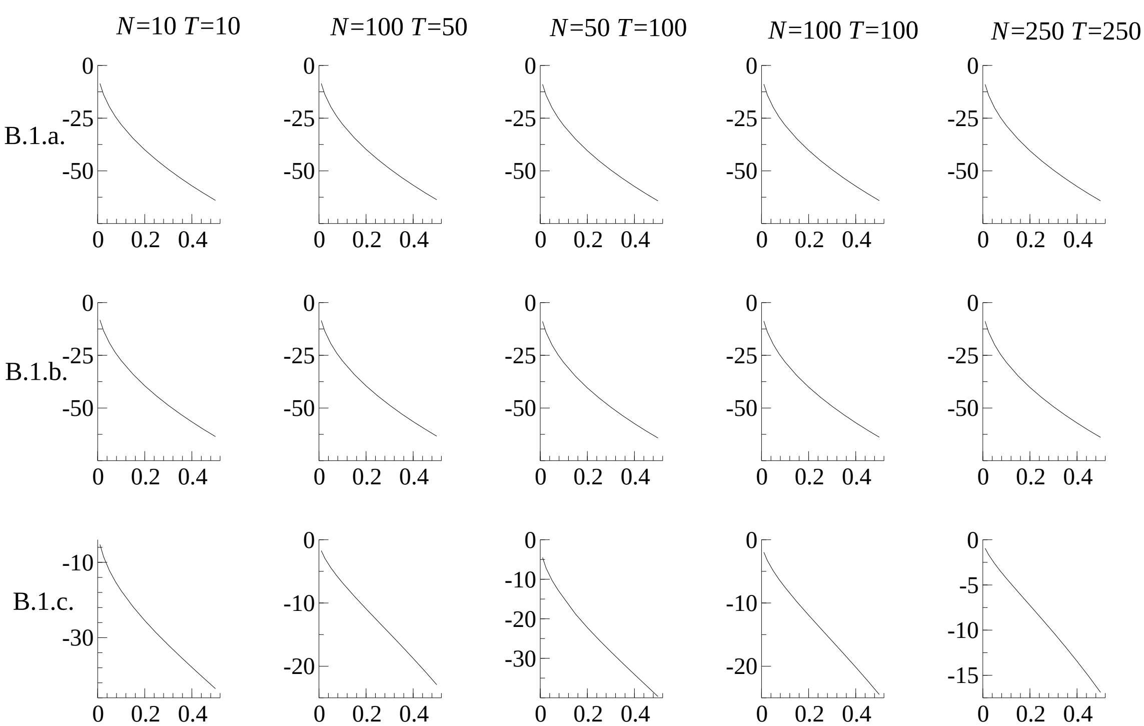

Figure 4: Diagnostic tests Binomial signal 1

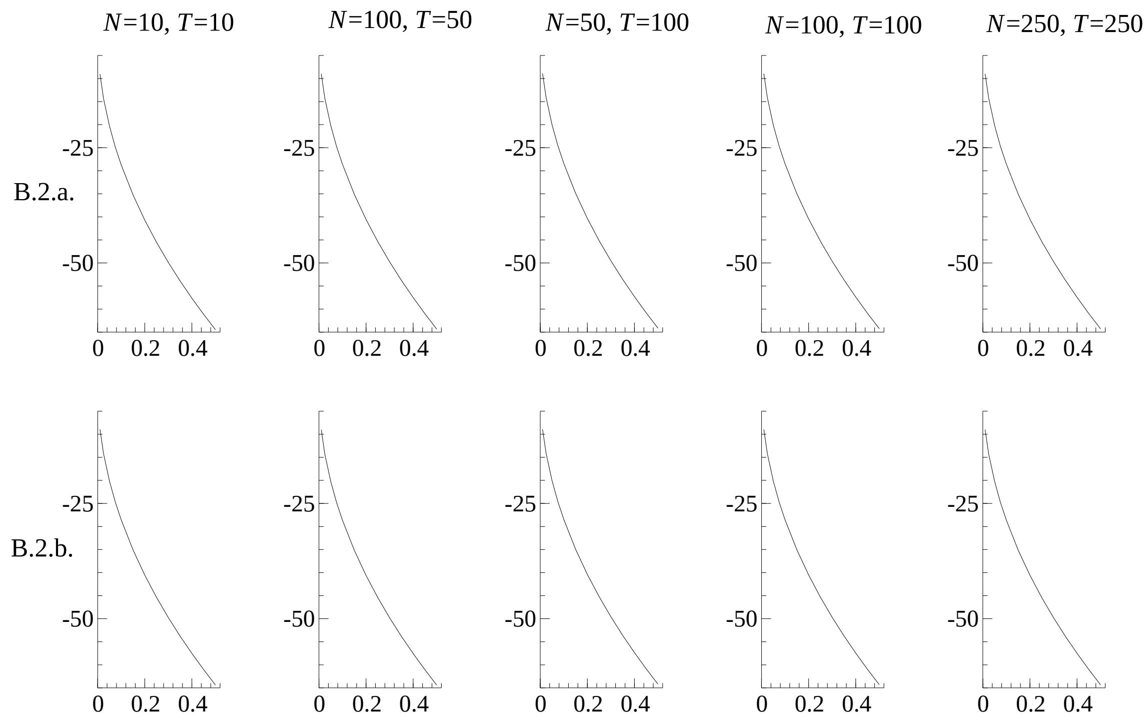

Figure 5: Diagnostic tests Binomial signal 2

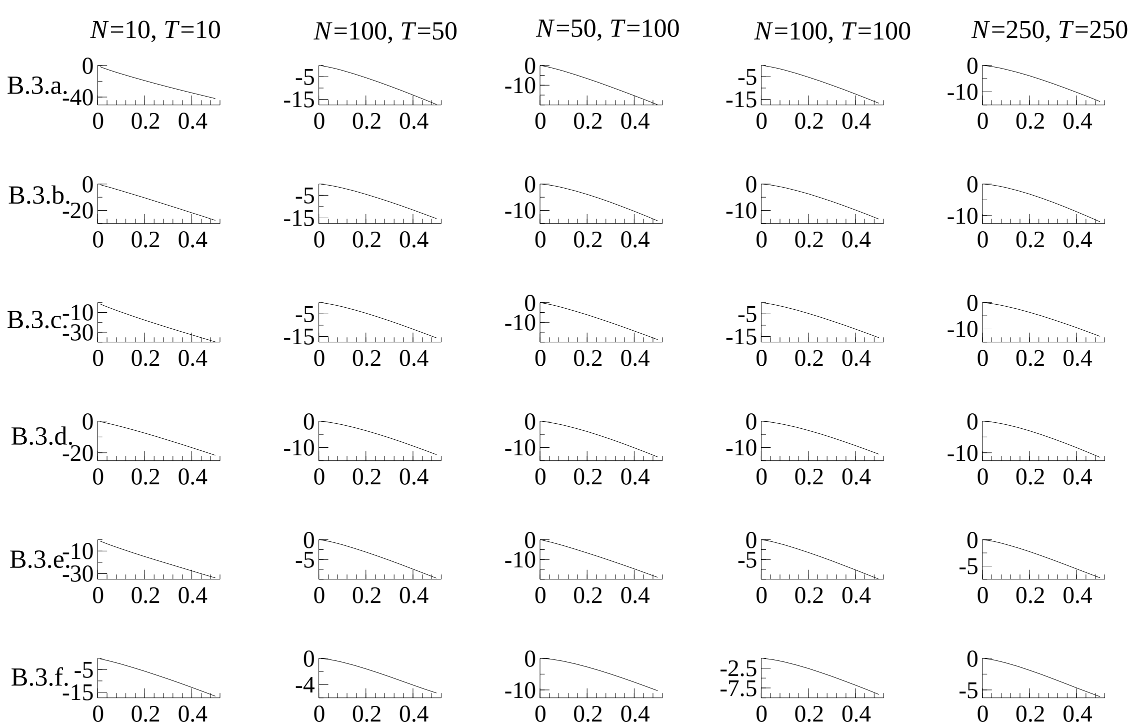

Figure 6: Diagnostic tests Binomial signal 3

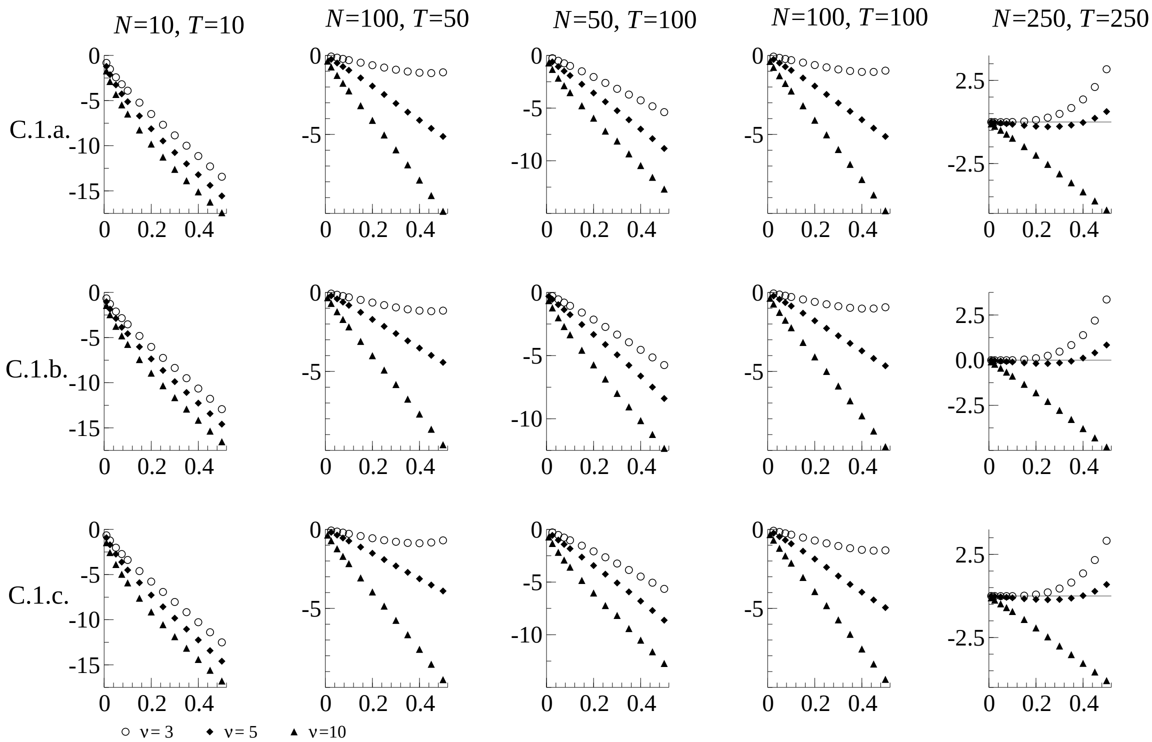

Figure 7: Diagnostic tests Student's t signal 1

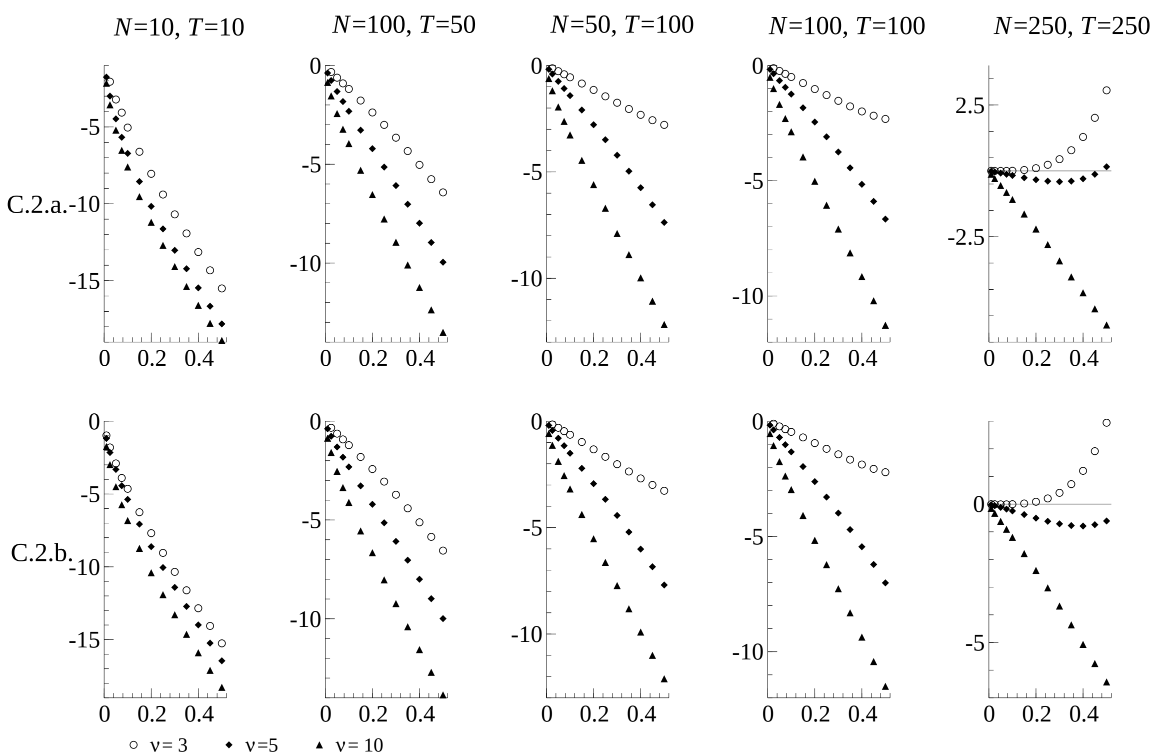

Figure 8: Diagnostic tests Student's t signal 2

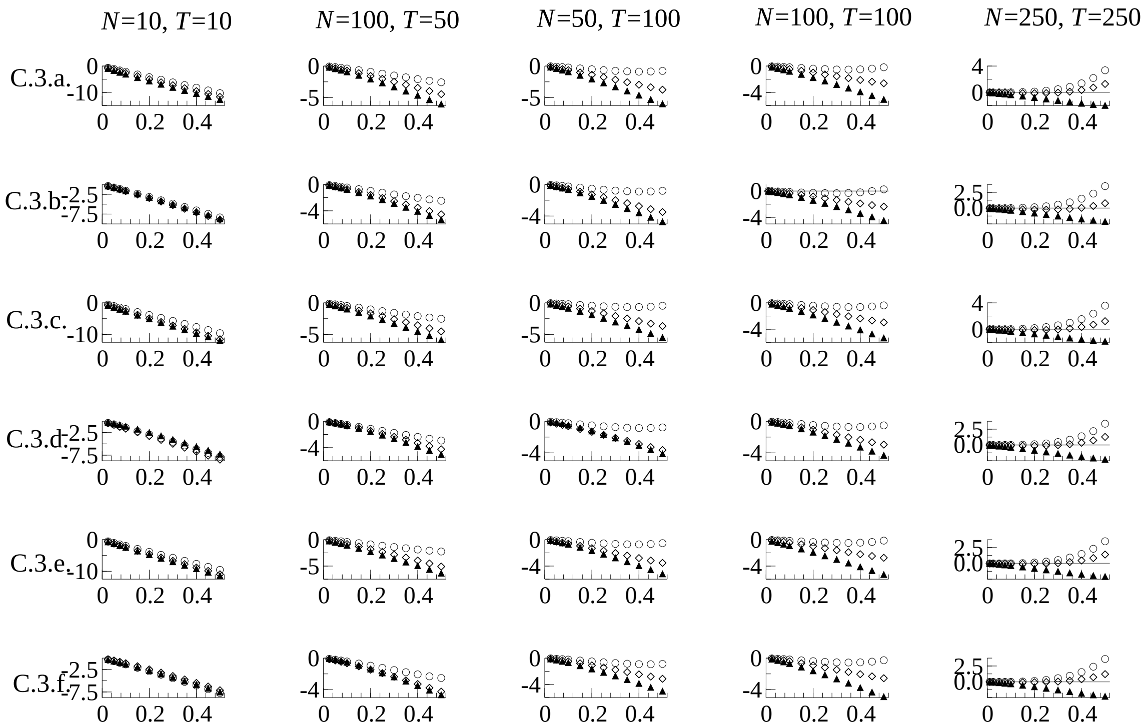

Figure 9: Diagnostic tests Student's t signal 3

2.2 Efficiency loss

We discussed in Mesters & Koopman (2014, Section 3) that our proposed implementation of importance sampling for the generalized dynamic panel data model is different from the standard implementation of Shephard & Pitt (1997) and Durbin & 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 (3). The corresponding likelihood is then estimated by

$$\hat{p}(y) = g(y) M^{-1} \sum_{i=1}^{M} w^{(i)}, \qquad w^{(i)} = \frac{p(y|\mu^{(i)}, \xi^{(i)}; x)}{g(y|\mu^{(i)}, \xi^{(i)})}, \tag{18}$$

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 1.

It is anticipated that the variance of the importance weight function increases when the SPDK method in (18) is replaced by our method in (1) because Algorithm A does not account for the dependence between μ and ξ in $g(\mu, \xi|y)$. The efficiency loss due to Algorithm A can also be investigated for the small panel. It may provide some insight into a possible necessary increase in the number of draws M compared to SPDK. For this purpose we carry out the the following simulation experiment. We generate fifty data panels based on signal 3 and for each combination of observation density and parameter values reported in Table 1.

For each of these simulated data panels, we obtain likelihood estimates at the true parameter values under a hundred different random seeds. We are interested in the standard deviation of the likelihood estimates for different values of M and computed by

$$SD_M = (1/50)\sum_{i=1}^{50}\left[(1/100)\sum_{j=1}^{100}(\log \hat{p}_j(y^i) - \log \bar{p}(y^i))^2\right]^{0.5},$$

where $\log \bar{p}(y^i) = (1/100)\sum_{j=1}^{100}\log \hat{p}_j(y^i)$ with $\hat{p}_j(y^i)$ as the likelihood estimate based on M draws, for the jth random seed and for the ith panel y^i.

The resulting values for SD_M are reported in Table 2 for $M = 50, 100, 1000$. For the Binary and Binomial densities, we observe a substantial increase in the variance of the likelihood estimate from Algorithm A when compared to the SPDK implementation, for each value of M. The increase decreases rapidly when the variance of the individual-specific effect σ_μ^2 increases. For the Student's t densities the increase in the variance is smaller and sometimes, for $\nu = 3$, Algorithm A is more efficient than SPDK. These results are only indicative since we can carry out these comparisons only for small panel dimensions. However, the increase of SD_M can be offset by increasing the number of draws M. Any degree of accuracy can be achieved and given the large computational improvements, as documented below, this seems a minor difficulty. We notice that SPDK is clearly not feasible for even modest panel dimensions.

2.3 Likelihood evaluation times

A major improvement for our simulation based estimator is proposed in Mesters & Koopman (2014, Section 3.2) where it is shown how large panel and time series data vectors can be collapsed to much smaller dimensions. We investigate the gains in computing time by simulating 100 data panels for each possible design given in Table 1. For each simulated data panel, we evaluate the likelihood by Algorithm B as described in Section 1 with $M = 1000$. We consider the standard implementation based on vector series y_t and $\bar{y}_i$ as well as the collapsed implementation based on y_t^l and $\bar{y}_i^l$ of Section 1. The average evaluation time for the likelihood for the collapsed implementation and the ratio between the evaluation times for the collapsed and standard implementations is presented in Table 3 for signal 1.b, 2.b and 3.d. The reduction in evaluation times are of course the same for different parameter values. The likelihood evaluation procedure based on the collapsed vectors is between 2 and 10 times faster compared to evaluation without collapsing the vectors. Most computational gains are due to the collapse of panel dimension N for the sampling of time-varying effect ξ.

2.4 Parameter estimation results

We report the average bias and standard deviation of the parameter estimation results in Tables 4 - 8 for all densities and signals. See Mesters & Koopman (2014, Section 4.4) for further details.

M	Signal	A; Binary	B; Binomial	C; Student's t ($\nu = 3$)	C; Student's t ($\nu = 5$)	C; Student's t ($\nu = 10$)
50	3.a	[0.003 , 0.028]	[0.003 , 0.061]	[0.170 , 0.184]	[0.097 , 0.113]	[0.050 , 0.074]
	3.b	[0.005 , 0.056]	[0.004 , 0.077]	[0.175 , 0.190]	[0.102 , 0.122]	[0.053 , 0.085]
	3.c	[0.011 , 0.040]	[0.006 , 0.068]	[0.189 , 0.202]	[0.117 , 0.132]	[0.061 , 0.084]
	3.d	[0.011 , 0.068]	[0.006 , 0.084]	[0.197 , 0.212]	[0.121 , 0.145]	[0.063 , 0.097]
	3.e	[0.049 , 0.064]	[0.046 , 0.085]	[0.196 , 0.207]	[0.124 , 0.138]	[0.065 , 0.086]
	3.f	[0.050 , 0.096]	[0.046 , 0.105]	[0.257 , 0.216]	[0.135 , 0.152]	[0.067 , 0.101]
100	3.a	[0.002 , 0.020]	[0.002 , 0.042]	[0.136 , 0.140]	[0.071 , 0.084]	[0.034 , 0.052]
	3.b	[0.003 , 0.040]	[0.003 , 0.050]	[0.139 , 0.138]	[0.076 , 0.085]	[0.037 , 0.056]
	3.c	[0.007 , 0.027]	[0.004 , 0.045]	[0.157 , 0.157]	[0.089 , 0.100]	[0.043 , 0.060]
	3.d	[0.008 , 0.044]	[0.004 , 0.056]	[0.158 , 0.154]	[0.092 , 0.101]	[0.044 , 0.065]
	3.e	[0.033 , 0.043]	[0.031 , 0.056]	[0.165 , 0.161]	[0.098 , 0.107]	[0.047 , 0.063]
	3.f	[0.034 , 0.061]	[0.031 , 0.071]	[0.169 , 0.159]	[0.101 , 0.108]	[0.048 , 0.068]
1000	3.a	[0.001 , 0.006]	[0.001 , 0.014]	[0.047 , 0.049]	[0.024 , 0.026]	[0.012 , 0.016]
	3.b	[0.001 , 0.010]	[0.001 , 0.016]	[0.048 , 0.049]	[0.025 , 0.027]	[0.012 , 0.017]
	3.c	[0.002 , 0.008]	[0.001 , 0.014]	[0.058 , 0.061]	[0.031 , 0.032]	[0.015 , 0.017]
	3.d	[0.003 , 0.011]	[0.001 , 0.017]	[0.058 , 0.060]	[0.032 , 0.035]	[0.015 , 0.019]
	3.e	[0.011 , 0.014]	[0.010 , 0.018]	[0.062 , 0.063]	[0.035 , 0.035]	[0.016 , 0.018]
	3.f	[0.012 , 0.020]	[0.010 , 0.023]	[0.070 , 0.063]	[0.035 , 0.038]	[0.018 , 0.019]

Table 2: We present the standard deviations of the Monte Carlo likelihood estimator for the proposed importance sampler (1) and the standard SPDK importance sampler (18). The results presented are [standard deviation estimator (18) , standard deviation estimator (1)]. The standard deviations are computed as discussed in Section 2.1. The proposed estimator (1) is implemented using the importance densities $g(\xi|y^l;\hat{\mu})$ and $g(\mu|y^l;\hat{\xi})$, which are based on the collapsed panels. The panel sizes are $N = 10$ and $T = 10$.

	$N = 100$ $T = 50$	$N = 50$ $T = 100$	$N = 100$ $T = 100$	$N = 250$ $T = 250$
		A. Binary		
1.b	[0.286, 0.635]	[0.281,0.532]	[0.584,0.568]	[3.387,0.594]
2.b	[0.334, 0.169]	[0.369,0.164]	[0.665,0.161]	[3.705,0.160]
3.d	[0.591, 0.172]	[0.607,0.178]	[1.240,0.210]	[9.688,0.223]
		B. Binomial		
1.b	[0.184,0.733]	[0.178,0.675]	[0.361,0.702]	[2.165,0.692]
2.b	[0.207,0.143]	[0.226,0.146]	[0.409,0.145]	[2.359,0.155]
3.d	[0.386,0.180]	[0.401,0.187]	[0.852,0.187]	[7.006,0.308]
		C. Student's t		
1.b	[0.115,0.657]	[0.111,0.590]	[0.228,0.621]	[1.339,0.597]
2.b	[0.191,0.100]	[0.207,0.102]	[0.379,0.102]	[2.129,0.098]
3.d	[0.663,0.164]	[0.797,0.222]	[1.692,0.221]	[11.504,0.280]

Table 3: Average likelihood evaluation time for the collapsed likelihood implementation (left in seconds) and the average ratio between the the collapsed and standard implementations of the likelihood evaluation as discussed above (the average is over 100 panels). The reduction is achieved by sampling from $\mu^{(i)}$ and $\xi^{(i)}$ from $g(\xi|y^l;\hat{\mu})$ and $g(\mu|y^l;\hat{\xi})$ instead of from $g(\xi|y;\hat{\mu})$ and $g(\mu|y;\hat{\xi})$, respectively. The signals are taken as in Table 1. For each model the likelihood is evaluated as discussed in Section 1 and by using $M = 1000$ samples from the importance densities. For the Student's t density the 5 degrees of freedom were used.

3 Weights empirical studies

In Figure 10 we present the diagnostic tests for the weights pertaining to the empirical studies. The weights are computed as discussed in Mesters & Koopman (2014, Section 4.2).

Code	N	T	γ	β	σ_μ	h	σ_η
A.1.a.	100	50	-0.003 $_{0.059}$	0.003 $_{0.039}$	-0.006 $_{0.059}$		
	50	100	-0.007 $_{0.059}$	-0.001 $_{0.038}$	-0.002 $_{0.056}$		
	100	100	-0.008 $_{0.047}$	-0.001 $_{0.026}$	0.002 $_{0.041}$		
	250	250	0.002 $_{0.017}$	-0.001 $_{0.012}$	0.004 $_{0.024}$		
A.1.b.	100	50	0.001 $_{0.076}$	0.001 $_{0.038}$	-0.001 $_{0.084}$		
	50	100	-0.009 $_{0.062}$	-0.001 $_{0.038}$	-0.001 $_{0.101}$		
	100	100	-0.008 $_{0.051}$	-0.001 $_{0.027}$	0.004 $_{0.074}$		
	250	250	0.000 $_{0.018}$	-0.002 $_{0.011}$	0.008 $_{0.044}$		
A.1.c.	100	50	0.001 $_{0.113}$	0.002 $_{0.045}$	-0.019 $_{0.230}$		
	50	100	-0.003 $_{0.094}$	-0.005 $_{0.049}$	-0.008 $_{0.352}$		
	100	100	-0.003 $_{0.067}$	-0.000 $_{0.038}$	0.022 $_{0.203}$		
	250	250	0.003 $_{0.026}$	0.000 $_{0.013}$	0.015 $_{0.133}$		
A.2.a.	100	50	-0.006 $_{0.052}$	-0.000 $_{0.038}$		-0.039 $_{0.333}$	-0.021 $_{0.053}$
	50	100	0.003 $_{0.045}$	0.004 $_{0.036}$		-0.106 $_{0.353}$	-0.009 $_{0.053}$
	100	100	-0.003 $_{0.042}$	0.000 $_{0.025}$		-0.020 $_{0.188}$	-0.005 $_{0.035}$
	250	250	0.001 $_{0.014}$	-0.002 $_{0.011}$		-0.004 $_{0.097}$	0.000 $_{0.013}$
A.2.b.	100	50	0.002 $_{0.057}$	-0.002 $_{0.036}$		-0.038 $_{0.114}$	-0.009 $_{0.059}$
	50	100	-0.004 $_{0.062}$	0.003 $_{0.033}$		-0.021 $_{0.070}$	0.001 $_{0.041}$
	100	100	-0.005 $_{0.051}$	-0.000 $_{0.026}$		-0.025 $_{0.070}$	0.004 $_{0.034}$
	250	250	0.001 $_{0.018}$	-0.002 $_{0.011}$		-0.008 $_{0.028}$	0.001 $_{0.013}$
A.3.a.	100	50	-0.003 $_{0.061}$	0.001 $_{0.039}$	0.003 $_{0.055}$	-0.026 $_{0.290}$	-0.012 $_{0.053}$
	50	100	-0.008 $_{0.063}$	-0.002 $_{0.039}$	-0.009 $_{0.056}$	-0.007 $_{0.263}$	-0.009 $_{0.055}$
	100	100	0.006 $_{0.041}$	0.005 $_{0.031}$	-0.006 $_{0.044}$	-0.049 $_{0.227}$	0.002 $_{0.038}$
	250	250	0.000 $_{0.020}$	-0.000 $_{0.010}$	-0.003 $_{0.024}$	-0.005 $_{0.090}$	-0.002 $_{0.015}$
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.c.	100	50	-0.002 $_{0.070}$	0.003 $_{0.039}$	-0.003 $_{0.091}$	-0.031 $_{0.311}$	-0.014 $_{0.050}$
	50	100	-0.013 $_{0.074}$	-0.006 $_{0.043}$	-0.016 $_{0.100}$	-0.002 $_{0.280}$	-0.012 $_{0.058}$
	100	100	0.007 $_{0.045}$	0.004 $_{0.032}$	-0.007 $_{0.073}$	-0.032 $_{0.234}$	-0.001 $_{0.042}$
	250	250	-0.001 $_{0.021}$	-0.000 $_{0.011}$	-0.005 $_{0.047}$	-0.010 $_{0.093}$	-0.001 $_{0.015}$
A.3.d.	100	50	-0.003 $_{0.076}$	0.002 $_{0.037}$	-0.004 $_{0.094}$	-0.064 $_{0.131}$	0.002 $_{0.046}$
	50	100	-0.008 $_{0.078}$	-0.005 $_{0.045}$	-0.022 $_{0.104}$	-0.044 $_{0.095}$	0.001 $_{0.043}$
	100	100	0.005 $_{0.048}$	0.004 $_{0.032}$	-0.010 $_{0.074}$	-0.036 $_{0.080}$	-0.001 $_{0.038}$
	250	250	-0.000 $_{0.020}$	-0.001 $_{0.010}$	-0.006 $_{0.043}$	-0.009 $_{0.037}$	-0.001 $_{0.015}$
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 $_{0.013}$	-0.021 $_{0.138}$	-0.019 $_{0.112}$	-0.000 $_{0.018}$
A.3.f.	100	50	-0.001 $_{0.102}$	0.008 $_{0.049}$	-0.018 $_{0.273}$	-0.126 $_{0.216}$	0.009 $_{0.069}$
	50	100	-0.008 $_{0.104}$	-0.005 $_{0.050}$	-0.055 $_{0.301}$	-0.091 $_{0.144}$	0.005 $_{0.083}$
	100	100	-0.000 $_{0.064}$	0.002 $_{0.039}$	-0.032 $_{0.245}$	-0.035 $_{0.089}$	-0.005 $_{0.054}$
	250	250	0.001 $_{0.030}$	-0.000 $_{0.012}$	-0.026 $_{0.141}$	-0.014 $_{0.041}$	-0.000 $_{0.018}$

Table 4: Simulation results for Binary dynamic panel data models.

	N	T	γ	β	σ_μ	h	σ_η
B.1.a.	100	50	$0.000_{0.010}$	$0.001_{0.018}$	$-0.001_{0.041}$		
	50	100	$-0.000_{0.012}$	$-0.001_{0.017}$	$-0.007_{0.050}$		
	100	100	$0.000_{0.007}$	$0.002_{0.014}$	$0.002_{0.043}$		
	250	250	$-0.000_{0.003}$	$-0.001_{0.005}$	$0.000_{0.020}$		
B.1.b.	100	50	$0.001_{0.012}$	$0.003_{0.018}$	$0.000_{0.077}$		
	50	100	$-0.001_{0.014}$	$-0.001_{0.018}$	$-0.014_{0.097}$		
	100	100	$0.000_{0.008}$	$0.003_{0.013}$	$0.001_{0.082}$		
	250	250	$-0.000_{0.004}$	$-0.000_{0.006}$	$0.000_{0.039}$		
B.1.c.	100	50	$0.002_{0.019}$	$0.003_{0.026}$	$-0.001_{0.231}$		
	50	100	$-0.000_{0.016}$	$-0.002_{0.023}$	$-0.042_{0.309}$		
	100	100	$-0.000_{0.011}$	$0.001_{0.017}$	$0.010_{0.259}$		
	250	250	$-0.001_{0.005}$	$-0.000_{0.008}$	$0.001_{0.122}$		
B.2.a.	100	50	$-0.001_{0.010}$	$0.000_{0.017}$		$-0.045_{0.168}$	$-0.006_{0.027}$
	50	100	$-0.001_{0.009}$	$0.001_{0.018}$		$-0.036_{0.146}$	$-0.002_{0.025}$
	100	100	$0.001_{0.006}$	$0.001_{0.015}$		$-0.032_{0.123}$	$-0.000_{0.021}$
	250	250	$-0.000_{0.003}$	$-0.001_{0.005}$		$0.001_{0.061}$	$-0.000_{0.009}$
B.2.b.	100	50	$-0.001_{0.012}$	$-0.001_{0.018}$		$-0.032_{0.079}$	$-0.003_{0.029}$
	50	100	$-0.002_{0.010}$	$0.002_{0.018}$		$-0.023_{0.068}$	$-0.002_{0.024}$
	100	100	$0.001_{0.008}$	$0.002_{0.014}$		$-0.025_{0.053}$	$0.001_{0.022}$
	250	250	$-0.000_{0.003}$	$-0.001_{0.005}$		$-0.007_{0.026}$	$0.001_{0.010}$
B.3.a.	100	50	$0.002_{0.012}$	$0.001_{0.020}$	$-0.008_{0.040}$	$-0.025_{0.178}$	$-0.005_{0.027}$
	50	100	$0.002_{0.011}$	$0.002_{0.023}$	$-0.005_{0.047}$	$-0.031_{0.164}$	$-0.000_{0.025}$
	100	100	$0.000_{0.008}$	$0.001_{0.014}$	$0.002_{0.038}$	$-0.034_{0.129}$	$-0.000_{0.020}$
	250	250	$0.000_{0.003}$	$0.001_{0.006}$	$-0.000_{0.024}$	$0.002_{0.065}$	$-0.001_{0.009}$
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	$0.000_{0.003}$	$0.001_{0.005}$	$-0.001_{0.024}$	$-0.007_{0.031}$	$-0.001_{0.010}$
B.3.c.	100	50	$0.002_{0.014}$	$0.000_{0.022}$	$-0.011_{0.073}$	$-0.029_{0.179}$	$-0.004_{0.029}$
	50	100	$0.003_{0.014}$	$0.001_{0.022}$	$-0.010_{0.098}$	$-0.045_{0.167}$	$-0.001_{0.026}$
	100	100	$0.001_{0.009}$	$0.000_{0.015}$	$0.003_{0.076}$	$-0.038_{0.131}$	$-0.003_{0.021}$
	250	250	$0.000_{0.003}$	$0.000_{0.006}$	$-0.002_{0.048}$	$0.003_{0.066}$	$-0.001_{0.010}$
B.3.d.	100	50	$0.003_{0.015}$	$0.001_{0.023}$	$-0.014_{0.072}$	$-0.048_{0.101}$	$-0.001_{0.031}$
	50	100	$0.002_{0.014}$	$0.001_{0.020}$	$-0.017_{0.095}$	$-0.025_{0.063}$	$0.001_{0.028}$
	100	100	$0.000_{0.009}$	$0.001_{0.015}$	$0.001_{0.075}$	$-0.028_{0.064}$	$-0.002_{0.022}$
	250	250	$0.000_{0.004}$	$0.000_{0.006}$	$-0.003_{0.045}$	$-0.009_{0.033}$	$-0.001_{0.010}$
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	$0.000_{0.004}$	$0.000_{0.007}$	$-0.008_{0.142}$	$0.000_{0.073}$	$-0.002_{0.010}$
B.3.f.	100	50	$0.003_{0.017}$	$0.004_{0.028}$	$-0.021_{0.225}$	$-0.068_{0.107}$	$0.002_{0.039}$
	50	100	$0.005_{0.020}$	$0.001_{0.026}$	$-0.031_{0.304}$	$-0.040_{0.071}$	$0.002_{0.033}$
	100	100	$0.002_{0.011}$	$0.000_{0.020}$	$-0.000_{0.215}$	$-0.047_{0.080}$	$-0.002_{0.027}$
	250	250	$0.000_{0.005}$	$0.001_{0.007}$	$-0.007_{0.142}$	$-0.012_{0.034}$	$-0.001_{0.010}$

Table 5: Simulation results for Binomial dynamic panel data models.

	N	T	γ	β	σ_μ	h	σ_η	ν
C.1.a.	100	50	$-0.001_{0.008}$	$0.003_{0.008}$	$0.012_{0.035}$			$0.028_{0.034}$
	50	100	$-0.001_{0.006}$	$-0.004_{0.012}$	$0.011_{0.054}$			$0.018_{0.050}$
	100	100	$0.000_{0.007}$	$0.001_{0.007}$	$-0.002_{0.030}$			$0.014_{0.027}$
	250	250	$0.001_{0.001}$	$0.001_{0.002}$	$0.004_{0.016}$			$-0.002_{0.008}$
	100	50	$0.000_{0.008}$	$0.003_{0.008}$	$0.040_{0.076}$			$0.028_{0.034}$
C.1.b.	50	100	$-0.001_{0.006}$	$-0.004_{0.012}$	$0.024_{0.104}$			$0.018_{0.051}$
	100	100	$0.000_{0.007}$	$0.001_{0.007}$	$-0.002_{0.054}$			$0.014_{0.027}$
	250	250	$0.001_{0.001}$	$0.001_{0.002}$	$0.007_{0.032}$			$-0.002_{0.008}$
C.1.c	100	50	$0.000_{0.008}$	$0.003_{0.008}$	$0.122_{0.224}$			$0.028_{0.035}$
	50	100	$-0.001_{0.006}$	$-0.004_{0.012}$	$0.075_{0.305}$			$0.018_{0.051}$
	100	100	$-0.000_{0.006}$	$0.001_{0.007}$	$-0.000_{0.153}$			$0.014_{0.027}$
	250	250	$0.001_{0.001}$	$0.001_{0.002}$	$0.021_{0.096}$			$-0.002_{0.008}$
C.2.a.	100	50	$0.001_{0.005}$	$-0.004_{0.013}$		$0.030_{0.122}$	$0.003_{0.025}$	$0.007_{0.045}$
	50	100	$-0.001_{0.008}$	$0.004_{0.007}$		$-0.040_{0.138}$	$0.005_{0.013}$	$0.025_{0.033}$
	100	100	$-0.000_{0.007}$	$-0.002_{0.007}$		$-0.030_{0.093}$	$-0.003_{0.014}$	$0.011_{0.032}$
	250	250	$0.001_{0.001}$	$0.001_{0.002}$		$-0.005_{0.051}$	$0.000_{0.007}$	$-0.004_{0.007}$
C.2.b.	100	50	$0.001_{0.006}$	$-0.003_{0.014}$		$-0.007_{0.036}$	$0.003_{0.025}$	$0.010_{0.044}$
	50	100	$-0.001_{0.008}$	$0.004_{0.007}$		$-0.019_{0.044}$	$0.005_{0.014}$	$0.024_{0.033}$
	100	100	$-0.000_{0.007}$	$-0.002_{0.007}$		$-0.021_{0.043}$	$-0.004_{0.013}$	$0.011_{0.031}$
	250	250	$0.001_{0.001}$	$0.001_{0.002}$		$-0.003_{0.031}$	$-0.000_{0.006}$	$-0.004_{0.007}$
C.3.a.	100	50	$0.003_{0.006}$	$-0.000_{0.013}$	$-0.002_{0.050}$	$-0.046_{0.211}$	$-0.006_{0.029}$	$0.024_{0.045}$
	50	100	$0.002_{0.005}$	$-0.001_{0.013}$	$0.003_{0.075}$	$-0.032_{0.117}$	$-0.003_{0.017}$	$0.021_{0.043}$
	100	100	$-0.001_{0.006}$	$0.003_{0.008}$	$0.000_{0.030}$	$0.010_{0.067}$	$-0.009_{0.017}$	$0.018_{0.031}$
	250	250	$0.000_{0.001}$	$0.001_{0.003}$	$0.004_{0.015}$	$0.014_{0.062}$	$-0.002_{0.007}$	$-0.003_{0.008}$
C.3.b.	100	50	$0.003_{0.006}$	$-0.000_{0.013}$	$0.006_{0.046}$	$-0.088_{0.155}$	$-0.002_{0.030}$	$0.017_{0.046}$
	50	100	$0.002_{0.005}$	$-0.001_{0.012}$	$-0.003_{0.073}$	$-0.044_{0.084}$	$-0.003_{0.013}$	$0.019_{0.042}$
	100	100	$-0.001_{0.006}$	$0.003_{0.008}$	$-0.001_{0.030}$	$-0.019_{0.042}$	$-0.008_{0.016}$	$0.017_{0.031}$
	250	250	$0.000_{0.001}$	$0.001_{0.003}$	$0.004_{0.015}$	$-0.008_{0.030}$	$-0.001_{0.008}$	$-0.003_{0.008}$
C.3.c.	100	50	$0.003_{0.006}$	$-0.000_{0.013}$	$0.020_{0.093}$	$-0.041_{0.209}$	$-0.006_{0.029}$	$0.017_{0.046}$
	50	100	$0.002_{0.005}$	$-0.001_{0.013}$	$0.004_{0.148}$	$-0.032_{0.114}$	$-0.003_{0.017}$	$0.021_{0.043}$
	100	100	$-0.001_{0.006}$	$0.003_{0.008}$	$0.003_{0.056}$	$0.011_{0.064}$	$-0.009_{0.017}$	$0.018_{0.031}$
	250	250	$0.000_{0.001}$	$0.001_{0.003}$	$0.009_{0.032}$	$0.013_{0.061}$	$-0.002_{0.007}$	$-0.003_{0.008}$
C.3.d.	100	50	$0.003_{0.006}$	$-0.000_{0.013}$	$0.017_{0.095}$	$-0.089_{0.154}$	$-0.002_{0.030}$	$0.017_{0.046}$
	50	100	$0.002_{0.005}$	$-0.001_{0.012}$	$-0.006_{0.140}$	$-0.045_{0.083}$	$-0.003_{0.013}$	$0.020_{0.043}$
	100	100	$-0.001_{0.006}$	$0.002_{0.008}$	$0.002_{0.055}$	$-0.025_{0.045}$	$-0.008_{0.017}$	$0.017_{0.031}$
	250	250	$0.000_{0.001}$	$0.001_{0.003}$	$0.008_{0.024}$	$-0.009_{0.031}$	$-0.001_{0.008}$	$-0.003_{0.008}$
C.3.e.	100	50	$0.004_{0.006}$	$-0.000_{0.013}$	$0.069_{0.285}$	$-0.044_{0.207}$	$-0.006_{0.029}$	$0.017_{0.046}$
	50	100	$0.002_{0.004}$	$-0.001_{0.013}$	$0.008_{0.431}$	$-0.033_{0.117}$	$-0.003_{0.017}$	$0.021_{0.043}$
	100	100	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.300_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$
	250	250	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.300_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$
C.3.f.	100	50	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.900_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$
	50	100	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.900_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$
	100	100	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.900_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$
	250	250	$-0.200_{0.000}$	$-1.000_{0.000}$	$-3.000_{0.000}$	$-0.900_{0.000}$	$-0.200_{0.000}$	$-3.000_{0.000}$

Table 6: Simulation results for Student's dynamic panel data models with $\nu = 3$.

	N	T	γ	β	σ_μ	h	σ_η	ν
C.1.a.	100	50	0.002 $_{0.010}$	-0.002 $_{0.010}$	-0.014 $_{0.047}$			0.286 $_{0.347}$
	50	100	-0.004 $_{0.008}$	-0.001 $_{0.014}$	0.004 $_{0.045}$			0.135 $_{0.274}$
	100	100	-0.001 $_{0.005}$	-0.003 $_{0.010}$	-0.010 $_{0.036}$			0.140 $_{0.159}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.003}$	0.002 $_{0.022}$			0.028 $_{0.067}$
C.1.b.	100	50	-0.001 $_{0.010}$	-0.001 $_{0.010}$	-0.007 $_{0.105}$			0.286 $_{0.348}$
	50	100	-0.004 $_{0.008}$	-0.001 $_{0.014}$	0.008 $_{0.091}$			0.136 $_{0.275}$
	100	100	-0.001 $_{0.005}$	-0.003 $_{0.010}$	-0.012 $_{0.067}$			0.141 $_{0.159}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.003}$	0.007 $_{0.044}$			0.026 $_{0.067}$
C.1.c.	100	50	-0.001 $_{0.009}$	-0.001 $_{0.010}$	-0.008 $_{0.321}$			0.284 $_{0.349}$
	50	100	-0.004 $_{0.008}$	-0.001 $_{0.014}$	0.021 $_{0.279}$			0.138 $_{0.274}$
	100	100	-0.001 $_{0.004}$	-0.003 $_{0.010}$	-0.020 $_{0.195}$			0.140 $_{0.159}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.003}$	0.019 $_{0.131}$			0.029 $_{0.067}$
C.2.a.	100	50	-0.003 $_{0.008}$	-0.001 $_{0.013}$		-0.025 $_{0.157}$	0.004 $_{0.021}$	0.059 $_{0.234}$
	50	100	0.002 $_{0.009}$	-0.003 $_{0.012}$		-0.031 $_{0.100}$	0.001 $_{0.024}$	0.209 $_{0.295}$
	100	100	-0.001 $_{0.005}$	-0.003 $_{0.009}$		0.095 $_{0.061}$	-0.004 $_{0.024}$	0.070 $_{0.134}$
	250	250	0.000 $_{0.003}$	0.000 $_{0.003}$		-0.006 $_{0.075}$	-0.002 $_{0.010}$	0.013 $_{0.063}$
C.2.b.	100	50	-0.004 $_{0.009}$	0.002 $_{0.014}$		-0.027 $_{0.058}$	0.004 $_{0.025}$	0.068 $_{0.232}$
	50	100	0.002 $_{0.008}$	-0.002 $_{0.012}$		-0.028 $_{0.052}$	0.003 $_{0.027}$	0.205 $_{0.295}$
	100	100	-0.001 $_{0.005}$	-0.003 $_{0.009}$		0.015 $_{0.037}$	-0.001 $_{0.020}$	0.069 $_{0.133}$
	250	250	0.000 $_{0.003}$	0.000 $_{0.003}$		0.000 $_{0.021}$	-0.003 $_{0.010}$	0.012 $_{0.065}$
C.3.a.	100	50	-0.005 $_{0.008}$	0.002 $_{0.007}$	-0.009 $_{0.037}$	-0.069 $_{0.178}$	0.003 $_{0.027}$	0.083 $_{0.306}$
	50	100	-0.001 $_{0.011}$	0.003 $_{0.007}$	0.000 $_{0.056}$	-0.073 $_{0.175}$	-0.008 $_{0.017}$	0.079 $_{0.282}$
	100	100	-0.001 $_{0.005}$	-0.000 $_{0.009}$	-0.004 $_{0.034}$	-0.044 $_{0.069}$	0.009 $_{0.019}$	0.080 $_{0.138}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.006 $_{0.017}$	-0.003 $_{0.046}$	-0.002 $_{0.010}$	0.015 $_{0.062}$
C.3.b.	100	50	-0.006 $_{0.008}$	0.001 $_{0.008}$	-0.008 $_{0.037}$	-0.036 $_{0.076}$	0.008 $_{0.025}$	0.112 $_{0.299}$
	50	100	-0.001 $_{0.010}$	0.003 $_{0.007}$	-0.004 $_{0.056}$	-0.003 $_{0.051}$	-0.005 $_{0.019}$	0.080 $_{0.283}$
	100	100	-0.001 $_{0.005}$	-0.000 $_{0.010}$	-0.006 $_{0.034}$	-0.041 $_{0.076}$	0.011 $_{0.022}$	0.081 $_{0.138}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.005 $_{0.017}$	-0.005 $_{0.028}$	-0.001 $_{0.011}$	0.015 $_{0.063}$
C.3.c.	100	50	-0.006 $_{0.008}$	0.001 $_{0.008}$	-0.011 $_{0.066}$	-0.064 $_{0.194}$	0.004 $_{0.027}$	0.115 $_{0.304}$
	50	100	-0.001 $_{0.011}$	0.003 $_{0.007}$	0.003 $_{0.115}$	-0.076 $_{0.172}$	-0.008 $_{0.017}$	0.081 $_{0.283}$
	100	100	-0.001 $_{0.005}$	-0.000 $_{0.009}$	-0.003 $_{0.064}$	-0.048 $_{0.066}$	0.008 $_{0.019}$	0.082 $_{0.138}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.012 $_{0.036}$	-0.004 $_{0.046}$	-0.002 $_{0.010}$	0.015 $_{0.062}$
C.3.d.	100	50	-0.006 $_{0.008}$	0.001 $_{0.008}$	-0.016 $_{0.067}$	-0.034 $_{0.076}$	0.008 $_{0.025}$	0.114 $_{0.300}$
	50	100	-0.001 $_{0.010}$	0.002 $_{0.007}$	-0.005 $_{0.115}$	-0.008 $_{0.065}$	-0.005 $_{0.019}$	0.082 $_{0.285}$
	100	100	-0.001 $_{0.005}$	-0.000 $_{0.010}$	-0.007 $_{0.063}$	-0.049 $_{0.075}$	0.010 $_{0.022}$	0.082 $_{0.138}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.011 $_{0.035}$	-0.006 $_{0.028}$	-0.001 $_{0.011}$	0.015 $_{0.063}$
C.3.e.	100	50	-0.006 $_{0.008}$	0.001 $_{0.008}$	-0.032 $_{0.188}$	-0.072 $_{0.189}$	0.004 $_{0.027}$	0.114 $_{0.305}$
	50	100	-0.001 $_{0.010}$	0.003 $_{0.007}$	0.011 $_{0.358}$	-0.078 $_{0.171}$	-0.008 $_{0.017}$	0.083 $_{0.286}$
	100	100	-0.002 $_{0.005}$	-0.000 $_{0.009}$	0.000 $_{0.182}$	-0.048 $_{0.066}$	0.008 $_{0.019}$	0.083 $_{0.140}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.034 $_{0.108}$	-0.005 $_{0.045}$	-0.002 $_{0.010}$	0.017 $_{0.064}$
C.3.f.	100	50	-0.006 $_{0.007}$	0.001 $_{0.008}$	-0.045 $_{0.189}$	-0.044 $_{0.081}$	0.008 $_{0.025}$	0.113 $_{0.301}$
	50	100	-0.001 $_{0.010}$	0.002 $_{0.007}$	-0.004 $_{0.361}$	-0.012 $_{0.067}$	-0.005 $_{0.019}$	0.084 $_{0.286}$
	100	100	-0.002 $_{0.005}$	-0.000 $_{0.010}$	-0.006 $_{0.180}$	-0.060 $_{0.075}$	0.010 $_{0.021}$	0.082 $_{0.139}$
	250	250	-0.000 $_{0.002}$	-0.000 $_{0.005}$	0.031 $_{0.106}$	-0.009 $_{0.027}$	-0.001 $_{0.011}$	0.015 $_{0.064}$

Table 7: Simulation results for Student's dynamic panel data models with $\nu = 5$.

	N	T	γ	β	σ_μ	h	σ_η	ν
C.1.a.	100	50	-0.005 $_{0.011}$	0.001 $_{0.015}$	-0.016 $_{0.042}$			1.053 $_{1.085}$
	50	100	-0.005 $_{0.009}$	0.006 $_{0.012}$	0.000 $_{0.043}$			0.130 $_{0.982}$
	100	100	-0.001 $_{0.006}$	-0.004 $_{0.008}$	-0.011 $_{0.034}$			1.025 $_{0.527}$
	250	250	0.000 $_{0.003}$	0.001 $_{0.004}$	0.006 $_{0.024}$			0.241 $_{0.295}$
C.1.b.	100	50	-0.006 $_{0.011}$	0.001 $_{0.015}$	-0.013 $_{0.078}$			1.145 $_{1.048}$
	50	100	-0.004 $_{0.009}$	0.006 $_{0.012}$	-0.008 $_{0.073}$			0.141 $_{0.992}$
	100	100	-0.001 $_{0.006}$	-0.004 $_{0.008}$	-0.025 $_{0.068}$			1.036 $_{0.524}$
	250	250	0.000 $_{0.003}$	0.001 $_{0.004}$	0.011 $_{0.048}$			0.242 $_{0.296}$
C.1.c.	100	50	-0.005 $_{0.011}$	0.001 $_{0.015}$	-0.031 $_{0.221}$			1.177 $_{1.057}$
	50	100	-0.004 $_{0.009}$	0.006 $_{0.012}$	-0.045 $_{0.196}$			0.145 $_{0.999}$
	100	100	0.000 $_{0.006}$	-0.004 $_{0.008}$	-0.085 $_{0.209}$			1.041 $_{0.523}$
	250	250	0.000 $_{0.003}$	0.001 $_{0.004}$	0.032 $_{0.142}$			0.243 $_{0.297}$
C.2.a.	100	50	-0.001 $_{0.008}$	0.007 $_{0.012}$		-0.039 $_{0.194}$	-0.009 $_{0.021}$	-0.164 $_{0.806}$
	50	100	-0.006 $_{0.011}$	0.001 $_{0.015}$		-0.048 $_{0.101}$	-0.002 $_{0.021}$	0.684 $_{0.959}$
	100	100	-0.002 $_{0.006}$	-0.002 $_{0.009}$		-0.009 $_{0.101}$	-0.004 $_{0.021}$	0.560 $_{0.744}$
	250	250	0.000 $_{0.003}$	0.000 $_{0.004}$		-0.021 $_{0.057}$	0.001 $_{0.009}$	0.137 $_{0.306}$
C.2.b.	100	50	-0.002 $_{0.009}$	0.009 $_{0.012}$		-0.035 $_{0.087}$	-0.007 $_{0.024}$	-0.295 $_{0.841}$
	50	100	-0.006 $_{0.011}$	0.001 $_{0.014}$		0.000 $_{0.026}$	-0.001 $_{0.021}$	0.658 $_{0.963}$
	100	100	-0.002 $_{0.006}$	-0.002 $_{0.009}$		-0.009 $_{0.047}$	-0.003 $_{0.022}$	0.575 $_{0.744}$
	250	250	0.000 $_{0.003}$	0.000 $_{0.004}$		-0.014 $_{0.033}$	-0.001 $_{0.009}$	0.136 $_{0.303}$
C.3.a.	100	50	-0.007 $_{0.013}$	0.001 $_{0.018}$	0.013 $_{0.040}$	0.025 $_{0.112}$	0.010 $_{0.019}$	0.290 $_{1.058}$
	50	100	-0.008 $_{0.012}$	0.001 $_{0.018}$	0.011 $_{0.055}$	-0.039 $_{0.074}$	0.009 $_{0.021}$	0.210 $_{1.036}$
	100	100	-0.000 $_{0.005}$	0.005 $_{0.006}$	-0.006 $_{0.038}$	-0.041 $_{0.155}$	-0.005 $_{0.015}$	0.700 $_{0.609}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.001 $_{0.023}$	0.007 $_{0.075}$	-0.000 $_{0.011}$	0.132 $_{0.297}$
C.3.b.	100	50	-0.008 $_{0.013}$	0.000 $_{0.018}$	0.019 $_{0.038}$	-0.024 $_{0.059}$	0.006 $_{0.029}$	0.194 $_{1.069}$
	50	100	-0.008 $_{0.012}$	0.001 $_{0.018}$	0.003 $_{0.057}$	-0.035 $_{0.058}$	0.009 $_{0.025}$	0.214 $_{1.040}$
	100	100	-0.000 $_{0.004}$	0.005 $_{0.006}$	-0.009 $_{0.037}$	-0.015 $_{0.030}$	-0.003 $_{0.009}$	0.686 $_{0.600}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.000 $_{0.024}$	-0.015 $_{0.033}$	0.001 $_{0.011}$	0.132 $_{0.298}$
C.3.c.	100	50	-0.008 $_{0.013}$	0.001 $_{0.018}$	0.042 $_{0.072}$	-0.018 $_{0.095}$	0.008 $_{0.021}$	0.217 $_{1.079}$
	50	100	-0.008 $_{0.012}$	0.001 $_{0.018}$	0.021 $_{0.102}$	-0.044 $_{0.076}$	0.009 $_{0.021}$	0.211 $_{1.028}$
	100	100	0.000 $_{0.005}$	0.005 $_{0.006}$	-0.014 $_{0.075}$	-0.045 $_{0.152}$	-0.006 $_{0.015}$	0.706 $_{0.608}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.003 $_{0.043}$	0.004 $_{0.075}$	-0.000 $_{0.010}$	0.133 $_{0.298}$
C.3.d.	100	50	-0.007 $_{0.013}$	0.000 $_{0.018}$	0.035 $_{0.072}$	-0.025 $_{0.062}$	0.006 $_{0.029}$	0.215 $_{1.073}$
	50	100	-0.008 $_{0.012}$	0.001 $_{0.018}$	0.010 $_{0.109}$	-0.036 $_{0.058}$	0.009 $_{0.025}$	0.216 $_{1.033}$
	100	100	0.000 $_{0.004}$	0.005 $_{0.006}$	-0.019 $_{0.075}$	-0.017 $_{0.030}$	-0.003 $_{0.009}$	0.701 $_{0.589}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.000 $_{0.046}$	-0.014 $_{0.033}$	0.001 $_{0.011}$	0.133 $_{0.299}$
C.3.e.	100	50	-0.006 $_{0.012}$	0.001 $_{0.018}$	0.117 $_{0.213}$	-0.026 $_{0.093}$	0.007 $_{0.021}$	0.229 $_{1.076}$
	50	100	-0.008 $_{0.011}$	0.001 $_{0.018}$	0.056 $_{0.296}$	-0.045 $_{0.076}$	0.008 $_{0.021}$	0.210 $_{1.023}$
	100	100	0.001 $_{0.005}$	0.005 $_{0.006}$	-0.051 $_{0.227}$	-0.048 $_{0.151}$	-0.006 $_{0.015}$	0.709 $_{0.607}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.011 $_{0.126}$	0.001 $_{0.076}$	-0.001 $_{0.010}$	0.133 $_{0.299}$
C.3.f.	100	50	-0.006 $_{0.011}$	0.000 $_{0.018}$	0.107 $_{0.207}$	-0.052 $_{0.094}$	0.005 $_{0.029}$	0.224 $_{1.070}$
	50	100	-0.008 $_{0.011}$	0.001 $_{0.018}$	0.057 $_{0.300}$	-0.055 $_{0.053}$	0.010 $_{0.025}$	0.210 $_{1.028}$
	100	100	0.001 $_{0.004}$	0.005 $_{0.006}$	-0.052 $_{0.229}$	-0.033 $_{0.028}$	-0.003 $_{0.009}$	0.705 $_{0.586}$
	250	250	0.000 $_{0.003}$	-0.000 $_{0.002}$	0.002 $_{0.129}$	-0.017 $_{0.034}$	0.001 $_{0.011}$	0.127 $_{0.297}$

Table 8: Simulation results for Student's dynamic panel data models with $\nu = 10$.

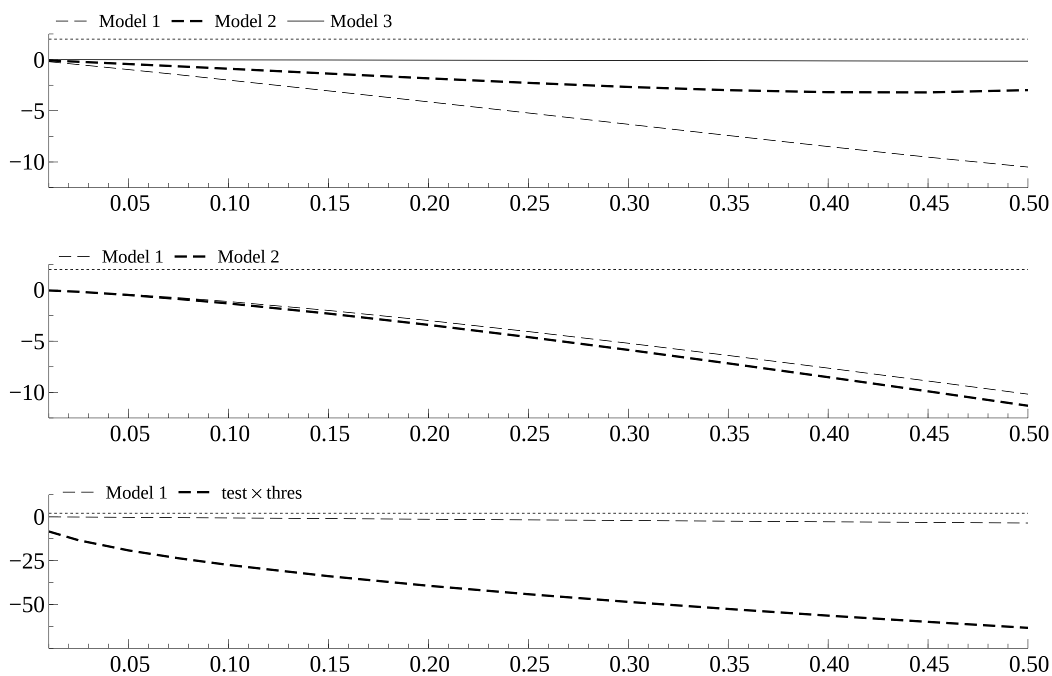

Figure 10: Diagnostic checks for the empirical studies. The t-test statistics correspond to the estimated models presented for: (i) the union choice of male, (ii) the crime rates of families and (iii) economic growth rates. All tests are implemented as discussed in Section 2.1.

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.

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. (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*, 2 edn, Oxford University Press, Oxford.

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.

Koopman, S. J. & Durbin, J. (2003), 'Filtering and smoothing of of state vector for diffuse state space models', *Journal of Time Series Analysis* **24**, 85–98.

Mesters, G. & Koopman, S. J. (2014), 'Generalized Dynamic Panel Data Models with Random Effects for Cross-Section and Time'. Tinbergen Institute Working paper TI 2012-009/4.

Nocedal, J. & Wright, S. J. (1999), *Numerical Optimization*, Springer Verlag, 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.