I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13214

Robust Dynamic Panel Data Models Using ε-Contamination

Badi H. Baltagi
Georges Bresson
Anoop Chaturvedi
Guy Lacroix

MAY 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13214

Robust Dynamic Panel Data Models Using ε-Contamination

Badi H. Baltagi
Syracuse University and IZA

Georges Bresson
Université Paris II

Anoop Chaturvedi
University of Allahabad

Guy Lacroix
Université Laval and IZA

MAY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Robust Dynamic Panel Data Models Using ε-Contamination*

This paper extends the work of Baltagi et al. (2018) to the popular dynamic panel data model. We investigate the robustness of Bayesian panel data models to possible misspecification of the prior distribution. The proposed robust Bayesian approach departs from the standard Bayesian framework in two ways. First, we consider the ε-contamination class of prior distributions for the model parameters as well as for the individual effects. Second, both the base elicited priors and the ε-contamination priors use Zellner (1986)'s g-priors for the variance-covariance matrices. We propose a general "toolbox" for a wide range of specifications which includes the dynamic panel model with random effects, with cross-correlated effects *à la* Chamberlain, for the Hausman-Taylor world and for dynamic panel data models with homogeneous/heterogeneous slopes and cross-sectional dependence. Using a Monte Carlo simulation study, we compare the finite sample properties of our proposed estimator to those of standard classical estimators. The paper contributes to the dynamic panel data literature by proposing a general robust Bayesian framework which encompasses the conventional frequentist specifications and their associated estimation methods as special cases.

JEL Classification: C11, C23, C26

Keywords: dynamic model, ε-contamination, g-priors, type-II maximum likelihood posterior density, panel data, robust Bayesian estimator, two-stage hierarchy

Corresponding author:
Badi H. Baltagi
Department of Economics and Center for Policy Research
Syracuse University
426 Eggers Hall
Syracuse, NY, 13244-1020
USA

E-mail: bbaltagi@maxwell.syr.edu

* We would like to thank Jean-Michel Etienne for help and support with Stata Mata codes. The usual disclaimers apply.

1. Introduction

The dynamic panel data model allows for feedback from lagged endogenous values and have been used in many empirical studies. The most popular estimation method is the generalized method of moments (GMM) with many variants, the best known being the Arellano-Bond difference GMM (Arellano and Bond (1991)) and the Blundell-Bond system GMM (Blundell and Bond (1998)) (see the surveys by Harris et al. (2008) and Bun and Sarafidis (2015) to mention a few). Despite its optimal asymptotic properties, Bun and Sarafidis (2015) and Moon et al. (2015), among others, argue that the finite sample behavior of the GMM estimator can be poor due to weakness and/or abundance of moment conditions and dependence on crucial nuisance parameters. Several alternative inference methods derived from inconsistent least squares (LS) or likelihood based procedures have been proposed. These include modifications of the profile likelihood (Lancaster (2002), Dhaene and Jochmans (2011, 2016)) or estimation methods based on the likelihood function of the first differences (Hsiao et al. (2002), Binder et al. (2005), Hayakawa and Pesaran (2015)).

While GMM estimation is very attractive because of its flexibility, other promising methods remain underrepresented in empirical work. Examples are bias-correction procedures for the fixed-effects dynamic panel estimator proposed by Kiviet (1995), Bun (2003), Bun and Kiviet (2003), Everaert and Pozzi (2007), and Everaert (2013) among others. Estimation of dynamic panel data models with heterogeneous slopes and/or cross-sectional dependence has also been investigated by Chudik and Pesaran (2015a,b), using the common correlated effects (CCE) approach of Pesaran (2006), and by Moon and Weidner (2015, 2017), who studied linear models with interactive fixed effects.

Quasi-maximum likelihood (QML) methods have been also proposed to circumvent this bias by modeling the unconditional likelihood function instead of conditioning on the initial observations. While this requires additional assumptions on the marginal distribution of the initial observations, the QML estimators are an attractive alternative to other estimation approaches in terms of efficiency and finite-sample performance if all the assumptions are satisfied. QML estimators can be characterized as limited-information maximum likelihood estimators that are special cases of a structural equation modeling or full information maximum-likelihood approach with many cross-equation restrictions[1]. For dynamic models with random effects, we must be explicit about the non-zero correlation between the individual-specific effects and the initial conditions (see Anderson and Hsiao (1982), Bhargava and Sargan (1983), Alvarez and Arellano (2003), Hsiao and Pesaran (2008), Moral-Benito (2012, 2013), Kripfganz (2016), Bun et al. (2017), Moral-Benito et al. (2019)).[2]

The widely used difference GMM estimator suffers from finite sample bias when the number of cross-section observations is small. Moreover, some have expressed concern in recent years that many instrumental variables of the type considered in panel GMM estimators such as Arellano and Bond (1991) may be invalid, weak or both (see for instance Bazzi and Clemens (2013) and Kraay (2015)). Based on the same identifying assumption, some alternatives have been proposed in the literature (e.g. Ahn and Schmidt (1995), Hansen et al. (1996), Hsiao et al. (2002), Moral-

[1] For the dynamic fixed-effects model, see for instance Hsiao et al. (2002).

[2] In a Gaussian dynamic linear mixed model: $y_{it} = \rho y_{it-1} + X'_{it}\beta + W'_{it}b_i + u_{it}$, $i = 1, ..., N$, $t = 2, ..., T$, as in our case (see eq.(8) in section (2)), maximum likelihood analysis is subject to an initial condition problem if the permanent subject effects b_i and the initial observations are correlated. In case of such correlation, possible options are a joint random prior (*e.g.*, bivariate normal) involving b_i and the first disturbance u_{i1} (Dorsett (1999)), or a prior for b_i that is conditional on y_{i1}, such as $b_i \mid y_{i1} \sim N(\varphi y_{i1}, \sigma_1^2)$ (see Hirano (2002), Wooldridge (2005)) and Congdon (2010)).

Benito (2013), to mention a few). Maximum likelihood estimators, asymptotically equivalent to the Arellano and Bond (1991) estimator, have recently been proposed and are strongly preferred in terms of finite sample performance (Moral-Benito et al. (2019)).

Bayesian analysis for dynamic panel data models have also been proposed (see for instance Hsiao et al. (1999), Lancaster (2002), Hsiao and Pesaran (2008), Koop et al. (2008), Juárez and Steel (2010), Tsai (2016), Liu et al. (2017), Liu et al. (2018), Bretó et al. (2019), Pacifico (2019)). Some consider that the process which generates the initial observation y_{i0} of the dependent variable for each individual i has started a long time ago (*e.g.*, Juárez and Steel (2010). Others derive the estimators under the assumption that y_{i0} are fixed constants (*e.g.*, Hsiao et al. (1999), Hsiao and Pesaran (2008)). Yet others consider that the initial value is generated from the finite past using state space forms (*e.g.*, Liu et al. (2017)), or use the Prais-Winsten transformation for the initial period. A simplifying approach, more feasible for large T, is to condition on the first observation in a model involving a first-order lag in y, so that y_{i1} is nonstochastic (Hjellvik and Tjstheim (1999), Bauwens et al. (2005)). Geweke and Keane (2000) and Lancaster (2002) consider Bayesian approaches to the dynamic linear panel model in which the model for period 1 is not necessarily linked to those for subsequent periods in a way consistent with stationarity (see also Congdon (2010)).

This brief overview seems to confirm the strong comeback of ML methods and associated Bayesian approaches for dynamic panel data models. MCMC holds some advantages over ML or QML estimation. Su and Yang (2015) and Yu et al. (2008) have discussed issues involved in maximizing a concentrated version of the likelihood function that could involve trivariate optimization over the parameters and subject to stationarity restrictions. This type of constrained optimization may lead to local optima and may produce misleading inference. In our earlier paper (Baltagi et al., 2018), which considered a static panel data model, we argued that the Bayesian approach rests upon hypothesized prior distributions (and possibly on their hyperparameters). The choice of specific distributions is often made out of convenience. Yet, it is well-known that the estimators can be sensitive to misspecification of the latter. Fortunately, this difficulty can be partly circumvented by use of the robust Bayesian approach which relies upon a class of prior distributions and selects an appropriate one in a data dependent fashion. This paper extends our earlier paper to the popular dynamic panel data model and studies the robustness of Bayesian panel data models to possible misspecification of the prior distribution in the spirit of the works of Good (1965), Dempster (1977), Rubin (1977), Hill (1980), Berger (1985), Berger and Berliner (1984) and Berger and Berliner (1986). In particular, it is concerned with the posterior robustness which is different from the robustness à la White (1980). The objective of our paper is to propose a robust Bayesian approach for dynamic panel data models which departs from the standard Bayesian one in two ways. First, we consider the ε-contamination class of prior distributions for the model parameters (and for the individual effects). Second, both the base elicited priors and the ε-contamination priors use Zellner (1986)'s g-priors rather than the standard Wishart distributions for the variance-covariance matrices. We propose a general "toolbox" for a wide range of specifications such as the dynamic panel model with random effects, or with cross-correlated effects *à la* Mundlak or *à la* Chamberlain, for the Hausman-Taylor world or for dynamic panel data models with homogeneous/heterogeneous slopes and cross-sectional dependence. The paper contributes to the dynamic panel data literature by proposing a general robust Bayesian framework which encompasses all the above-mentioned conventional frequentist specifications and their associated estimation methods as special cases.

Section 2 gives the general framework of a robust linear dynamic panel data model using ε-contamination and derives the Type-II maximum likelihood posterior mean and the variance-covariance matrix of the coefficients in a two-stage hierarchy model. Section 3 investigates the finite sample performance of our robust Bayesian estimator through extensive Monte Carlo experiments. The simulation results underscore the relatively good performance of the two-stage hierarchy estimator as compared to the standard frequentist estimation methods. Section 4 gives our conclusion.

2. A robust linear dynamic panel data model

2.1. The static framework

Baltagi et al. (2018) considered the following Gaussian static linear mixed model:

$$y_{it} = X'_{it}\beta + W'_{it}b_i + u_{it} \ , \ i = 1, ..., N \ , \ t = 1, ..., T, \tag{1}$$

where X'_{it} is a $(1 \times K_x)$ vector of explanatory variables including the intercept, and β is a $(K_x \times 1)$ vector of parameters. t is the faster index (primal pooling). Furthermore, let W'_{it} denote a $(1 \times K_2)$ vector of covariates and b_i a $(K_2 \times 1)$ vector of parameters. The subscript i of b_i indicates that the model allows for heterogeneity on the W variables. The distribution of u_{it} is parametrized in terms of its precision τ rather than its variance $\sigma_u^2 \, (= 1/\tau)$.

Following the seminal papers of Lindley and Smith (1972) and Smith (1973), various authors including Chib and Carlin (1999), Koop (2003), Chib (2008), Greenberg (2008), Zheng et al. (2008), and Rendon (2013) have proposed a very general three-stage hierarchy framework

$$\begin{array}{rl} \text{First stage :} & y = X\beta + Wb + u, \, u \sim N(0, \Sigma), \Sigma = \tau^{-1} I_{NT} \\ \text{Second stage :} & \beta \sim N\left(\beta_0, \Lambda_\beta\right) \text{ and } b \sim N\left(b_0, \Lambda_b\right) \\ \text{Third stage :} & \Lambda_b^{-1} \sim Wish\left(\nu_b, R_b\right) \text{and } \tau \sim G(\cdot), \end{array} \tag{2}$$

where $y = (y_{1,1}, ..., y_{1,T}, ..., y_{N,1}, ..., y_{N,T})'$ is $(NT \times 1)$. X is $(NT \times K_x)$, W is $(NT \times K_2)$, u is $(NT) \times 1)$ and I_{NT} is a $(NT \times NT)$ identity matrix. The parameters depend upon hyperparameters which themselves follow random distributions. The second stage (also called *fixed effects model* in the Bayesian literature) updates the distribution of the parameters. The third stage (also called *random effects model* in the Bayesian literature) updates the distribution of the hyperparameters. The *random effects model* simply updates the distribution of the hyperparameters. The precision τ is assumed to follow a Gamma distribution and Λ_b^{-1} is assumed to follow a Wishart distribution with ν_b degrees of freedom and a hyperparameter matrix R_b which is generally chosen close to an identity matrix. In that case, the hyperparameters only concern the variance-covariance matrix of the b coefficients and the precision τ. As is well-known, Bayesian methods are sensitive to misspecification of the distributions of the priors. Conventional proper priors in the normal linear model have been based on the conjugate Normal-Gamma family because they allow closed form calculations of all marginal likelihoods. Likewise, rather than specifying a Wishart distribution for the variance-covariance matrices as is customary, Zellner's g-prior ($\Lambda_\beta = (\tau g X'X)^{-1}$ for β or $\Lambda_b = (\tau h W'W)^{-1}$ for b) has been widely adopted because of its computational efficiency in evaluating marginal likelihoods and because of its simple interpretation arising from the design matrix of observables in the sample. Since the calculation of marginal likelihoods using a mixture of g-priors involves only a one-dimensional integral, this approach provides an attractive computational solution that made the original g-priors popular while insuring robustness to misspecification of g (see Zellner (1986) and Fernández et al. (2001)).

To guard against mispecifying the distributions of the priors, Baltagi et al. (2018) considered the ε-contamination class of prior distributions for (β, b, τ):

$$\Gamma = \left\{\pi\left(\beta, b, \tau|g_0, h_0\right) = (1-\varepsilon)\,\pi_0\left(\beta, b, \tau|g_0, h_0\right) + \varepsilon q\left(\beta, b, \tau|g_0, h_0\right)\right\}, \tag{3}$$

where $\pi_0(\cdot)$ is the base elicited prior, $q(\cdot)$ is the contamination belonging to some suitable class Q of prior distributions, and $0 \le \varepsilon \le 1$ reflects the amount of error in $\pi_0(\cdot)$. τ is assumed to have a vague prior, $p(\tau) \propto \tau^{-1}$, $0 < \tau < \infty$, and $\pi_0\left(\beta, b, \tau|g_0, h_0\right)$ is the base prior assumed to be a specific g-prior with

$$\left\{\begin{array}{lcl} \beta & \sim & N\left(\beta_0\iota_{K_x}, \left(\tau g_0 \Lambda_X\right)^{-1}\right) \text{ with } \Lambda_X = X'X \\ b & \sim & N\left(b_0\iota_{K_2}, \left(\tau h_0 \Lambda_W\right)^{-1}\right) \text{ with } \Lambda_W = W'W, \end{array}\right. \tag{4}$$

where ι_{K_x} is a $(K_x \times 1)$ vector of ones. Here, β_0, b_0, g_0 and h_0 are known scalar hyperparameters of the base prior $\pi_0\left(\beta, b, \tau|g_0, h_0\right)$. The probability density function (henceforth pdf) of the base prior $\pi_0(.)$ is given by:

$$\pi_0\left(\beta, b, \tau|g_0, h_0\right) = p\left(\beta|b, \tau, \beta_0, b_0, g_0, h_0\right) \times p\left(b|\tau, b_0, h_0\right) \times p(\tau). \tag{5}$$

The possible class of contamination Q is defined as:

$$Q = \left\{\begin{array}{c} q\left(\beta, b, \tau|g_0, h_0\right) = p\left(\beta|b, \tau, \beta_q, b_q, g_q, h_q\right) \times p\left(b|\tau, b_q, h_q\right) \times p(\tau) \\ \text{with } 0 < g_q \le g_0,\ 0 < h_q \le h_0 \end{array}\right\}, \tag{6}$$

with

$$\left\{\begin{array}{lcl} \beta & \sim & N\left(\beta_q\iota_{K_x}, \left(\tau g_q \Lambda_X\right)^{-1}\right) \\ b & \sim & N\left(b_q\iota_{K_2}, \left(\tau h_q \Lambda_W\right)^{-1}\right), \end{array}\right. \tag{7}$$

where β_q, b_q, g_q and h_q are unknown. The restrictions $g_q \le g_0$ and $h_q \le h_0$ imply that the base prior is the best possible so that the precision of the base prior is greater than any prior belonging to the contamination class. The ε-contamination class of prior distributions for (β, b, τ) is then conditional on known g_0 and h_0.

Following Baltagi et al. (2018) for the static panel model, we use a two-step strategy because it simplifies the derivation of the predictive densities (or marginal likelihoods).[3] This will be extended to the dynamic panel model introduced in the next section.

2.2. The dynamic framework

This paper considers the Gaussian dynamic linear mixed model:

$$y_{it} = \rho y_{it-1} + X_{it}'\beta + W_{it}'b_i + u_{it} = Z_{it}'\theta + W_{it}'b_i + u_{it}\ ,\ i = 1, ..., N\ ,\ t = 2, ..., T, \tag{8}$$

[3]One could also use a one-step estimation of the ML-II posterior distribution. But in the one-step approach, the pdf of y and the pdf of the base prior $\pi_0\left(\beta, b, \tau|g_0, h_0\right)$ need to be combined to get the predictive density. It thus leads to a complex expression whose integration with respect to (β, b, τ) may be involved.

where $Z'_{it} = [y_{it-1}, X'_{it}]'$ and $\theta' = [\rho, \beta']'$ is a $(1 \times K_1)$ vector with $K_1 = K_x + 1$. The likelihood is conditional on the first period observations y_1. In that case, the first period is assumed exogenous and known. In the spirit of eq(2), we have the following:

$$\begin{array}{ll} \text{First stage :} & y = \rho y_{-1} + X\beta + Wb + u,\ u \sim N(0,\Sigma), \Sigma = \tau^{-1} I_{N(T-1)} \\ \text{Second stage :} & \beta \sim N\left(\beta_0, \Lambda_\beta\right) \text{ and } b \sim N\left(b_0, \Lambda_b\right) \\ \text{with} & p\left(\tau\right) \propto \tau^{-1}\ ,\ \Lambda_\beta = \left(\tau g X'X\right)^{-1} \text{and } \Lambda_b = \left(\tau h W'W\right)^{-1}. \end{array} \tag{9}$$

where $y = (y_{1,2}, ..., y_{1,T}, ..., y_{N,2}, ..., y_{N,T})'$ and $y_{-1} = (y_{1,1}, ..., y_{1,T-1}, ..., y_{N,1}, ..., y_{N,T-1})'$ are $(N(T-1) \times 1)$. X is $(N(T-1) \times K_x)$, W is $(N(T-1) \times K_2)$, u is $(N(T-1) \times 1)$ and $I_{N(T-1)}$ is a $(N(T-1) \times N(T-1))$ identity matrix.

There is an extensive literature on autoregressive processes using Bayesian methods. The stationarity assumption implies that the autoregressive time dependence parameter space for ρ is a compact subset of $(-1, 1)$. For the pros and cons of imposing a stationarity hypothesis in a Bayesian setup see Phillips (1991). Ghosh and Heo (2003) proposed a comparative study using some selected noninformative (objective) priors for the $AR(1)$ model. Ibazizen and Fellag (2003), assumed a noninformative prior for the autoregressive parameter without considering the stationarity assumption for the AR(1) model. However, most papers consider a noninformative (objective) prior for the Bayesian analysis of an $AR(1)$ model without considering the stationarity assumption. See for example DeJong and Whiteman (1991), Schotman and van Dijk (1991), and Sims and Uhlig (1991). For the dynamic random coefficients panel data model, Hsiao and Pesaran (2008) do not impose any constraint on the coefficients of the lag dependent variable, ρ_i. But, following Liu and Tiao (1980), they suggest that one way to impose the stability condition on individual units would be to assume that ρ_i follows a rescaled Beta distribution on $(0, 1)$. In the time series framework, and for an $AR(1)$ model, Karakani et al. (2016) have performed a posterior sensitivity analysis based on Gibbs sampling with four different priors: natural conjugate prior, Jeffreys' prior, truncated normal prior and g-prior. Their respective performances are compared in terms of the highest posterior density region criterion. They show that the truncated normal distribution outperforms very slightly the g-prior and more strongly the other priors especially when the time dimension is small. On the other hand, for a larger time span, there is no significant difference between the truncated normal distribution and the g-prior.

Nevertheless, introducing a truncated normal distribution for ρ poses very complex integration problems due to the presence of the normal cdf function as integrand in the marginal likelihoods with ε-contamination class of prior distributions. To avoid these problems, ρ is assumed to be $U(-1, 1)$. In that case, its mean (0) and its variance (1/3) are exactly defined and we do not need to introduce an ε-contamination class of prior distributions for ρ at the second stage of the hierarchy. This was initially our first goal (see appendix A.1 in the supplementary material). Unfortunately, the results using Monte Carlo simulations showed biased estimates of ρ, β and residual variances (see Appendix A.2 in the supplementary material). Consequently, we assume a Zellner g-prior, for $\theta\big(= [\rho, \beta']'\big)$ which encompasses the coefficient of the lagged dependent variable $y_{i,t-1}$ and those of the explanatory variables X'_{it}. In other words, the two-stage hierarchy becomes.

$$\begin{array}{ll} \text{First stage :} & y = Z'_{it}\theta + Wb + u,\ u \sim N(0,\Sigma), \Sigma = \tau^{-1} I_{N(T-1)} \\ \text{Second stage :} & \theta \sim N\left(\theta_0, \Lambda_\theta\right) \text{ and } b \sim N\left(b_0, \Lambda_b\right) \\ \text{with} & p\left(\tau\right) \propto \tau^{-1}\ ,\ \Lambda_\theta = \left(\tau g Z'Z\right)^{-1} \text{and } \Lambda_b = \left(\tau h W'W\right)^{-1}. \end{array} \tag{10}$$

Thus, we do not impose stationarity constraints like many authors and we respect the philosophy of ε-contamination class using data-driven priors.

2.3. The robust dynamic linear model in the two-stage hierarchy

Using a two-step approach, we can integrate first with respect to (θ, τ) given b and then, conditional on θ, we integrate with respect to (b, τ).

1. Let $y^* = (y - Wb)$. Derive the conditional ML-II posterior distribution of θ given the specific effects b.
2. Let $\widetilde{y} = (y - Z\theta)$. Derive the conditional ML-II posterior distribution of b given the coefficients θ.

Thus, the marginal likelihoods (or predictive densities) corresponding to the base priors are:

$$m\left(y^*|\pi_0, b, g_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\theta, \tau|g_0\right) \times p\left(y^*|Z, b, \tau\right)\ d\theta\ d\tau$$

and

$$m\left(\widetilde{y}|\pi_0, \theta, h_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_2}} \pi_0\left(b, \tau|h_0\right) \times p\left(\widetilde{y}|W, \theta, \tau\right)\ db\ d\tau,$$

with

$$\begin{aligned}
\pi_0\left(\theta, \tau|g_0\right) &= \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} \left|\Lambda_Z\right|^{1/2} \exp\left(-\frac{\tau g_0}{2}\left(\theta - \theta_0\iota_{K_1}\right)'\Lambda_Z(\theta - \theta_0\iota_{K_1})\right)\right), \\
\pi_0\left(b, \tau|h_0\right) &= \left(\frac{\tau h_0}{2\pi}\right)^{\frac{K_2}{2}} \tau^{-1} \left|\Lambda_W\right|^{1/2} \exp\left(-\frac{\tau h_0}{2}(b - b_0\iota_{K_2})'\Lambda_W(b - b_0\iota_{K_2})\right).
\end{aligned}$$

Solving these equations is considerably easier than solving the equivalent expression in the one-step approach.

2.3.1. The first step of the robust Bayesian estimator

Let $y^* = y - Wb$. Combining the pdf of y^* and the pdf of the base prior, we get the predictive density corresponding to the base prior[4]:

$$\begin{aligned}
m\left(y^*|\pi_0, b, g_0\right) &= \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\theta, \tau|g_0\right) \times p\left(y^*|Z, b, \tau\right)\ d\theta\ d\tau \qquad (11) \\
&= \widetilde{H}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}\left(1 + \left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\theta_0}}{1 - R^2_{\theta_0}}\right)\right)^{-\frac{NT}{2}}
\end{aligned}$$

with $\widetilde{H} = \frac{\Gamma\left(\frac{NT}{2}\right)}{\pi^{\left(\frac{NT}{2}\right)} v(b)^{\left(\frac{NT}{2}\right)}}$, $R^2_{\theta_0} = \frac{(\widehat{\theta}(b) - \theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b) - \theta_0\iota_{K_1})}{(\widehat{\theta}(b) - \theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b) - \theta_0\iota_{K_1}) + v(b)}$, $\widehat{\theta}(b) = \Lambda_Z^{-1}Z'y^*$ and $v(b) = (y^* - Z\widehat{\theta}(b))'(y^* - Z\widehat{\theta}(b))$, and where $\Gamma(\cdot)$ is the Gamma function.

[4] Derivation can be found in the supplementary appendix of Baltagi et al. (2018).

Likewise, we can obtain the predictive density corresponding to the contaminated prior for the distribution $q(\theta, \tau | g_0, h_0) \in Q$ from the class Q of possible contamination distributions:

$$m(y^*|q, b, g_0) = \widetilde{H}\left(\frac{g_q}{g_q+1}\right)^{\frac{K_1}{2}}\left(1+\left(\frac{g_q}{g_q+1}\right)\left(\frac{R^2_{\theta_q}}{1-R^2_{\theta_q}}\right)\right)^{-\frac{NT}{2}}, \tag{12}$$

where

$$R^2_{\theta_q} = \frac{(\widehat{\theta}(b) - \theta_q \iota_{K_1})'\Lambda_Z(\widehat{\theta}(b) - \theta_q \iota_{K_1})}{(\widehat{\theta}(b) - \theta_q \iota_{K_1})'\Lambda_Z(\widehat{\theta}(b) - \theta_q \iota_{K_1}) + v(b)}.$$

As the ε-contamination of the prior distributions for (θ, τ) is defined by $\pi(\theta, \tau | g_0) = (1-\varepsilon)\pi_0(\theta, \tau | g_0) + \varepsilon q(\theta, \tau | g_0)$, the corresponding predictive density is given by:

$$m(y^*|\pi, b, g_0) = (1-\varepsilon)\, m(y^*|\pi_0, b, g_0) + \varepsilon m(y^*|q, b, g_0)$$

and

$$\sup_{\pi \in \Gamma} m(y^*|\pi, b, g_0) = (1-\varepsilon)\, m(y^*|\pi_0, b, g_0) + \varepsilon \sup_{q \in Q} m(y^*|q, b, g_0).$$

The maximization of $m(y^*|\pi, b, g_0)$ requires the maximization of $m(y^*|q, b, g_0)$ with respect to θ_q and g_q. The first-order conditions lead to

$$\widehat{\theta}_q = \left(\iota'_{K_1}\Lambda_Z \iota_{K_1}\right)^{-1} \iota'_{K_1}\Lambda_Z \widehat{\theta}(b) \tag{13}$$

and

$$\begin{aligned} \widehat{g}_q &= \min(g_0, g^*), \qquad (14)\\ \text{with } g^* &= \max\left[\left(\frac{(NT-K_1)}{K_1}\frac{(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1})'\Lambda_Z(\widehat{\theta}(b) - \widehat{\theta}_q \iota_{K_1})}{v(b)} - 1\right)^{-1}, 0\right] \\ &= \max\left[\left(\frac{(NT-K_1)}{K_1}\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right) - 1\right)^{-1}, 0\right]. \end{aligned}$$

Denote $\sup_{q\in Q} m(y^*|q, b, g_0) = m(y^*|\widehat{q}, b, g_0)$. Then

$$m(y^*|\widehat{q}, b, g_0) = \widetilde{H}\left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)^{\frac{K_1}{2}}\left(1+\left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right)\right)^{-\frac{NT}{2}}.$$

Let $\pi_0^*(\theta, \tau | g_0)$ denote the posterior density of (θ, τ) based upon the prior $\pi_0(\theta, \tau | g_0)$. Also, let $q^*(\theta, \tau | g_0)$ denote the posterior density of (θ, τ) based upon the prior $q(\theta, \tau | g_0)$. The ML-II posterior density of θ is thus given by:

$$\begin{aligned} \widehat{\pi}^*(\theta | g_0) &= \int_0^\infty \widehat{\pi}^*(\theta, \tau | g_0)\, d\tau \\ &= \widehat{\lambda}_{\theta, g_0}\int_0^\infty \pi_0^*(\theta, \tau | g_0)\, d\tau + \left(1-\widehat{\lambda}_{\theta, g_0}\right)\int_0^\infty q^*(\theta, \tau | g_0)\, d\tau \\ &= \widehat{\lambda}_{\theta, g_0}\pi_0^*(\theta | g_0) + \left(1-\widehat{\lambda}_{\theta, g_0}\right)\widehat{q}^*(\theta | g_0) \qquad (15) \end{aligned}$$

with

$$\widehat{\lambda}_{\theta,g_0} = \left[1 + \frac{\varepsilon}{1-\varepsilon} \left(\frac{\frac{\widehat{g}_q}{\widehat{g}_q+1}}{\frac{g_0}{g_0+1}} \right)^{K_1/2} \left(\frac{1 + \left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\theta_0}}{1-R^2_{\theta_0}}\right)}{1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right)} \right)^{\frac{NT}{2}} \right]^{-1}.$$

Note that $\widehat{\lambda}_{\theta,g_0}$ depends upon the ratio of the $R^2_{\theta_0}$ and $R^2_{\widehat{\theta}_q}$, but primarily on the sample size NT. Indeed, $\widehat{\lambda}_{\theta,g_0}$ tends to 0 when $R^2_{\theta_0} > R^2_{\widehat{\theta}_q}$ and tends to 1 when $R^2_{\theta_0} < R^2_{\widehat{\theta}_q}$, irrespective of the model fit (*i.e*, the absolute values of $R^2_{\theta_0}$ or $R^2_{\widehat{\theta}_q}$). Only the relative values of $R^2_{\widehat{\theta}_q}$ and $R^2_{\theta_0}$ matter.

It can be shown that $\pi_0^*(\theta|g_0)$ is the pdf (see the supplementary appendix of Baltagi et al. (2018)) of a multivariate t-distribution with mean vector $\theta_*(b|g_0)$, variance-covariance matrix $\left(\frac{\xi_{0,\theta} M_{0,\theta}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) with

$$M_{0,\theta} = \frac{(g_0+1)}{v(b)}\Lambda_Z \text{ and } \xi_{0,\theta} = 1 + \left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\theta_0}}{1-R^2_{\theta_0}}\right). \tag{16}$$

$\theta_*(b|g_0)$ is the Bayes estimate of θ for the prior distribution $\pi_0(\theta,\tau)$:

$$\theta_*(b|g_0) = \frac{\widehat{\theta}(b) + g_0\theta_0\iota_{K_1}}{g_0+1}. \tag{17}$$

Likewise $\widehat{q}^*(\theta)$ is the pdf of a multivariate t-distribution with mean vector $\widehat{\theta}_{EB}(b|g_0)$, variance-covariance matrix $\left(\frac{\xi_{q,\theta} M_{q,\theta}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) with

$$\xi_{q,\theta} = 1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right) \text{ and } M_{q,\theta} = \left(\frac{(\widehat{g}_q+1)}{v(b)}\right)\Lambda_Z, \tag{18}$$

where $\widehat{\theta}_{EB}(b|g_0)$ is the empirical Bayes estimator of θ for the contaminated prior distribution $q(\theta,\tau)$ given by:

$$\widehat{\theta}_{EB}(b|g_0) = \frac{\widehat{\theta}(b) + \widehat{g}_q\widehat{\theta}_q\iota_{K_1}}{\widehat{g}_q+1}. \tag{19}$$

The mean of the ML-II posterior density of θ is then:

$$\begin{aligned} \widehat{\theta}_{ML-II} &= E[\widehat{\pi}^*(\theta|g_0)] \\ &= \widehat{\lambda}_{\theta,g_0} E[\pi_0^*(\theta|g_0)] + \left(1-\widehat{\lambda}_{\theta,g_0}\right) E[\widehat{q}^*(\theta|g_0)] \\ &= \widehat{\lambda}_{\theta,g_0}\theta_*(b|g_0) + \left(1-\widehat{\lambda}_{\theta,g_0}\right)\widehat{\theta}_{EB}(b|g_0). \end{aligned} \tag{20}$$

The ML-II posterior density of θ, given b and g_0 is a shrinkage estimator. It is a weighted average of the Bayes estimator $\theta_*(b|g_0)$ under base prior g_0 and the data-dependent empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. If the base prior is consistent with the data, the weight $\widehat{\lambda}_{\theta,g_0} \to 1$ and the ML-II posterior density of θ gives more weight to the posterior $\pi_0^*(\theta|g_0)$ derived from the elicited prior.

In this case $\widehat{\theta}_{ML-II}$ is close to the Bayes estimator $\theta_*(b|g_0)$. Conversely, if the base prior is not consistent with the data, the weight $\widehat{\lambda}_{\theta,g_0} \to 0$ and the ML-II posterior density of θ is then close to the posterior $\widehat{q}^*(\theta|g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. The ability of the ε-contamination model to extract more information from the data is what makes it superior to the classical Bayes estimator based on a single base prior.[5]

2.3.2. The second step of the robust Bayesian estimator

Let $\widetilde{y} = y - Z\theta$. Moving along the lines of the first step, the ML-II posterior density of b is given by:

$$\widehat{\pi}^*(b|h_0) = \widehat{\lambda}_{b,h_0}\pi_0^*(b|h_0) + \left(1 - \widehat{\lambda}_{b,h_0}\right)\widehat{q}^*(b|h_0)$$

with

$$\widehat{\lambda}_{b,h_0} = \left[1 + \frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{h}}{\widehat{h}+1}}{\frac{h_0}{h_0+1}}\right)^{K_2/2}\left(\frac{1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_0}}{1-R^2_{b_0}}\right)}{1+\left(\frac{\widehat{h}}{\widehat{h}+1}\right)\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right)}\right)^{\frac{NT}{2}}\right]^{-1},$$

where

$$R^2_{b_0} = \frac{(\widehat{b}(\theta) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - b_0\iota_{K_2})}{(\widehat{b}(\theta) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - b_0\iota_{K_2}) + v(\theta)},$$

$$R^2_{\widehat{b}_q} = \frac{(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2})}{(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2}) + v(\theta)},$$

with $\widehat{b}(\theta) = \Lambda_W^{-1}W'\widetilde{y}$ and $v(\theta) = (\widetilde{y} - W\widehat{b}(\theta))'(\widetilde{y} - W\widehat{b}(\theta))$,

$$\widehat{b}_q = \left(\iota'_{K_2}\Lambda_W\iota_{K_2}\right)^{-1}\iota'_{K_2}\Lambda_W\widehat{b}(\theta)$$

and

$$\begin{aligned}\widehat{h}_q &= \min(h_0, h^*)\\ \text{with } h^* &= \max\left[\left(\frac{(NT-K_2)}{K_2}\frac{(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - \widehat{b}_q\iota_{K_2})}{v(\theta)} - 1\right)^{-1}, 0\right]\\ &= \max\left[\left(\frac{(NT-K_2)}{K_2}\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right) - 1\right)^{-1}, 0\right].\end{aligned}$$

$\pi_0^*(b|h_0)$ is the pdf of a multivariate t-distribution with mean vector $b_*(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{0,b}M_{0,b}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) with

$$M_{0,b} = \frac{(h_0+1)}{v(\theta)}\Lambda_W \text{ and } \xi_{0,b} = 1 + \left(\frac{h_0}{h_0+1}\right)\frac{(\widehat{b}(\theta) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}(\theta) - b_0\iota_{K_2})}{v(\theta)}.$$

[5] Following Berger (1985), Baltagi et al. (2018) derived the analytical ML-II posterior variance-covariance matrix of θ (see the supplementary appendix of Baltagi et al. (2018)).

$b_*(\theta|h_0)$ is the Bayes estimate of b for the prior distribution $\pi_0(b, \tau|h_0)$:

$$b_*(\theta|h_0) = \frac{\widehat{b}(\theta) + h_0 b_0 \iota_{K_2}}{h_0 + 1}.$$

$q^*(b|h_0)$ is the pdf of a multivariate t-distribution with mean vector $\widehat{b}_{EB}(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{1,b} M_{1,b}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) with

$$\xi_{1,b} = 1 + \left(\frac{\widehat{h}_q}{\widehat{h}_q + 1}\right) \frac{(\widehat{b}(\theta) - \widehat{b}_q \iota_{K_2})' \Lambda_W (\widehat{b}(\theta) - \widehat{b}_q \iota_{K_2})}{v(\theta)} \text{ and } M_{1,b} = \left(\frac{\widehat{h} + 1}{v(\theta)}\right) \Lambda_W$$

$\widehat{b}_{EB}(\theta|h_0)$ is the empirical Bayes estimator of b for the contaminated prior distribution $q(b, \tau|h_0)$:

$$\widehat{b}_{EB}(\theta|h_0) = \frac{\widehat{\theta}(b) + \widehat{h}_q \widehat{b}_q \iota_{K_2}}{\widehat{h}_q + 1}.$$

The mean of the ML-II posterior density of b is hence given by:

$$\widehat{b}_{ML-II} = \widehat{\lambda}_b b_*(\theta|h_0) + \left(1 - \widehat{\lambda}_\theta\right) \widehat{b}_{EB}(\theta|h_0). \tag{21}$$

The ML-II posterior variance-covariance matrix of b can be derived in a similar fashion[6] to that of $\hat{\theta}_{ML-II}$.

2.4. Estimating the ML-II posterior variance-covariance matrix

Many have raised concerns about the unbiasedness of the posterior variance-covariance matrices of $\widehat{\theta}_{ML-II}$ and $\widehat{b}_{ML-II}$. Indeed, they will both be biased towards zero as $\widehat{\lambda}_{\theta,g_0}$ and $\widehat{\lambda}_{b,h_0} \to 0$ and converge to the empirical variance which is known to underestimate the true variance (see *e.g.* Berger and Berliner (1986); Gilks et al. (1997); Robert (2007)). Consequently, the assessment of the performance of either $\widehat{\theta}_{ML-II}$ or $\widehat{b}_{ML-II}$ using standard quadratic loss functions can not be conducted using the analytical expressions. What is needed is an unbiased estimator of the true ML-II variances. Baltagi et al. (2018) proposed two different strategies to approximate these, each with different desirable properties: MCMC with multivariate t-distributions or block resampling bootstrap. Simulations show that one needs as few as 20 bootstrap samples to achieve acceptable results[7]. Here, we will use the same individual block resampling bootstrap method. Following Bellman et al. (1989); Andersson and Karlsson (2001), and Kapetanios (2008), individual block resampling consists of drawing an $(N \times T)$ matrix Y^{BR} whose rows are obtained by resampling those of an $(N \times T)$ matrix Y with replacement. Conditionally on Y, the rows of Y^{BR} are independent and identically distributed. The following algorithm is used to approximate the variance matrices:

1. Loop over BR samples

[6] See the supplementary appendix of Baltagi et al. (2018).

[7] Increasing the number of bootstrap samples does not change the results but increases the computation time considerably.

2. In the first step, compute the mean of the ML-II posterior density of θ using our initial shrinkage procedure

$$\begin{aligned}\widehat{\theta}_{ML-II,br} &= E\left[\widehat{\pi}^*\left(\theta|g_0\right)\right] \\ &= \widehat{\lambda}_{\theta,g_0}\theta_*(b|g_0) + \left(1-\widehat{\lambda}_{\theta,g_0}\right)\widehat{\theta}_{EB}\left(b|g_0\right).\end{aligned}$$

3. In the second step, compute the mean of the ML-II posterior density of b:

$$\widehat{b}_{ML-II,br} = \widehat{\lambda}_b b_*(\theta|h_0) + \left(1-\widehat{\lambda}_\theta\right)\widehat{b}_{EB}\left(\theta|h_0\right)$$

4. Once the BR bootstraps are completed, use the $(K_1 \times BR)$ matrix of coefficients $\theta^{(BR)}$ and the $(N \times BR)$ matrix of coefficients $b^{(BR)}$ to compute:

$$\begin{aligned}\widehat{\theta}_{ML-II} &= E\left[\theta^{(BR)}\right], & \widehat{\sigma}_{\theta_{ML-II}} &= \sqrt{diag\left(Var\left[\theta^{(BR)}\right]\right)} \\ \widehat{b}_{ML-II} &= E\left[b^{(BR)}\right], & \widehat{\sigma}_{b_{ML-II}} &= \sqrt{diag\left(Var\left[b^{(BR)}\right]\right)}\end{aligned}$$

3. Monte Carlo simulation study

In what follows, we compare the finite sample properties of our proposed estimator with those of standard classical estimators.

3.1. The DGP of the Monte Carlo simulation study

For the random effects (RE), the Chamberlain (1982)-type fixed effects (FE) world and the Hausman and Taylor (1981) (HT) worlds, we use the same DGP as that of Baltagi et al. (2018) extended to the dynamic case. For the dynamic homogeneous/heterogeneous panel data model with common trends or with common correlated effects, we will follow Chudik and Pesaran (2015a,b).

$$\begin{aligned}y_{it} &= \rho y_{i,t-1} + x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + V_{1,i}\eta_1 + V_{2,i}\eta_2 + \mu_i + u_{it}, \quad (22)\\ &\quad \text{for } i=1,...,N\ ,\ t=2,...,T, \text{ with}\end{aligned}$$

$$\begin{aligned}x_{1,1,it} &= 0.7x_{1,1,it-1} + \delta_i + \zeta_{it} \\ x_{1,2,it} &= 0.7x_{1,2,it-1} + \theta_i + \varsigma_{it} \\ u_{it} &\sim N\left(0,\tau^{-1}\right),\ (\delta_i,\theta_i,\zeta_{it},\varsigma_{it}) \sim U(-2,2) \\ \text{and } \rho &= 0.75,\ \beta_{1,1}=\beta_{1,2}=\beta_2=1.\end{aligned}$$

1. For a random effects (RE) world, we assume that:

$$\begin{aligned}\eta_1 &= \eta_2 = 0 \\ x_{2,it} &= 0.7x_{2,it-1} + \kappa_i + \vartheta_{it}\ ,\ (\kappa_i,\vartheta_{it}) \sim U(-2,2) \\ \mu_i &\sim N\left(0,\sigma^2_\mu\right)\ ,\ \sigma^2_\mu = 4\tau^{-1}.\end{aligned}$$

Furthermore, $x_{1,1,it}$, $x_{1,2,it}$ and $x_{2,it}$ are assumed to be exogenous in that they are not correlated with μ_i and u_{it}.

2. For a Chamberlain-type fixed effects (FE) world, we assume that:

$$
\begin{aligned}
\eta_1 &= \eta_2 = 0; \\
x_{2,it} &= \delta_{2,i} + \omega_{2,it} \ , \ \delta_{2,i} \sim N(m_{\delta_2}, \sigma^2_{\delta_2}), \ \omega_{2,it} \sim N(m_{\omega_2}, \sigma^2_{\omega_2}); \\
m_{\delta_2} &= m_{\omega_2} = 1, \ \sigma^2_{\delta_2} = 8, \ \sigma^2_{\omega_2} = 2; \\
\mu_i &= x_{2,i1}\pi_1 + x_{2,i2}\pi_2 + ... + x_{2,iT}\pi_T + \nu_i, \ \nu_i \sim N(0, \sigma^2_\nu); \\
\sigma^2_\nu &= 1, \ \pi_t = (0.8)^{T-t} \text{ for } t = 1, ..., T.
\end{aligned}
$$

$x_{1,1,it}$ and $x_{1,2,it}$ are assumed to be exogenous but $x_{2,it}$ is correlated with the μ_i and we assume an exponential growth for the correlation coefficient π_t.

3. For a Hausman-Taylor (HT) world, we assume that:

$$
\begin{aligned}
\eta_1 &= \eta_2 = 1; \\
x_{2,it} &= 0.7x_{2,it-1} + \mu_i + \vartheta_{it} \ , \ \vartheta_{it} \sim U(-2, 2); \\
V_{1,i} &= 1, \forall i; \\
V_{2,i} &= \mu_i + \delta_i + \theta_i + \xi_i, \ \xi_i \sim U(-2, 2); \\
\mu_i &\sim N\left(0, \sigma^2_\mu\right) \text{ and } \sigma^2_\mu = 4\tau^{-1}.
\end{aligned}
$$

$x_{1,1,it}$ and $x_{1,2,it}$ and $V_{1,i}$ are assumed to be exogenous while $x_{2,it}$ and $V_{2,i}$ are endogenous because they are correlated with the μ_i but not with the u_{it}.

4. For the homogeneous panel data world with common trends, we follow Chudik and Pesaran (2015a,b) and assume that

$$
y_{it} = \rho y_{i,t-1} + x_{it}\beta_1 + x_{i,t-1}\beta_2 + f'_t\gamma_i + u_{it}, \text{ for } i = 1, ..., N \ , \ t = 2, ..., T, \tag{23}
$$

with

$$
\begin{aligned}
x_{it} &= \alpha_{x_i} y_{i,t-1} + f'_t\gamma_{x_i} + \omega_{x_{it}} \\
\omega_{x_{it}} &= \varrho_{x_i}\omega_{x_{it-1}} + \zeta_{x_{it}} \\
\gamma_{il} &= \gamma_l + \eta_{i,\gamma_l}, \text{ for } l = 1, ..., m \\
\gamma_{x_{il}} &= \gamma_{x_l} + \eta_{i,\gamma_{x_l}}, \text{ for } l = 1, ..., m
\end{aligned}
$$

where

$$
\begin{array}{lll}
\zeta_{x_{it}} \sim N(0, \sigma^2_{\omega_{x_i}}) \ , & \sigma^2_{\omega_{x_i}} = \beta_1\sqrt{1 - [E(\varrho_{x_i})]^2} \ , & \varrho_{x_i} \sim N(0, 0.9), \text{ for } i = 1, ..., N \\
\eta_{i,\gamma_l} \sim N(0, \sigma^2_{\gamma_l}) \ , & \eta_{i,\gamma_{x_l}} \sim N(0, \sigma^2_{\gamma_{x_l}}), \text{ for } l = 1, ..., m \ , & \sigma^2_{\gamma_l} = \sigma^2_{\gamma_{x_l}} = 0.2^2 \\
\gamma_l = \sqrt{l \times c_\gamma} \ , & \gamma_{x_l} = \sqrt{l \times c_{x,l}} & c_\gamma = (1/m) - \sigma^2_{\gamma_l} \\
c_{x,l} = \dfrac{2}{m(m+1)} - \dfrac{2\sigma^2_{\gamma_{x_l}}}{(m+1)} \ , & \rho = 0.75, \ \beta_1 = \beta_2 = 1 & \text{and } u_{it} \sim N\left(0, \tau^{-1}\right).
\end{array}
$$

f_t and γ_i are $(m \times 1)$ vectors. We consider $m = 2$ deterministic known common trends: one linear trend $f_{t,1} = t/T$ and one polynomial trend: $f_{t,2} = t/T + 1.4(t/T)^2 - 3(t/T)^3$ for $t = 1, ..., T$. The feedback coefficients follow a uniform distribution $\alpha_{x_i} \sim U(0, 0.15)$ and are non-zero for all i ($\alpha_{x_i} \neq 0$). They lead to weakly exogenous regressors x_{it}.

5. For the homogeneous panel data world with correlated common effects, we follow Chudik and Pesaran (2015a,b), and assume that the m common trends, f_t (23), are replaced with unobserved common factors:

$$f_{tl} = \rho_{fl} f_{t-1,l} + \xi_{ftl},\ \xi_{ftl} \sim N(0, 1-\rho_{fl}^2),\ l = 1, ..., m$$

   We assume that the common factors are independent stationary $AR(1)$ processes with $\rho_{fl} = 0.6$ for $l = 1, ..., m$.
6. For the heterogeneous panel data world with correlated common effects, we follow Chudik and Pesaran (2015a,b) and assume that ρ (resp. β_1) in the model (23) is replaced by individual coefficients $\rho_i \sim U(0.6, 0.9)$ (resp. $\beta_{1i} \sim U(0.5, 1)$) for $i = 1, ..., N$ and we keep the m unobserved common factors as defined previously.

For each set-up, we vary the size of the sample and the length of the panel. We choose several (N, T) pairs with $N = 100$, 200 and $T = 10$, 30 for cases 1 to 3 and $N = (50, 100)$ and $T = (30, 50)$ for cases 4 to 6. The autoregressive coefficient is set as $\rho = 0.75$. We set the initial values of y_{it}, $x_{1,1,it}$, $x_{1,2,it}$ and $x_{2,it}$, x_{it} to zero. We next generate all the $x_{1,1,it}$, $x_{1,2,it}$, $x_{1,2,it}$, x_{it}, y_{it}, u_{it}, ζ_{it}, ς_{it}, $\omega_{2,it}$, $\ldots$ over $T + T_0$ time periods and we drop the first $T_0 (= 50)$ observations to reduce the dependence on the initial values. The robust Bayesian estimators for the two-stage hierarchy are estimated with $\varepsilon = 0.5$, though we investigate the robustness of our results to various values of ε.[8]

We must set the hyperparameters values $\theta_0, b_0, g_0, h_0, \tau$ for the initial distributions of $\theta \sim N\left(\theta_0 \iota_{K_1}, (\tau g_0 \Lambda_Z)^{-1}\right)$ and $b \sim N\left(b_0 \iota_{K_2}, (\tau h_0 \Lambda_W)^{-1}\right)$ where $\theta = [\rho, \beta_{1,1}, \beta_{1,2}, \beta_2]'$ for the first three cases and $\theta = [\rho, \beta_1, \beta_2]'$ for the last three cases. While we can choose arbitrary values for θ_0, b_0 and τ, the literature generally recommends using the unit information prior (UIP) to set the g-priors.[9] In the normal regression case, and following Kass and Wasserman (1995), the UIP corresponds to $g_0 = h_0 = 1/NT$, leading to Bayes factors that behave like the Bayesian Information Criterion (BIC).

For the 2S robust estimators, we use $BR = 20$ samples in the block resampling bootstrap. For each experiment, we run $R = 1,000$ replications and we compute the means, standard errors and root mean squared errors (RMSEs) of the coefficients and the residual variances.

3.1.1. The random effects world

Rewrite the general dynamic model (8) as follows:

$$\begin{aligned} y =& Z\theta + Wb + u = Z\theta + Z_\mu \mu + u \\ & \text{with } Z'_{it} = [y_{it-1}, X'_{it}]\ ,\ \theta' = [\rho, \beta']' \text{ and } X'_{it} = [x_{1,1,it}, x_{1,2,it}, x_{2,it}]\,, \end{aligned}$$

where $u \sim N(0, \Sigma)$, $\Sigma = \tau^{-1} I_{N(T-1)}$, $Z_\mu = I_N \otimes \iota_{T-1}$ is $(N(T-1) \times N)$, $\otimes$ is the Kronecker product, ι_{T-1} is a $(T-1 \times 1)$ vector of ones and $\mu (\equiv b)$ is a $(N \times 1)$ vector of idiosyncratic parameters. When $W \equiv Z_\mu$, the random effects, $\mu \sim N\left(0, \sigma_\mu^2 I_N\right)$, are associated with the error term $\nu = Z_\mu \mu + u$ with $\operatorname{Var}(\nu) = \sigma_\mu^2 (I_N \otimes J_{T-1}) + \sigma_u^2 I_{N(T-1)}$, where $J_{T-1} = \iota_{T-1} \iota'_{T-1}$. This

[8] $\varepsilon = 0.5$ is an arbitrary value. This implicitly assumes that the amount of error in the base elicited prior is 50%. In other words, $\varepsilon = 0.5$ means that we elicit the π_0 prior but feel we could be as much as 50% off (in terms of implied probability sets).

[9] We chose: $\theta_0 = 0, b_0 = 0$ and $\tau = 1$.

model is usually estimated using GMM (see Arellano and Bond (1991); Blundell and Bond (1998), amongst others). It could also be estimated using the quasi-maximum likelihood (QML) estimator (see Bhargava and Sargan (1983), Kripfganz (2016), Bun et al. (2017), Moral-Benito et al. (2019)). Thus we compare our Bayesian two stage estimator with the Arellano-Bond GMM and the QML estimators.[10]

Table 1 reports the results of fitting the Bayesian two stage model with block resampling bootstrap (*2S bootstrap*)[11] along with those from the GMM and QMLE, each in a separate panel respectively for ($N = 100, T = 10$) and ($N = 200, T = 30$). The true parameter values appear in the first row of the Table. The last column reports the computation time in seconds.[12] Note that the computation time increases significantly as we move from a small sample to a larger one (the QMLE being the fastest).

The first noteworthy feature of the Table is that all the estimators yield parameter estimates, standard errors[13] and RMSEs that are very close. For the coefficient of the lagged dependent variable, ρ, the RMSE is the lowest for the *2S bootstrap* when $N = 100$ and $T = 10$, but this RMSE is the lowest for the QMLE when $N = 200$ and $T = 30$, although the differences are small. GMM yields higher RMSEs for all coefficients. For the β coefficients, results are mixed in terms of RMSE for $N = 100$ and $T = 10$, but QMLE is the best for $N = 200$ and $T = 30$ with still small differences. The *2S bootstrap* has better RMSEs than the frequentist estimators (GMM and QMLE) for the residual disturbances (σ_u^2) and the random effects (σ_μ^2). Table 1 confirms that the base prior is not consistent with the data since $\hat{\lambda}_{\theta,g_0}$ is close to zero. The ML-II posterior density of θ is close to the posterior $\widehat{q}^*(\theta|g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(b|g_0)$. In contrast, $\widehat{\lambda}_\mu$ is close to 0.5 so the Bayes estimator $b_*(\theta|h_0)$ under the base prior h_0 and the empirical Bayes estimator $\widehat{b}_{EB}(\theta|h_0)$ each contribute similarly to the random effects $b_i(\equiv \mu_i)$.

Table B.2 in the supplementary material gives the results when the coefficient ρ of the lagged dependent variable is increased from 0.75 to 0.98 (close to the unit root) for $N = 100$ and $T = 10$. The GMM estimator performs the worst as compared to the two other estimators. Even with a coefficient ρ very close to the unit root, the 95% Highest Posterior Density Interval (HPDI) of the Bayesian estimator confirm the stationarity of the $AR(1)$ process. It does not therefore seem necessary to impose a stationarity constraint on the prior distribution of ρ. QMLE has the lowest RMSE (although the differences are small) except for σ_u^2 and σ_μ^2 where *2S bootstrap* is the best as reported in Table 1.

[10]We use our own R codes for our Bayesian estimator, the R package "plm" for the Arellano-Bond GMM estimator and the "xtdpdqml" Stata package. We use the same DGP set under R and Stata environments to compare the three methods. We thank Jean-Michel Etienne for his help and support with the full-blown programming language Mata of Stata.

[11]Recall that we use only $BR = 20$ individual block bootstrap samples. Fortunately, the results are very robust to the value of BR. For instance, increasing BR from 20 to 200 in the random effects world increases the computation time tenfold but yields practically the same results.

[12]The simulations were conducted using R version 3.3.2 on a MacBook Pro, 2.8 GHz core i7 with 16Go 1600 MGz DDR3 ram.

[13]Strictly speaking, we should mention "posterior means" and "posterior standard errors" whenever we refer to Bayesian estimates and "coefficients" and "standard errors" when discussing frequentist ones. For the sake of brevity, we will use "coefficients" and "standard errors" in both cases.

3.1.2. The Chamberlain-type fixed effects world

For the Chamberlain (1982)-type specification, the individual effects are given by $\mu = \underline{X}\Pi + \varpi$, where $\underline{X}$ is a $(N \times (T-1)K_x)$ matrix with $\underline{X}_i = (X'_{i2}, ..., X'_{iT})$ and $\Pi = (\pi'_2, ..., \pi'_T)'$ is a $((T-1)K_x \times 1)$ vector. Here π_t is a $(K_x \times 1)$ vector of parameters to be estimated. The model can be rewritten as: $y = Z\theta + Z_\mu \underline{X}\Pi + Z_\mu \varpi + u$. We concatenate $[Z, Z_\mu \underline{X}]$ into a single matrix of observables Z^* and let $Wb \equiv Z_\mu \varpi$.

For the Chamberlain world, we compare the QML estimator to our Bayesian estimator. These are based on the transformed model: $y_{it} = \rho y_{i,t-1} + x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + \sum_{t=2}^{T} x_{2,it}\pi_t + \varpi_i + u_{it}$ or $y = Z^*\theta^* + Wb + u = Z^*\theta^* + Z_\mu \varpi + u$ where $Z^* = \left[y_{-1}, x_{1,1}, x_{1,2}, x_2, \underline{x_2}\right]$, $W = Z_\mu$ and $b = \varpi$.

Table 2 once again shows that the results of the *2S bootstrap* are very close to those of the QML estimator. *2S bootstrap* has the lowest RMSE for $N = 100$ and $T = 10$ except for σ^2_μ. QMLE has the lowest RMSEs for all the parameters when $N = 200$ and $T = 30$. Table B.3 in the supplementary appendix gives the estimates of the π_t coefficients. The RMSEs are lower for QMLE than *2S bootstrap*, although again the differences are very small. Table B.4 in the supplementary material report the results for $N = 200$ and $T = 30$.

3.1.3. The Hausman-Taylor world

The static Hausman-Taylor model (henceforth HT, see Hausman and Taylor (1981)) posits that $y = X\beta + V\eta + Z_\mu \mu + u$, where V is a vector of time-invariant variables, and that subsets of X (*e.g.*, $X'_{2,i}$) and V (*e.g.*, V'_{2i}) may be correlated with the individual effects μ, but leaves the correlations unspecified. Hausman and Taylor (1981) proposed a two-step IV estimator.

For our dynamic general model (8) and for equation (22): $y = Z\theta + Wb + u = \rho y_{-1} + X\beta + V\eta + Z_\mu \mu + u$, we assume that ($\overline{X'_{2,i}}$, V'_{2i} and μ_i) are jointly normally distributed:

$$\begin{pmatrix} \mu_i \\ \begin{pmatrix} \overline{X'_{2,i}} \\ V'_{2i} \end{pmatrix} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \begin{pmatrix} E_{\overline{X'_2}} \\ E_{V'_2} \end{pmatrix} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

where $\overline{X'_{2,i}}$ is the individual mean of $X'_{2,it}$. The conditional distribution of $\mu_i \mid \overline{X'_{2,i}}, V'_{2i}$ is given by:

$$\mu_i \mid \overline{X'_{2,i}}, V'_{2i} \sim N\left(\Sigma_{12}\Sigma_{22}^{-1}. \begin{pmatrix} \overline{X'_{2,i}} - E_{\overline{X'_2}} \\ V'_{2i} - E_{V'_2} \end{pmatrix}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

Since we do not know the elements of the variance-covariance matrix Σ_{js}, we can write:

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right)\theta_X + \left(V'_{2i} - E_{V'_2}\right)\theta_V + \varpi_i,$$

where $\varpi_i \sim N\left(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$ is uncorrelated with u_{it}, and where θ_X and θ_V are vectors of parameters to be estimated. In order to identify the coefficient vector of V'_{2i} and to avoid possible collinearity problems, we assume that the individual effects are given by:

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right)\theta_X + f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)\right]\theta_V + \varpi_i, \tag{24}$$

where $\odot$ is the Hadamard product and $f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)\right]$ can be a nonlinear function of $\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(V'_{2i} - E_{V'_2}\right)$. The first term on the right-hand side of equation (24) corresponds to the Mundlak (1978) transformation while the middle term captures the correlation between V'_{2i} and μ_i. The individual effects, μ, are a function of PX and $(f\,[PX \odot V])$, *i.e.,* a function of the column-by-column Hadamard product of PX and V where $P = (I_N \otimes J_{T-1})/(T-1))$ is the between transformation. We can once again concatenate $[y_{-1}, X, PX, f\,[PX \odot V]]$ into a single matrix of observables Z^* and let $Wb \equiv Z_\mu \varpi$.

For our model (22), $y_{it} = \rho y_{i,t-1} + x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + V_{1,i}\eta_1 + V_{2,i}\eta_2 + \mu_i + u_{it}$ or $y = \rho y_{-1} + X_1\beta_1 + x_2\beta_2 + V_1\eta_1 + V_2\eta_2 + Z_\mu\mu + u$. Then, we assume that

$$\mu_i = (\overline{x_{2,i}} - E_{\overline{x_2}})\,\theta_X + f\,[(\overline{x_{2,i}} - E_{\overline{x_2}}) \odot (V_{2i} - E_{V_2})]\,\theta_V + \varpi_i. \tag{25}$$

We propose adopting the following strategy: If the correlation between μ_i and V_{2i} is quite large (> 0.2), use $f\,[.] = (\overline{x_{2,i}} - E_{\overline{x_2}})^2 \odot (V_{2i} - E_{V_2})^s$ with $s = 1$. If the correlation is weak, set $s = 2$. In real-world applications, we do not know the correlation between μ_i and V_{2i} *a priori.* We can use a proxy of μ_i defined by the OLS estimation of μ: $\widehat{\mu} = \left(Z'_\mu Z_\mu\right)^{-1} Z'_\mu \widehat{y}$ where $\widehat{y}$ are the fitted values of the pooling regression $y = \rho y_{-1} + X_1\beta_1 + x_2\beta_2 + V_1\eta_1 + V_2\eta_2 + \zeta$. Then, we compute the correlation between $\widehat{\mu}$ and V_2. In our simulation study, it turns out the correlations between μ and V_2 are large: 0.65. Hence, we choose $s = 1$. In this specification, $Z = [y_{-1}, x_{1,1}, x_{1,2}, x_2, V_1, V_2, Px_2, f\,[Px_2 \odot V_2]]$, $W = Z_\mu$ and $b = \varpi$.

Our *2S bootstrap* estimation method is compared with the two-stage quasi-maximum likelihood sequential approach proposed by Kripfganz and Schwarz (2019). In the first stage, they estimate the coefficients of the time-varying regressors without relying on coefficient estimates for the time-invariant regressors using the quasi-maximum likelihood (QML) estimator of Hsiao et al. (2002) with the "xtdpdqml" Stata command. Subsequently, they regress the first-stage residuals on the time-invariant regressors. They achieve identification by using instrumental variables in the spirit of Hausman and Taylor (1981), and they adjust the second-stage standard errors to account for the first-stage estimation error.[14] They have proposed a new "xtseqreg" Stata command which implements the standard error correction for two-stage dynamic linear panel data models.[15]

Table 3 compares results of the *2S bootstrap* estimator to those of the two-stage QML sequential approach. Once again, the estimates are very close to one another. The RMSE is smaller for the two-stage QML than for the *2S bootstrap* for all the parameters for $N = 200$ and $T = 10$. On the other hand, the *2S bootstrap* has a lower RMSE for η_2 (for $N = 100$ and $T = 10$) and ($N = 100$ and $T = 30$). This is true despite the fact that the *2S bootstrap* estimator yields a slightly upward biased estimate of η_2, the coefficient associated with the time-invariant variable $Z_{2,i}$ which is itself correlated with μ_i. This bias decreases as T increases (from 16% for $T = 10$ to 4.7% for $T = 30$). Interestingly, the standard errors of that same coefficient η_2 are smaller when using the Bayesian estimator as compared to the two-stage QMLE, and especially when T is larger. Even with a slight bias, the 95% confidence intervals of the Bayesian estimator are narrower and entirely nested within those obtained with the two-stage QML sequential approach. We also reached the same conclusion

[14] For the following specification: $y_{it} = \rho y_{i,t-1} + x'_{it}\beta + V'_i\eta + \mu_i + u_{it}$, the first stage model is $y_{it} = \rho y_{i,t-1} + x'_{it}\beta + \overline{\kappa} + e_{it}$, where $e_{it} = \kappa_i - \overline{\kappa} + u_{it}$, $\kappa_i = V'_i\eta + \mu_i$, $\overline{\kappa} = E[\kappa_i]$ and is estimated in first differences. In the second stage, Kripfganz and Schwarz (2019) estimate the coefficients η based on the level relationship: $y_{it} - \widehat{\rho} y_{i,t-1} - x'_{it}\widehat{\beta} = V'_i\eta + \vartheta_{it}$ where $\vartheta_{it} = \mu_i + u_{it} + (\widehat{\rho} - \rho)y_{i,t-1} - x'_{it}(\widehat{\beta} - \beta)$ and compute proper standard errors with an analytical correction term.

[15] Following Kripfganz and Schwarz (2019), we use successively these two Stata commands ("xtdpdqml" and "xtseqreg"). Unfortunately, these Stata commands do not give the residual variance of specific effects σ^2_μ but only σ^2_u.

in a static model (see Baltagi et al. (2018)). Finally, note that the *2S bootstrap* estimator yields slightly biased estimates of σ_μ^2 but this bias decreases rapidly as the time span is increased (from 17.5% for $T = 10$ to 2.5% for $T = 30$).

3.1.4. The dynamic homogeneous panel data world with common trends

The dynamic homogeneous panel data world with common trends is defined as:

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta_1 + x_{i,t-1}\beta_2 + f_t'\gamma_i + u_{it}$$

Since the m common trends, f_t, are known, we can rewrite the general dynamic model (8) as follows:

$$\begin{aligned} y =& Z\theta + Wb + u = Z\theta + F\Gamma + u \\ & \text{with } Z_{it}' = \left[y_{it-1}, X_{it}'\right] \text{ , } \theta' = \left[\rho, \beta'\right]' \text{ and } X_{it}' = \left[x_{i,t}, x_{i,t-1}\right], \end{aligned}$$

where $u \sim N(0, \Sigma)$, $\Sigma = \tau^{-1}I_N$. The $(N(T-1) \times Nm)$ matrix F of the m common trends is a blockdiagonal matrix where each $(T-1 \times m)$ sub block f is replicated N times and Γ is the $(Nm \times 1)$ individual varying coefficients vector:

$$F = I_N \otimes f \text{ with } f = \begin{pmatrix} f_{21} & \dots & f_{2m} \\ \dots & \dots & \dots \\ f_{T1} & \dots & f_{Tm} \end{pmatrix} \text{ and } \Gamma = vec\begin{pmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{N1} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{N2} \\ \dots & \dots & \dots & \dots \\ \gamma_{1m} & \gamma_{2m} & \cdots & \gamma_{Nm} \end{pmatrix}$$

This model is usually estimated using the common correlated effects pooled estimator (CCEP) (see Pesaran (2006) and Chudik and Pesaran (2015a,b)). It can also be estimated using the quasi-maximum likelihood (QML) estimator. We compare our *2S bootstrap* estimator with the CCEP estimator.[16] We chose samples in which the time span is large $T = 30$ or $T = 50$ with small ($N = 50$) or medium ($N = 100$) number of individuals (in the spirit of Chudik and Pesaran (2015a) who vary N and T between 40 and 200 in their simulations).

Table 4 shows that the results of the *2S bootstrap* estimator are close to those of the CCEP estimator. The results on RMSEs are mixed. *2S bootstrap* gives a lower RMSE for ρ than CCEP for $N = 100$ and $T = 30$, *i.e.* 0.002 compared to 0.007, but this difference is reduced for $N = 50$ and $T = 50$, *i.e.* 0.0058 compared to 0.0066. CCEP gives a lower RMSE for β_1. For β_2, it depends on the sample sizes. Finally, *2S bootstrap* gives a lower RMSE for σ_u^2. The computation time is a bit longer with our estimator given the bootstrap procedure. However, all estimators yield essentially the same parameter estimates, no matter what sample size.

3.1.5. The dynamic homogeneous panel data world with correlated common effects

Again, this model is usually estimated using the common correlated effects pooled estimator (CCEP) (see Pesaran (2006); Chudik and Pesaran (2015a,b)) or with the principal components estimators using quasi-maximum likelihood (QML) estimator (see Bai (2009) or Song (2013)). Since the m common correlated effects, f_t, are unknown, we need to rewrite the general dynamic

[16] We use our own R codes for our Bayesian estimator and the "xtdcce2" Stata package for the CCEP estimator. We use the same DGP set under R and Stata environments to compare the two methods.

model (8) as follows:

$$
\begin{aligned}
y =& Z\theta + Wb + u = Z\theta + F\Gamma + u \\
& \text{with } Z'_{it} = [y_{it-1}, X'_{it}] \ , \theta' = [\rho, \beta']' \text{ and } X'_{it} = [x_{i,t}, x_{i,t-1}] \, ,
\end{aligned}
$$

where the $(N(T-1) \times Nm)$ matrix F of the m unobserved factors is still a blockdiagonal matrix where each $((T-1) \times m)$ sub block f is replicated N times but f should be approximated by known variables. Similar to the Hausman-Taylor case (see eq(24)), we can approximate the $((T-1) \times m)$ f matrix with a $((T-1) \times K_1)$ f^* matrix of the within time transformation[17] of Z_{it}:

$$
f^* = \begin{pmatrix} f_2^* \\ \dots \\ f_T^* \end{pmatrix} \qquad \text{where } f_t^* = \left[\left(\overline{y}_{-1,t} - \overline{\overline{y}}_{-1} \right), \left(\overline{x}_t - \overline{\overline{x}} \right), \left(\overline{x}_{-1,t} - \overline{\overline{x}}_{-1} \right) \right]
$$

$$
\text{with } \overline{x}_t = (1/N) \sum_{i=1}^{N} x_{it}, \, \overline{\overline{x}} = (1/NT) \sum_{i=1}^{N} \sum_{t=2}^{T} x_{it},
$$

or as Chudik and Pesaran (2015a) by the time means of the dependent and explanatory variables: $f_t^* = \left[\overline{y}_t, \overline{y}_{-1,t}, \overline{x}_t, \overline{x}_{-1,t} \right]$.[18] We follow the method of Chudik and Pesaran (2015a,b) by introducing the time means of the dependent and explanatory variables instead of introducing only the within time transformation of the explanatory variables Z'_{it}. Then, the product $F\Gamma$ is approximated with the product $F^*\Gamma^*$ where the factor loadings Γ^* is a $(NK_1 \times 1)$ vector and F^* is a $(N(T-1) \times NK_1)$ matrix of the time means of Y and Z.

Table 5 shows that the results of the *2S bootstrap* are very close to those of the dynamic CCEP estimator. However, the RMSE is smaller for CCEP than *2S bootstrap*, but not by much for most parameters.

3.1.6. The dynamic heterogeneous panel data world with correlated common effects

The dynamic heterogeneous panel data world with common factors is defined as:

$$
y_{it} = \rho_i y_{i,t-1} + x_{it}\beta_{1i} + x_{i,t-1}\beta_{2i} + f'_t\gamma_i + u_{it} = Z'_{it}\theta_i + f'_t\gamma_i + u_{it}
$$

where $Z'_{it} = [y_{it-1}, X'_{it}]$, $\theta'_i = [\rho_i, \beta'_i]'$ and $X'_{it} = [x_{i,t}, x_{i,t-1}]$. This model is usually estimated using the common correlated effects mean group estimator (CCEMG) (see Pesaran (2006) and Chudik and Pesaran (2015a,b)). It could also be estimated using the quasi-maximum likelihood (QML) estimator. So we compare the mean coefficients $\widehat{\overline{\theta}} = (1/N)\sum_{i=1}^{N} \widehat{\theta}_i$ of our *2S bootstrap* estimator with the CCEMG estimator.[19]

While the bottom panel of Table 6 gives insights on the distribution of ρ_i and β_{1i} for different sample sizes, the top panel of Table 6 gives the estimated values of the mean coefficients $\overline{\rho}$, $\overline{\beta}_1$, the estimated values of β_2 and σ_u^2, their standard deviations and their RMSE's. Table 6 shows that

[17] *i.e.*, the demeaned time means.

[18] The dynamic CCEP estimator is defined as: $y_{it} = \rho y_{i,t-1} + x_{it}\beta_1 + x_{i,t-1}\beta_2 + \sum_{j=0}^{p_T} f^*_{t-j}\gamma_{i,j} + u_{it}$ where $p_T = T^{1/3}$ (see Chudik and Pesaran (2015b) pp. 26). Then, $p_T \approx 3$ when $T = 30$ or $T = 50$. In the simulations, we use $p_T = 0$.

[19] We use our own R codes for our Bayesian estimator and the "xtdcce2" Stata package for the CCEMG estimator. We use the same DGP set under R and Stata environments to compare the two methods.

the results of the *2S bootstrap* estimator are close to those of the CCEMG estimator. The RMSEs results are mixed. *2S bootstrap* gives a smaller RMSE for ρ than CCEMG, but CCEMG gives a smaller RMSE for $\overline{\beta}_1$ and σ_u^2. The results for β_2 depend on the sample size. For $\overline{\beta}_1$ the bias is (7.64% for $N = 100$, $T = 30$), (resp. 16.47% for $N = 50$, $T = 50$) for the *2S bootstrap* estimator as compared to those of the CCEMG estimator (4.37%, resp.(10.21%)). For the residuals' variance σ_u^2, the bias increases with the time dimension for both estimators. However, all the estimators yield roughly the same parameter estimates. Computation time is a bit longer with our estimator given the bootstrap procedure.

3.1.7. Sensitivity to ε-contamination values

Tables B.5 and B.6 (in the supplementary material) investigate the sensitivity of the *2S bootstrap* estimator for the random effects world and for the heterogeneous panel data world with correlated common effects[20] with respect to ε, the contamination part of the prior distributions, which varies between 0 and 90%. As shown in Table B.5 for the random effects world when $N = 100$ and $T = 10$, all the parameter estimates are insensitive to ε. The only noteworthy change concerns the estimated value of $\lambda_\mu (\equiv \lambda_{b,h_0})$. It more or less corresponds to $(1-\varepsilon)$. This particular relation may occur whenever $\widehat{h}/\left(\widehat{h}+1\right) = h_0/(h_0+1)$ and $R_{b_0}^2/\left(1-R_{b_0}^2\right) = R_{\widehat{b}_q}^2/\left(1-R_{\widehat{b}_q}^2\right)$ (see the definition of $\widehat{\lambda}_{b,h_0}$ in section 2.3.2). The observed stability of the coefficients estimates stems from the fact that the base prior is not consistent with the data as the weight $\widehat{\lambda}_\theta \to 0$. The ML-II posterior mean of θ is thus close to the posterior $\widehat{q}^*(\theta \mid g_0)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}(\mu \mid g_0)$. Hence, the numerical value of the ε-contamination, for $\varepsilon \neq 0$, does not seem to play an important role in our simulated worlds. Table B.5 also reports the results when ε is very close to zero ($\varepsilon = 10^{-17}$) and we get similar results. Lastly, we have also checked the extreme case when $\varepsilon = 0$. The restricted ML-II estimator ($\varepsilon = 0$) constrains the model to rely exclusively on a base elicited prior which is implicitly assumed error-free. This is a strong assumption. This time, results are not strictly similar to those of $\varepsilon \neq 0$ but they are close to the true values except for σ_μ^2 which has a fairly large upward bias (11.4%) as well as a large RMSE.[21]

Table B.6 shows similar results. All the parameter estimates are insensitive to ε ($\varepsilon \neq 0$) for the heterogeneous panel data world with correlated common effects when $N = 100$ and $T = 30$. The only changes concern the estimated values of λ_{θ,g_0} and $\lambda_\mu (\equiv \lambda_{b,h_0})$. While $\widehat{\lambda}_\mu (\equiv \widehat{\lambda}_{b,h_0})$ changes inversely to ε, $\widehat{\lambda}_{\theta,g_0}$ has the shape of an inverted J as ε increases. As for the random effects world, when $\varepsilon = 0$, the results are not strictly similar to those of $\varepsilon \neq 0$ but they are close to the true values except for σ_μ^2 which has also a fairly large upward bias (14.5%) as well as large standard error and RMSE. Whatever the world tested, results are insensitive to the exact value of $\varepsilon \neq 0$. This stems from the fact that the *2S bootstrap* estimator is data driven and implicitly adjusts the weights to the different values of ε-contamination. This may by why, even though the choice of $\varepsilon = 0.5$ is somewhat arbitrary, the adjustment compensates for it not being optimal (see Berger (1985)).

[20] This exercise could be conducted for the other worlds as Chamberlain-type fixed effects or Hausman-Taylor world but we report the results for only two worlds for the sake of brevity.

[21] From a theoretical point of view, and under the null, $H_0 : \varepsilon = 0$, it follows that the weights $\widehat{\lambda}_{\theta,g_0} = 1$ and $\widehat{\lambda}_{b,h_0} = 1$ so that the restricted ML-II estimator of θ is given by $\widehat{\theta}_{restrict} = \theta_*(b \mid g_0)$. Under $H_1 : \varepsilon \neq 0$ the unrestricted estimator is $\widehat{\theta}_{unrestrict}\left(\equiv \widehat{\theta}_{ML-II}\right) = \widehat{\lambda}_{\theta,g_0}\theta_*(b \mid g_0) + \left(1-\widehat{\lambda}_{\theta,g_0}\right)\theta_{EB}(b \mid g_0)$. The restricted ML-II estimator $\theta_*(b \mid g_0)$ is the Bayes estimator under the base prior g_0.

3.1.8. Departure from normality

Tables B.7 and B.8 (in the supplementary material) investigate the robustness of the estimators to a non-normal framework for the random effects world and for the heterogeneous panel data world with correlated common effects. The remainder disturbances, u_{it}, are now assumed to follow a right-skewed t-distribution with mean = 0, degrees of freedom $\nu = 3$, and skewing parameter $\gamma = 2$ (see Fernández and Steel (1998), Baltagi et al. (2018)).[22] Our *2S bootstrap* estimator behaves pretty much like the GMM and the QMLE for the random effects world when $N = 100$ and $T = 10$ (see Table B.7). There is, however, a slight downward bias for the ρ coefficient (-4.7%) and a slight upward bias for β_{11} and β_2 (7%). But these biases are small. Compared to the GMM estimator, our *2S bootstrap* estimator provides better estimates of σ_u^2 and σ_μ^2 but it is the QML estimator that gives the estimates closest to the true values. Another interesting result concerns the standard errors and RMSEs of all the estimators. The presence of a right-skewed t-distribution greatly increases these values especially for σ_u^2.

Table B.8 investigates the robustness of the CCEMG and *2S bootstrap* estimators to the right-skewed t-distribution for the heterogeneous panel data world with correlated common effects when $N = 100$ and $T = 30$. There are slight downward biases for the $\overline{\rho}$ mean coefficient, with that of CCEMG being larger than that of *2S bootstrap* (-16% *vs* -7.6%) as well as slight upward bias for σ_u^2, that of CCEMG being larger than that of *2S bootstrap* (5.6% *vs* 0.5%). However, for the $\overline{\beta}_1$ mean coefficient, it is the *2S bootstrap* estimator which has a slightly larger bias (17.2% *vs* 7.5%). Finally, it can be noted that the RMSE of σ_u^2 is larger for the CCEMG estimator than for the *2S bootstrap* estimator.

4. Conclusion

To our knowledge, our paper is the first to analyze the dynamic linear panel data model using an ϵ-contamination approach within a two-stage hierarchical approach. The main benefit of this approach is its ability to extract more information from the data than the classical Bayes estimator with a single base prior. In addition, we show that our approach encompasses a variety of classical or frequentist specifications. We estimate the Type-II maximum likelihood (ML-II) posterior distribution of the slope coefficients and the individual effects using a two-step procedure. The posterior distribution is a convex combination of the conditional posterior densities derived from the elicited prior and the ϵ-contaminated prior. Thus if the base prior is consistent with the data, more weight is given to the conditional posterior density derived from the former. Otherwise, more weight is given to the latter.

The finite sample performance of the two-stage hierarchical models is investigated using extensive Monte Carlo experiments. The experimental design includes a random effects world, a Chamberlain type fixed effects world, a Hausman–Taylor-type world and worlds with homogeneous/heterogeneous slopes and cross-sectional dependence. The simulation results underscore the relatively good performance of the two-stage hierarchy estimator, irrespective of the data generating process considered. The biases and the RMSEs are close and sometimes smaller than those of

[22] The Skewed t distribution with ν degrees of freedom and skewing parameter γ has the following density:

$$pdf(x) = \frac{2}{\gamma + \frac{1}{\gamma}} f(z) \text{ where } z = \gamma x \text{ if } x < 0 \text{ or } z = x/\gamma \text{ if } x \geq 0$$

where $f(.)$ is the density of the t distribution with ν degrees of freedom.

the conventional (classical) estimators. We also investigate the sensitivity of the estimators to the contamination part of the prior distribution. It turns out that parameter estimates are relatively stable. Finally, we investigate the robustness of the estimators when the remainder disturbances are assumed to follow a right-skewed t-distribution. Compared to classical estimators, our robust estimators globally behave well in terms of precision and bias.

The robust Bayesian approach we propose is arguably a useful all-in-one panel data estimator. Because it embeds a variety of estimators, it can be used straightforwardly to estimate dynamic panel data models under many alternative stochastic specifications. Unlike classical estimators, there is no need to have a custom estimator for each possible DGP. Furthermore, it allows one to circumvent the difficulties faced by analysts who are oftentimes constrained to use those estimators that are readily available in standard software suites.

We reckon that our estimator contributes only marginally to those already available in the literature. Our main contribution is to propose an estimator that allows the analysts to focus on the stochastic specification of their model rather than finding the software best suited to their needs. This is because our estimator is easily amenable to many specifications in addition to those already presented in this paper. These include models with individual and time random effects in unknown common factors, spatial structures (autoregressive spatial), *etc.* We leave these for future work.

Table 1: Dynamic Random Effects World

$\varepsilon = 0.5$, $r = 0.8$, Replications=$1,000$

		ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1	1	4			
$N = 100$	2S boot coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	0.4943	521.89
$T = 10$	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			
	GMM coef	0.7554	0.9912	0.9135	0.9886	0.8837	4.6186			149.15
	se	0.0067	0.0273	0.0266	0.0279	0.0541	0.3857			
	rmse	0.0115	0.0392	0.0941	0.0403	0.1282	0.7289			
	QMLE coef	0.7474	1.0101	0.9761	1.0073	0.9925	4.1236			457.25
	se	0.0058	0.0255	0.0251	0.0252	0.0508	0.2499			
	rmse	0.0063	0.0274	0.0274	0.0274	0.0513	0.2787			
$N = 200$	2S boot coef	0.7521	1.0080	1.0060	1.0058	1.0003	4.0906	$< 10^{-4}$	0.4975	6054.19
$T = 30$	se	0.0026	0.0090	0.0090	0.0091	0.0186	0.0732			
	rmse	0.0028	0.0119	0.0108	0.0108	0.0186	0.1165			
	GMM coef	0.7645	0.9398	0.9222	0.9265	0.9927	3.8460			5305.03
	se	0.0024	0.0085	0.0088	0.0085	0.0272	0.2010			
	rmse	0.0159	0.0644	0.0812	0.0768	0.0282	0.2531			
	QMLE coef	0.7493	1.0006	0.9991	0.9993	0.9988	4.1385			2472.06
	se	0.0021	0.0088	0.0091	0.0093	0.0188	0.0815			
	rmse	0.0022	0.0088	0.0088	0.0088	0.0189	0.1607			

2S boot: two stage with individual block resampling bootstrap.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table 2: Dynamic Chamberlain-type Fixed Effects World ($\varepsilon = 0.5$)

		ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1	1	162.5341			
$N = 100$	2S boot coef	0.7414	1.0134	1.0132	0.9959	0.9952	173.8813	$< 10^{-4}$	0.4959	686.26
$T = 10$	se	0.0061	0.0234	0.0236	0.0231	0.0665	36.1240			
Replications=$1,000$	rmse	0.0105	0.0277	0.0274	0.0241	0.0666	37.8470			
	QMLE coef	0.7507	0.9979	0.9982	0.9988	1.0057	162.1588			1199.07
	se	0.0136	0.0426	0.0416	0.0247	0.1384	25.1817			
	rmse	0.01368	0.0426	0.0426	0.0426	0.1385	25.1721			
	true	0.75	1	1	1	1	204.3142			
$N = 200$	2S boot coef	0.7472	1.0060	1.0059	0.9992	1.0005	208.6873	$< 10^{-4}$	0.4984	4525.51
$T = 30$	se	0.0023	0.0086	0.0087	0.0091	0.0267	28.4166			
Replications=500$^{(*)}$	rmse	0.0036	0.0110	0.0107	0.0094	0.0266	28.7231			
	QMLE coef	0.7499	0.9999	0.9998	0.9995	0.9995	204.4599			4059.55
	se	0.0022	0.0090	0.0086	0.0095	0.0191	20.8648			
	rmse	0.0022	0.0090	0.0090	0.0090	0.0191	20.8444			

2S boot: two stage with individual block resampling bootstrap.

The parameters π_t are omitted from the table, see Table B.3 in appendix B in the supplementary material.

QMLE: quasi-maximum likelihood estimation.

The parameters π_t are omitted from the table, see Table B.4 in appendix B in the supplementary material.

$^{(*)}$ *When $T = 30$, we restrict the exercise to only 500 replications, not because of our estimator under R but because of the size of the simulated database and its reading under Stata. Indeed, for $1,000$ replications and $T = 30$, one must read, under Stata, $(1 + 4 + 30) \times 1,000 = 35,000$ variables of size $(NT, 1)$! Even with a 64-bit computer and Stata (MP and S), only $32,767$ variables can be read. On the other hand, with our code under R, there is no limitation of that order.*

Table 3: Dynamic Hausman-Taylor World

$\rho = 0.8$, $\varepsilon = 0.5$, replications=1,000

		ρ	β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	True	0.75	1	1	1	1	1	1	4			
$N = 100$	2S boot coef	0.7373	1.0081	1.0069	1.0066	1.0734	1.1687	0.9897	4.7146	$< 10^{-4}$	0.4999	568.02
$T = 10$	se	0.0091	0.0327	0.0327	0.0359	0.2291	0.1200	0.0631	1.1883			
	rmse	0.0156	0.0337	0.0335	0.0365	0.2405	0.2071	0.0639	1.3861			
	two-stage QML coef	0.7498	1.0003	0.9992	1.0011	1.0215	1.0007	0.9935	$n.a$			619.27
	se	0.0085	0.0278	0.0272	0.0275	0.2090	0.2189	0.0480	$n.a$			
	rmse	0.0085	0.0278	0.0278	0.0278	0.2100	0.2188	0.0485	$n.a$			
$N = 100$	2S boot coef	0.7463	1.0036	1.0044	1.0041	1.0340	1.0473	0.9973	4.1293	$< 10^{-4}$	0.5005	1650.26
$T = 30$	se	0.0044	0.0173	0.0172	0.0179	0.2021	0.0581	0.0383	0.7478			
	rmse	0.0057	0.0177	0.0177	0.0184	0.2049	0.0749	0.0384	0.7585			
	two-stage QML coef	0.7499	0.9999	1.0004	1.0005	1.0193	0.9963	0.9978	$n.a$			2423.60
	se	0.0032	0.0131	0.0127	0.0131	0.1999	0.1441	0.0273	$n.a$			
	rmse	0.0032	0.0132	0.0132	0.0132	0.2007	0.1441	0.0274	$n.a$			
$N = 200$	2S boot coef	0.7377	1.0091	1.0088	1.0047	1.0474	1.1611	0.9932	4.7739	$< 10^{-4}$	0.4999	2137.42
$T = 10$	se	0.0065	0.0236	0.0231	0.0255	0.1629	0.0840	0.0472	0.8531			
	rmse	0.0140	0.0253	0.0247	0.0259	0.1696	0.1817	0.0476	1.1515			
	two-stage QML coef	0.7500	1.0004	1.0005	0.9997	0.9987	1.0049	0.9965	$n.a$			1260.29
	se	0.0060	0.0200	0.0190	0.0198	0.1468	0.1457	0.0357	$n.a$			
	rmse	0.0061	0.0200	0.0200	0.0200	0.1468	0.1457	0.0359	$n.a$			

2S boot: two stage with individual block resampling bootstrap.

two-stage QML: two-stage quasi-maximum likelihood sequential approach with non available ($n.a$) estimate of σ_μ^2.

Table 4: Dynamic homogeneous panel data model with common trends

$\varepsilon = 0.5$, Replications=1,000

		ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1			
$N = 100$	2S boot coef	0.7490	0.9943	1.0134	1.0025	$< 10^{-4}$	0.4594	3744.05
$T = 30$	se	0.0022	0.0255	0.0326	0.0265			
	rmse	0.0024	0.0261	0.0353	0.0266			
	CCEP coef	0.7441	1.0092	1.0312	0.9903			1491.96
	se	0.0038	0.0225	0.0272	0.0270			
	rmse	0.0070	0.0243	0.0414	0.0286			
$N = 50$	2S boot coef	0.7469	1.0085	1.0178	1.0055	$< 10^{-4}$	0.4613	3502.82
$T = 50$	se	0.0049	0.0312	0.0376	0.0316			
	rmse	0.0058	0.0323	0.0416	0.0321			
	CCEP coef	0.7447	1.0045	1.0204	0.9800			1117.15
	se	0.0041	0.0245	0.0290	0.0285			
	rmse	0.0066	0.0249	0.0355	0.0348			

2S boot: two stage with individual block resampling bootstrap.

CCEP: Common Correlated Effects Pooled estimator.

Table 5: Dynamic homogeneous panel data model with common correlated effects

$\varepsilon = 0.5$, Replications=1,000

		ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1			
$N = 100$	2S boot coef	0.7439	1.0173	1.0279	1.0240	$< 10^{-4}$	0.4809	4706.92
$T = 30$	se	0.0060	0.0328	0.0390	0.0316			
	rmse	0.0086	0.0371	0.0480	0.0397			
	CCEP coef	0.7445	1.0040	1.0233	1.0211			1381.78
	se	0.0043	0.0231	0.0284	0.0314			
	rmse	0.0069	0.0234	0.0367	0.0378			
$N = 50$	2S boot coef	0.7469	1.0113	1.0126	1.0184	$< 10^{-4}$	0.4865	3858.11
$T = 50$	se	0.0052	0.0330	0.0393	0.0341			
	rmse	0.0061	0.0349	0.0413	0.0387			
	CCEP coef	0.7475	1.0033	1.0107	1.0147			1101.13
	se	0.0040	0.0248	0.0299	0.0324			
	rmse	0.0047	0.0250	0.0317	0.0356			

2S boot: two stage with individual block resampling bootstrap.

CCEP: Common Correlated Effects Pooled estimator.

Table 6: Dynamic heterogeneous panel data model with common correlated effects

$\varepsilon = 0.5$, Replications=1,000

		$\overline{\rho}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.7501	0.7498	1	1			
$N = 100$	2S boot coef	0.7347	0.8071	1.0759	1.0536	0.0011	0.3925	4695.86
$T = 30$	se	0.0194	0.0671	0.0531	0.0392			
	rmse	0.0248	0.0882	0.0926	0.0664			
	CCEMG coef	0.7205	0.7826	1.0856	1.0409			1618.05
	se	0.0221	0.0554	0.0453	0.0353			
	rmse	0.0369	0.0600	0.0968	0.0540			
	true	0.7501	0.7503	1	1			
$N = 50$	2S boot coef	0.7474	0.8739	1.0666	1.0838	0.0010	0.4407	2492.75
$T = 50$	se	0.0217	0.0973	0.0643	0.0474			
	rmse	0.0219	0.1572	0.0925	0.0963			
	CCEMG coef	0.7423	0.8269	1.0654	1.1450			1145.23
	se	0.0233	0.0848	0.0619	0.2180			
	rmse	0.0245	0.1159	0.0900	0.2618			

2S boot: two stage with individual block resampling bootstrap.

CCEMG: Common Correlated Effects Mean Group estimator.

mean coefficients: $\overline{\rho} = (1/N)\sum_{i=1}^{N}\rho_i$ and $\overline{\beta}_1 = (1/N)\sum_{i=1}^{N}\beta_{1i}$.

	$N = 100, T = 30$		$N = 50, T = 50$	
	ρ_i	β_{1i}	ρ_i	β_{1i}
min	0.6030	0.5052	0.6059	0.5095
mean	0.7501	0.7498	0.7501	0.7503
sd	0.0865	0.1442	0.0863	0.1442
max	0.8970	0.9951	0.8942	0.9905

References

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. Journal of Econometrics 68, 5–27.

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.

Anderson, T.W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18, 47–82.

Andersson, M., Karlsson, S., 2001. Bootstrapping error component models. Computational Statistics 16, 221–231.

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

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

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2018. Robust linear static panel data models using ε-contamination. Journal of Econometrics 202, 108–123.

Bauwens, L., Lubrano, M., Richard, J.F., 2005. Bayesian Inference in Dynamic Econometric Models. Advanced Text in Econometrics. Oxford University Press, Oxford, UK.

Bazzi, S., Clemens, M.A., 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. American Economic Journal: Macroeconomics 5, 152–86.

Bellman, L., Breitung, J., Wagner, J., 1989. Bias correction and bootstrapping of error component models for panel data: Theory and applications. Empirical Economics 14, 329–342.

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J., Berliner, M., 1984. Bayesian input in Stein estimation and a new minimax empirical Bayes estimator. Journal of Econometrics 25, 87–108.

Berger, J., Berliner, M., 1986. Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Annals of Statistics 14, 461–486.

Bhargava, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. Econometrica , 1635–1659.

Binder, M., Hsiao, C., Pesaran, M.H., 2005. Estimation and inference in short panel vector autoregressions with unit roots and cointegration. Econometric Theory 21, 795–837.

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

Bretó, C., Ionides, E.L., King, A.A., 2019. Panel data analysis via mechanistic models. Journal of the American Statistical Association *(just accepted)* , 1–35.

Bun, M.J.G., 2003. Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. Econometric Reviews 22, 29–58.

Bun, M.J.G., Carree, M.A., Juodis, A., 2017. On maximum likelihood estimation of dynamic panel data models. Oxford Bulletin of Economics and Statistics 79, 463–494.

Bun, M.J.G., Kiviet, J.F., 2003. On the diminishing returns of higher-order terms in asymptotic expansions of bias. Economics Letters 79, 145–152.

Bun, M.J.G., Sarafidis, V., 2015. Dynamic panel data models, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 76–110.

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18, 5–46.

Chib, S., 2008. Panel data modeling and inference: a Bayesian primer, in: Mátyás, L., Sevestre, P. (Eds.), The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice, pp. 479–516.

Chib, S., Carlin, B.P., 1999. On MCMC sampling in hierarchical longitudinal models. Statistics and Computing 9, 17–26.

Chudik, A., Pesaran, M.H., 2015a. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393–420.

Chudik, A., Pesaran, M.H., 2015b. Large panel data models with cross-sectional dependence: A survey, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 3–45.

Congdon, P.D., 2010. Applied Bayesian hierarchical methods. Chapman and Hall/CRC.

DeJong, D.N., Whiteman, C.H., 1991. Reconsidering 'trends and random walks in macroeconomic time series'. Journal of Monetary Economics 28, 221–254.

Dempster, A.P., 1977. Examples relevant to the robustness of applied inference, in: Gupta, S.S., Moore, D.S. (Eds.), Statistical Decision Theory and Related Topics II. Academic Press, New York, pp. 121–138.

Dhaene, G., Jochmans, K., 2011. An adjusted profile likelihood for non-stationary panel data models with fixed effects. Technical Report. hal-01073732.

Dhaene, G., Jochmans, K., 2016. Likelihood inference in an autoregression with fixed effects. Econometric Theory 32, 1178–1215.

Dorsett, R., 1999. An econometric analysis of smoking prevalence among lone mothers. Journal of Health Economics 18, 429–441.

Everaert, G., 2013. Orthogonal to backward mean transformation for dynamic panel data models. The Econometrics Journal 16, 179–221.

Everaert, G., Pozzi, L., 2007. Bootstrap-based bias correction for dynamic panels. Journal of Economic Dynamics and Control 31, 1160–1184.

Fernández, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100, 381–427.

Fernández, C., Steel, M.F.J., 1998. On Bayesian modeling of fat tails and skewness. Journal of The American Statistical Association 93, 359–371.

Geweke, J., Keane, M., 2000. An empirical analysis of earnings dynamics among men in the PSID: 1968-1989. Journal of Econometrics 96, 293–256.

Ghosh, M., Heo, J., 2003. Default Bayesian priors for regression models with first-order autoregressive residuals. Journal of Time Series Analysis 24, 269–282.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1997. Markov Chain Monte Carlo in Practice. 2nd ed., Chapman & Hall, London, UK.

Good, I.J., 1965. The Estimation of Probabilities. MIT Press, Cambridge, MA.

Greenberg, E., 2008. Introduction to Bayesian Econometrics. Cambridge University Press, Cambridge, UK.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics 14, 262–280.

Harris, M., Mátyás, L., Sevestre, P., 2008. Dynamic models for short panels, in: Mátyás, L., Sevestre, P. (Eds.), The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice, pp. 249–278.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.

Hayakawa, K., Pesaran, M.H., 2015. Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. Journal of Econometrics 188, 111–134.

Hill, B., 1980. Invariance and robustness of the posterior distribution of characteristics of a finite population with reference to contingency tables and the sampling of species, in: Zellner, A. (Ed.), Bayesian Analysis in Econometrics and Statistics, Essays in Honor of Harold Jeffreys. North-Holland, Amsterdam, pp. 383–391.

Hirano, K., 2002. Semiparametric Bayesian inference in autoregressive panel data models. Econometrica 70, 781–799.

Hjellvik, V., Tjstheim, D., 1999. Modelling panels of intercorrelated autoregressive time series. Biometrika 86, 573–590.

Hsiao, C., Pesaran, M.H., 2008. Random coefficient models, in: The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice. Mátyás, L. and Sevestre, P., pp. 185–214.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 1999. Bayes estimation of short-run coefficients in dynamic panel data models, in: Hsiao, C., Lahiri, K., Lee, L.F., Pesaran, M. (Eds.), Analysis of Panel Data and Limited Dependent Variable Models. Cambridge University Press, pp. 268–297.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics 109, 107–150.

Ibazizen, M., Fellag, H., 2003. Bayesian estimation of an AR(1) process with exponential white noise. Statistics 37, 365–372.

Juárez, M.A., Steel, M.F.J., 2010. Non-Gaussian dynamic Bayesian modelling for panel data. Journal of Applied Econometrics 25, 1128–1154.

Kapetanios, G., 2008. A bootstrap procedure for panel datasets with many cross-sectional units. Econometrics Journal 11, 377–395.

Karakani, H.M., van Niekerk, J., van Staden, P., 2016. Bayesian analysis of AR(1) model. arXiv preprint arXiv:1611.08747 .

Kass, R.E., Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. Journal of the American Statistical Association 90, 928–934.

Kiviet, J.F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 53–78.

Koop, G., 2003. Bayesian Econometrics. Wiley, New York.

Koop, G., Leon-Gonzalez, R., Strachan, R., 2008. Bayesian inference in a cointegrating panel data model, in: Advances in Econometrics. Bayesian Econometrics. Emerald Group Publishing Limited, pp. 433–469.

Kraay, A., 2015. Weak instruments in growth regressions: implications for recent cross-country evidence on inequality and growth. Policy Research Working Paper 7494. The World Bank.

Kripfganz, S., 2016. Quasi-maximum likelihood estimation of linear dynamic short-T panel-data models. The Stata Journal 16, 1013–1038.

Kripfganz, S., Schwarz, C., 2019. Estimation of linear dynamic panel data models with time-invariant regressors. Journal of Applied Econometrics 34, 526–546.

Lancaster, T., 2002. Orthogonal parameters and panel data. The Review of Economic Studies 69, 647–666.

Lindley, D., Smith, A., 1972. Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B (Methodological), 34, 1–41.

Liu, F., Zhang, P., Erkan, I., Small, D.S., 2017. Bayesian inference for random coefficient dynamic panel data models. Journal of Applied Statistics 44, 1543–1559.

Liu, L., Moon, H.R., Schorfheide, F., 2018. Forecasting with dynamic panel data models. Technical Report. National Bureau of Economic Research.

Liu, L.M., Tiao, G.C., 1980. Random coefficient first-order autoregressive models. Journal of Econometrics 13, 305–325.

Moon, H.R., Perron, B., Phillips, P.C.B., 2015. Incidental parameters and dynamic panel modeling, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 111–148.

Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543–1579.

Moon, H.R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33, 158–195.

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

Moral-Benito, E., 2013. Likelihood-based estimation of dynamic panels with predetermined regressors. Journal of Business & Economic Statistics 31, 451–472.

Moral-Benito, E., Allison, P., Williams, R., 2019. Dynamic panel data modelling using maximum likelihood: an alternative to Arellano-Bond. Applied Economics 51, 2221–2232.

Mundlak, Y., 1978. On the pooling of time series and cross-section data. Econometrica 46, 69–85.

Pacifico, A., 2019. Structural panel Bayesian VAR model to deal with model misspecification and unobserved heterogeneity problems. Econometrics 7, 1–24.

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

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

Rendon, S.R., 2013. Fixed and random effects in classical and Bayesian regression. Oxford Bulletin of Economics and Statistics, 75, 460–476.

Robert, C.P., 2007. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. 2nd ed., Springer, New York, USA.

Rubin, H., 1977. Robust Bayesian estimation, in: Gupta, S.S., Moore, D.S. (Eds.), Statistical Decision Theory and Related Topics II. Academic Press, New York, pp. 351–356.

Schotman, P.C., van Dijk, H.K., 1991. On Bayesian routes to unit roots. Journal of Applied Econometrics 6, 387–401.

Sims, C.A., Uhlig, H., 1991. Understanding unit rooters: A helicopter tour. Econometrica , 1591–1599.

Smith, A., 1973. A general Bayesian linear model. Journal of the Royal Statistical Society, Series B (Methodological) 35, 67–75.

Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Technical Report. Mimeo, January.

Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. Journal of Econometrics 185, 230–258.

Tsai, T.H., 2016. A Bayesian approach to dynamic panel models with endogenous rarely changing variables. Political Science Research and Methods 4, 595–620.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

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

Yu, J., De Jong, R., Lee, L.F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118–134.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distribution, in: Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, Studies in Bayesian Econometrics. North-Holland, Amsterdam. volume 6, pp. 389–399.

Zheng, Y., Zhu, J., Li, D., 2008. Analyzing spatial panel data of cigarette demand: a Bayesian hierarchical modeling approach. Journal of Data Science 6, 467–489.

Robust dynamic panel data models using ε-contamination

Supplementary material

Badi H. Baltagi[a,*], Georges Bresson[b], Anoop Chaturvedi[c], Guy Lacroix[d]

[a] *Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, U.S.A.*
[b] *Université Paris II, France*
[c] *University of Allahabad, India*
[d] *Départment d'économique, Université Laval, Québec, Canada*

May 4, 2020

*Corresponding author.

Email addresses: bbaltagi@maxwell.syr.edu (Badi H. Baltagi), georges.bresson@u-paris2.fr (Georges Bresson), anoopchaturv@gmail.com (Anoop Chaturvedi), Guy.Lacroix@ecn.ulaval.ca (Guy Lacroix)

Contents

A. Uniform distribution, derivation of the mean and variance of the ML-II posterior density of $\boldsymbol{\rho}$ and some Monte Carlo results

A.1. Uniform distribution and derivation of the mean and variance of the ML-II posterior density of ρ

Following Singh and Chaturvedi (2012) (see also Shrivastava et al. (2019)), and for deriving the posterior density of ρ, given (β, b), we write:

$$y^{\circ} = (y - X\beta - Wb) = \rho y_{-1} + u$$

the probability density function (pdf) of y°, given the observables and the parameters, is:

$$p\left(y^{\circ}|y_{-1}, \rho, \tau\right) = \left(\frac{\tau}{2\pi}\right)^{\frac{N(T-1)}{2}} \exp\left(-\frac{\tau}{2}(y^{\circ} - \rho y_{-1})'(y^{\circ} - \rho y_{-1})\right)$$

Let $\widehat{\rho}(\beta, b) = \left(y_{-1}' y_{-1}\right)^{-1} y_{-1}' y^{\circ} = \left(\Lambda_y\right)^{-1} y_{-1}' y^{\circ}$, then following the derivation of (eq.16) in the technical appendix (pp 6-7) of Baltagi et al. (2018), we can write:

$$(y^{\circ} - \rho y_{-1})'(y^{\circ} - \rho y_{-1}) = \varphi\left(\beta, b\right) + \Lambda_y \left\{\rho - \widehat{\rho}(\beta, b)\right\}^2$$

with

$$\varphi\left(\beta, b\right) = (y^{\circ} - \widehat{\rho}(\beta, b) y_{-1})'(y^{\circ} - \widehat{\rho}(\beta, b) y_{-1})$$

and

$$\widehat{\rho}\left(\beta, b\right) = \left(y_{-1}' y_{-1}\right)^{-1} y_{-1}' y^{\circ} = \Lambda_y^{-1} y_{-1}' y^{\circ}$$

then

$$p\left(y^{\circ}|y_{-1}, \rho, \tau\right) = \left(\frac{\tau}{2\pi}\right)^{\frac{N(T-1)}{2}} \exp\left(-\frac{\tau}{2}\left\{\varphi\left(\beta, b\right) + \Lambda_y \left\{\rho - \widehat{\rho}(\beta, b)\right\}^2\right\}\right)$$

As the precision τ is assumed to have a vague prior $p\left(\tau\right) \propto \tau^{-1}$ and as $|\rho| < 1$ is assumed to be $U\left(-1, 1\right)$, the conditional posterior density of ρ, given (β, b) is defined by:

$$\pi^{*}\left(\rho|\beta, b\right) = \frac{\int\limits_0^{\infty} p(\rho) p\left(\tau\right) p\left(y^{\circ}|y_{-1}, \rho, \tau\right) d\tau}{\int\limits_0^{\infty}\int\limits_{-1}^{1} p(\rho) p\left(\tau\right) p\left(y^{\circ}|y_{-1}, \rho, \tau\right) d\rho d\tau} \tag{A.1}$$

where

$$p\left(\rho\right) = \frac{1}{2} \text{ and } p\left(\tau\right) = \frac{1}{\tau}$$

So, the numerator of $\pi^*(\rho|\beta,b)$ can be written as:

$$\begin{aligned}\int_0^\infty p(\rho)p(\tau)\,p(y^\circ|y_{-1},\rho,\tau)\,d\tau &= \int_0^\infty \left(\frac{1}{2}\right)\left(\frac{1}{\tau}\right)\left(\frac{\tau}{2\pi}\right)^{\frac{N(T-1)}{2}} \\ &\quad\times \exp\left(-\frac{\tau}{2}\left\{\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right\}\right)d\tau \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{2\pi}\right)^{\frac{N(T-1)}{2}} \\ &\quad\times\int_0^\infty\left[\begin{array}{c}(\tau)^{\frac{N(T-1)}{2}-1} \\ \times\exp\left(-\frac{\tau}{2}\left\{\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right\}\right)\end{array}\right]d\tau\end{aligned}$$

As $\int_0^\infty \tau^{x-1}\exp\left[-\frac{\tau}{2}r\right]d\tau = (2/r)^x\Gamma(x)$, then

$$\int_0^\infty p(\rho)p(\tau)\,p(y^\circ|y_{-1},\rho,\tau)\,d\tau = \frac{\Gamma(\frac{N(T-1)}{2})}{2\,(\pi)^{\frac{N(T-1)}{2}}}\left[\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}}$$

The denominator of the conditional posterior density of ρ is

$$\int_0^\infty\int_{-1}^1 p(\rho)p(\tau)\,p(y^\circ|y_{-1},\rho,\tau)\,d\rho d\tau = \frac{\Gamma(\frac{N(T-1)}{2})}{2\,(\pi)^{\frac{N(T-1)}{2}}}\int_{-1}^1\left[\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}}d\rho$$

then, the conditional posterior density of ρ, given (β,b) is defined by:

$$\pi^*(\rho|\beta,b) = \frac{\left[\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}}}{\int_{-1}^1\left[\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}}d\rho} \tag{A.2}$$

Let us derive the denominator of the previous expression

$$
\begin{aligned}
A &= \int_{-1}^{1} \left[\varphi(\beta,b) + \Lambda_y \left\{\rho - \widehat{\rho}(\beta,b)\right\}^2\right]^{-\frac{N(T-1)}{2}} d\rho \\
&= \left[\varphi(\beta,b)\right]^{-\frac{N(T-1)}{2}} \int_{-1}^{1} \left[1 + \frac{\Lambda_y \left\{\rho - \widehat{\rho}(\beta,b)\right\}^2}{\varphi(\beta,b)}\right]^{-\frac{N(T-1)}{2}} d\rho \\
&= \left[\varphi(\beta,b)\right]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \int_{-\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt \\
&= \left[\varphi(\beta,b)\right]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[\int_{-\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))}^{0} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt + \int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt\right] \\
&= \left[\varphi(\beta,b)\right]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[\int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt + \int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt\right] \\
&= \left[\varphi(\beta,b)\right]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[I_1 + I_2\right]
\end{aligned}
$$

Now taking the transformation $\eta = \frac{t^2}{1+t^2}$, then

$$
\left(1+t^2\right) = \left(1-\eta\right)^{-1} \text{ and } dt = \frac{1}{2}\eta^{-\frac{1}{2}}\left(1-\eta\right)^{-3/2} d\eta
$$

and we obtain I_1 as[1]

$$\begin{aligned}
I_1 &= \int_0^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt = \frac{1}{2}\int_0^{\zeta_1} \eta^{-\frac{1}{2}}(1-\eta)^{\frac{N(T-1)-3}{2}} d\eta \text{ where } \zeta_1 = \frac{\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2}{\left[1+\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2\right]} \\
&= \frac{1}{2}\zeta_1^{\frac{1}{2}}\int_0^1 z^{-\frac{1}{2}}(1-\zeta_1 z)^{\frac{N(T-1)-3}{2}} dz = \frac{1}{2}\zeta_1^{\frac{1}{2}}\frac{\Gamma(1/2)}{\Gamma(3/2)} \times_2 F_1\left(-\frac{N(T-1)-3}{2};\frac{1}{2};\frac{3}{2};\zeta_1\right) \\
&= \zeta_1^{\frac{1}{2}} \times_2 F_1\left(-\frac{N(T-1)-3}{2};\frac{1}{2};\frac{3}{2};\zeta_1\right) \text{ as } \Gamma(1/2)/\Gamma(3/2)=2
\end{aligned} \tag{A.3}$$

Using the Pfaff's transformation:

$$_2F_1(a_1;a_2;a_3;z) = (1-z)^{-a_2} \times_2 F_1\left(a_3-a_1;a_2;a_3;\frac{z}{z-1}\right)$$

we obtain

$$_2F_1\left(-\frac{N(T-1)-3}{2};\frac{1}{2};\frac{3}{2};\zeta_1\right) = (1-\zeta_1)^{-\frac{1}{2}} \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};\frac{\zeta_1}{\zeta_1-1}\right)$$

Notice that

$$\frac{\zeta_1}{\zeta_1-1} = -\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2$$

Hence

$$\begin{aligned}
I_1 &= \zeta_1^{\frac{1}{2}}(1-\zeta_1)^{-\frac{1}{2}} \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};\frac{\zeta_1}{\zeta_1-1}\right) \\
&= -\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}\left(1+\widehat{\rho}(\beta,b)\right) \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2\right)
\end{aligned} \tag{A.4}$$

Similarly we obtain

$$\begin{aligned}
I_2 &= \int_0^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt = \frac{1}{2}\int_0^{\zeta_2} \eta^{-\frac{1}{2}}(1-\eta)^{\frac{N(T-1)-3}{2}} d\eta \text{ where } \zeta_2 = \frac{\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2}{\left[1+\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2\right]} \\
&= \zeta_2^{\frac{1}{2}} \times_2 F_1\left(-\frac{N(T-1)-3}{2};\frac{1}{2};\frac{3}{2};\zeta_2\right) = \zeta_2^{\frac{1}{2}}(1-\zeta_2)^{-\frac{1}{2}} \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};\frac{\zeta_2}{\zeta_2-1}\right)
\end{aligned}$$

[1] The Euler integral formula is given by:

$$\int_0^1 (t)^{a_2-1}(1-t)^{a_3-a_2-1}(1-zt)^{-a_1} dt = \frac{\Gamma(a_2)\Gamma(a_3-a_2)}{\Gamma(a_3)} \times_2 F_1(a_1;a_2;a_3;z)$$

where $_2F_1(a_1;a_2;a_3;z)$ is the Gaussian hypergeometric function with $_2F_1(a_1;a_2;a_3;z)$).

And as

$$\frac{\zeta_2}{\zeta_2 - 1} = -\frac{\Lambda_y}{\varphi(\beta, b)} (1 - \widehat{\rho}(\beta, b))^2 = \frac{\Lambda_y}{\varphi(\beta, b)} (\widehat{\rho}(\beta, b) - 1)^2$$

Hence

$$I_2 = \sqrt{\frac{\Lambda_y}{\varphi(\beta, b)}} (\widehat{\rho}(\beta, b) - 1) \times {}_2F_1\left(\frac{N(T-1)}{2}; \frac{1}{2}; \frac{3}{2}; \frac{\Lambda_y}{\varphi(\beta, b)} (\widehat{\rho}(\beta, b) - 1)^2\right) \quad \text{(A.5)}$$

Then, the denominator of the conditional posterior density of ρ, given (β, b), is

$$\begin{aligned} A &= [\varphi(\beta, b)]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta, b)}{\Lambda_y}} \left[I_1 + I_2\right] \\ &= [\varphi(\beta, b)]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta, b)}{\Lambda_y}} \left[\begin{array}{c} (\widehat{\rho}(\beta, b) - 1) \times {}_2F_1\left(\frac{N(T-1)}{2}; \frac{1}{2}; \frac{3}{2}; \frac{\Lambda_y}{\varphi(\beta,b)} (\widehat{\rho}(\beta, b) - 1)^2\right) \\ - (1 + \widehat{\rho}(\beta, b)) \times {}_2F_1\left(\frac{N(T-1)}{2}; \frac{1}{2}; \frac{3}{2}; -\frac{\Lambda_y}{\varphi(\beta,b)} (1 + \widehat{\rho}(\beta, b))^2\right) \end{array}\right] \end{aligned} \quad \text{(A.6)}$$

and the conditional posterior density of ρ, given (β, b), is

$$\begin{aligned} \pi^*(\rho|\beta, b) &= \frac{1}{A}\left[\varphi(\beta, b) + \Lambda_y \{\rho - \widehat{\rho}(\beta, b)\}^2\right]^{-\frac{N(T-1)}{2}} \\ &= \frac{\varphi(\beta, b)^{-\frac{N(T-1)}{2}}}{(AB)} B\left[1 + \frac{\Sigma^{-1}}{N(T-1) - 2} \{\rho - \widehat{\rho}(\beta, b)\}^2\right]^{-\frac{N(T-1)}{2}} \end{aligned} \quad \text{(A.7)}$$

where[2]

$$\begin{aligned} \Sigma^{-1} &= \frac{\Lambda_y [N(T-1) - 2]}{\varphi(\beta, b)} \\ \text{and } B &= \frac{\Gamma\left(\frac{N(T-1)}{2}\right)}{\Gamma\left(\frac{N(T-1)}{2} - 1\right)} \sqrt{\frac{\Lambda_y}{\varphi(\beta, b)}} = \frac{N(T-1)}{2\pi [N(T-1) - 2]} \sqrt{\frac{\Lambda_y}{\varphi(\beta, b)}} \end{aligned} \quad \text{(A.8)}$$

Then, the expression

$$B\left[1 + \frac{\Sigma^{-1}}{N(T-1) - 2} \{\rho - \widehat{\rho}(\beta, b)\}^2\right]^{-\frac{N(T-1)}{2}}$$

is the pdf of a t-distribution $t_\nu(\widehat{\rho}(\beta, b), \Sigma)$ with $\Sigma = \frac{\varphi(\beta,b)}{\Lambda_y[N(T-1)-2]}$ and $\nu = N(T-1) - 2$ degrees of freedom. As $\pi^*(\rho|\beta, b)$ is a linear combination of a t-distribution (*e.g.*, $C^{-1}t_\nu$ with

[2] A random variable $X \in \mathbb{R}^p$ has a multivariate Student distribution with location parameters μ, shape matrix Σ and ν degrees of freedom, $X \sim t_\nu(\mu, \Sigma)$ if its pdf is given by

$$\frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{p}{2}} \mid \Sigma \mid^{\frac{1}{2}}} \left[1 + \frac{1}{\nu}(x - \mu)' \Sigma^{-1} (x - \mu)\right]^{-\frac{\nu+p}{2}}$$

$C = A.B.\varphi(\beta, b)^{\frac{N(T-1)}{2}}$), we can use the results of Welch (1947) and Fairweather (1972) (see also Witkovskỳ (2004))[3].
As the mean (resp. the variance) of a $t_{\nu_S}(\widehat{\rho}(\beta, b), \Sigma)$ distribution is $\widehat{\rho}(\beta, b)$, for $\nu_S > 1$ (resp. $\frac{\nu_S}{\nu_S - 2}\Sigma$, for $\nu_S > 2$), then the mean of the posterior density of ρ is:

$$\widehat{\rho} = D\widehat{\rho}(\beta, b)$$

and the variance of the posterior density of ρ is

$$Var[\widehat{\rho}] = D^2.\frac{\nu_S}{\nu_S - 2}\Sigma$$

with

$$\begin{aligned} \nu_S &= 4 + C^2 (N(T-1) - 6) \\ \text{and } D &= \sqrt{\frac{\nu_S - 2}{\nu_S} \frac{N(T-1) - 2}{N(T-1) - 4}} \end{aligned}$$

If ρ is assumed to be $U(-1, 1)$, then we get a three-step approach. For the dynamic specification: $y = \rho y_{-1} + X\beta + Wb + u$, we can integrate first with respect to (β, τ) given b and ρ, and then, conditional on β and ρ, we can next integrate with respect to (b, τ) and last, we can integrate with respect to ρ given (β, b).

1. Let $y^* = (y - \rho y_{-1} - Wb)$. Derive the conditional ML-II posterior distribution of β given the specific effects b and ρ as in the section 2.3.1 of the main text.
2. Let $\widetilde{y} = (y - \rho y_{-1} - X\beta)$. Derive the conditional ML-II posterior distribution of b given the coefficients β and ρ as in the section 2.3.2 of the main text.
3. Let $y^\circ = (y - X\beta - Wb)$. Derive the conditional ML-II posterior distribution of ρ given the coefficients β and b as in the previous section A.1.

As the mean and variance of ρ are exactly defined, we do not need to introduce an ε-contamination class of prior distributions for ρ at the second stage of the hierarchy. This was initially our first goal. Unfortunately, the results obtained on a Monte Carlo simulation study (see section A.2) provide

[3] If $S = \sum_{j=1}^{k} \lambda_j t_{\nu_j}$ is a weighted linear combination of independent Student's t random variables with ν_j $(j = 1, ..., k)$ degrees of freedom, then S is approximated by the distribution of the D multiple of single Student's t random variables with ν_S degrees of freedom, say $D.t_{\nu_S}$ where ν_S and $D > 0$ are to be determinated by equating the second and the fourth moments of $D.t_{\nu_S}$ to those of S. In particular, if $\nu_j > 4$, $\forall j$, then:

$$\nu_S = 4 + \frac{1}{\sum_{j=1}^{k} \frac{\lambda_j^2}{\nu_j - 4}}, \; D = \sqrt{\frac{\nu_S - 2}{\nu_S} \frac{1}{\sum_{j=1}^{k} \frac{\nu_j - 2}{\nu_j}}}$$

In our case, $k = 1$, $\lambda_1 = (\frac{1}{C})$ with $C = A.B.\varphi(\beta, b)^{\frac{N(T-1)}{2}}$ and $\nu_1 = N(T-1) - 2$. So

$$\begin{aligned} \nu_S &= 4 + C^2 (\nu_1 - 4) = 4 + C^2 (N(T-1) - 6) \\ \text{and } D &= \sqrt{\frac{\nu_S - 2}{\nu_S} \frac{\nu_1}{\nu_1 - 2}} = \sqrt{\frac{\nu_S - 2}{\nu_S} \frac{N(T-1) - 2}{N(T-1) - 4}} \end{aligned}$$

biased estimates of ρ, β and residual variances. That is why we assume a Zellner g-prior, for the $\theta\big(= [\rho, \beta']'\big)$ vector encompassing the coefficient of the lagged dependent variable $y_{i,t-1}$ and those of the explanatory variables X'_{it}. Thus, we do not impose stationarity constraints like many authors and we respect the philosophy of ε-contamination class using data-driven priors.

A.2. Some Monte Carlo results

We run a Monte Carlo simulation study for the dynamic random effects world comparing different robust Bayesian estimators. As previously in the main text, we run the two stage approach with individual block resampling bootstrap assuming a Zellner g-prior, for the $\theta\big(= [\rho, \beta']'\big)$ vector encompassing the coefficient of the lagged dependent variable $y_{i,t-1}$ and those of the explanatory variables X'_{it}. We introduce a two stage three step approach when $\rho \sim U(-1, 1)$. When the initial value of ρ is drawn for a uniform distribution $U(-1, 1)$, results are strongly biased as shown on Table A.1. Even if we initialize ρ with its OLS estimator on the pooled model, the results, if they improve, are further biased.

Table A.1: Dynamic Random Effects World

$N = 100$, $T = 10$, $\varepsilon = 0.5$, Replications=1,000

		ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1	1	4			
2S boot	coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	0.4943	521.89
	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			
2S-3S boot	coef	0.2358	1.2131	1.1870	1.2510	5.4548	167.5906	0.0527	0.3821	697.44
$U(-1,1)$	se	0.3397	0.1923	0.1970	0.2063					
	rmse	0.5202	0.2412	0.2159	0.2749					
2S-3S boot	coef	0.7379	1.1174	1.0764	1.1148	1.0400	4.6650	0.0465	0.4920	700.85
$U(-1,1)$ init OLS	se	0.0065	0.0275	0.0273	0.0282					
	rmse	0.0131	0.1200	0.0801	0.1175					

2S bootstrap: two stage with individual block resampling bootstrap.

2S-3S boot $U(0,1)$: two stage - three steps with individual block resampling bootstrap and with $\rho \sim U(-1,1)$ where ρ init is drawn from $U(-1,1)$.

2S-3S boot $U(0,1)$ init OLS: two stage - three steps with individual block resampling bootstrap and with $\rho \sim U(-1,1)$ where ρ init is the OLS estimator of ρ.

B. Some extra Monte Carlo simulation results

1. Dynamic random effects world with $\rho = 0.98$.
2. Chamberlain-type Fixed Effects World with $\rho = 0.75$.
3. Sensitivity to ε-contamination values
4. Departure from normality

Table B.2: Dynamic Random Effects World

$N = 100$, $T = 10$, $\varepsilon = 0.5$, Replications=1,000

		ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.98	1	1	1	1	4			
2S boot	coef	0.9822	1.0151	0.9843	1.0087	0.9971	3.7994	$< 10^{-4}$	0.4929	1219.06
	se	0.0009	0.0242	0.0247	0.0250	0.0485	0.2003			
	rmse	0.0022	0.0339	0.0334	0.0306	0.0485	0.2833			
	hpdi_lower	0.9801	0.9482	0.9185	0.9365	0.9016	3.4361			
	hpdi_upper	0.9843	1.0832	1.0538	1.0766	1.0908	4.1940			
GMM	coef	0.9831	0.9300	0.8502	0.9279	0.9016	4.0194			155.51
	se	0.0009	0.0282	0.0273	0.0290	0.0561	0.4290			
	rmse	0.0036	0.0857	0.1568	0.0879	0.1132	0.4292			
QMLE	coef	0.9794	1.0110	0.9771	1.0083	0.9942	4.1953			626.90
	se	0.0007	0.0256	0.0252	0.0255	0.0508	0.2606			
	rmse	0.0009	0.0279	0.0279	0.0279	0.0511	0.3255			

2S boot: two stage with individual block resampling bootstrap.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

hpdi_lower (hpdi_upper): lower and upper bounds of the 95% Highest Posterior Density Interval (HPDI).

Table B.3: Dynamic Chamberlain-type Fixed Effects World

$N = 100$, $T = 10$, $\varepsilon = 0.5$, Replications=$1,000$

	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
true	0.75	1	1	1	1	162.5341			
2S boot coef	0.7414	1.0134	1.0132	0.9959	0.9952	173.8813	$< 10^{-4}$	0.4959	686.26
se	0.0061	0.0234	0.0236	0.0231	0.0665	36.1240			
rmse	0.0105	0.0277	0.0274	0.0241	0.0666	37.8470			
	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.6400	0.8000	1.0000
2S boot coef	0.1865	0.2267	0.2812	0.3465	0.4284	0.5327	0.6622	0.8224	1.0259
se	0.1068	0.1063	0.1064	0.1061	0.1075	0.1081	0.1072	0.1087	0.1089
rmse	0.0797	0.0825	0.0836	0.0796	0.0851	0.0822	0.0875	0.0842	0.0878

	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
QMLE coef	0.7507	0.9979	0.9982	0.9988	1.0057	162.1588			1199.07
se	0.0136	0.0426	0.0416	0.0247	0.1384	25.1817			
rmse	0.01368	0.0426	0.0426	0.0426	0.1385	25.1721			
	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.6400	0.8000	1.0000
QMLE coef	0.1689	0.2083	0.2616	0.3274	0.4083	0.5090	0.6388	0.7958	0.9961
se	0.0772	0.0777	0.0782	0.0759	0.0810	0.0807	0.0858	0.0888	0.0971
rmse	0.0772	0.0776	0.0782	0.0759	0.0810	0.0807	0.0857	0.0888	0.0971

2S boot: two stage with individual block resampling bootstrap.

QMLE: quasi-maximum likelihood estimation.

Table B.4: Dynamic Chamberlain-type Fixed Effects World

$N = 200$, $T = 30$, $\varepsilon = 0.5$, Replications=500$^{(*)}$

	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)	
true	0.75	1	1	1	1	204.3142				
2S boot coef	0.7472	1.0060	1.0059	0.9992	1.0005	208.6873	$< 10^{-4}$	0.4984	4525.51	
se	0.0023	0.0086	0.0087	0.0091	0.0267	28.4166				
rmse	0.0036	0.0110	0.0107	0.0094	0.0266	28.7231				
		π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true		0.0019	0.0024	0.0030	0.0038	0.0047	0.0059	0.0074	0.0092	0.0115
2S boot coef		0.0031	0.0019	0.0038	0.0037	0.0035	0.0083	0.0039	0.0100	0.0120
se		0.0775	0.0765	0.0769	0.0770	0.0771	0.0773	0.0773	0.0768	0.0763
rmse		0.0581	0.0557	0.0570	0.0603	0.0566	0.0594	0.0577	0.0605	0.0564
	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	π_{16}	π_{17}	π_{18}	π_{19}	π_{20}
true	0.0144	0.0180	0.0225	0.0281	0.0352	0.0440	0.0550	0.0687	0.0859	0.1074
2S boot coef	0.0228	0.0215	0.0210	0.0262	0.0354	0.0426	0.0599	0.0731	0.0876	0.1082
se	0.0768	0.0771	0.0774	0.0761	0.0761	0.0774	0.0759	0.0775	0.0766	0.0764
rmse	0.0585	0.0567	0.0587	0.0570	0.0528	0.0597	0.0575	0.0581	0.0593	0.0573
	π_{21}	π_{22}	π_{33}	π_{24}	π_{25}	π_{26}	π_{27}	π_{28}	π_{29}	π_{30}
true	0.1342	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.6400	0.8000	1.0000
2S boot coef	0.1388	0.1682	0.2100	0.2643	0.3295	0.4167	0.5188	0.6464	0.8070	1.0095
se	0.0764	0.0757	0.0759	0.0776	0.0761	0.0767	0.0770	0.0760	0.0765	0.0778
rmse	0.0560	0.0582	0.0602	0.0606	0.0598	0.0571	0.0606	0.0599	0.0564	0.0610

$^{(*)}$ *When* $T = 30$, *we restrict the exercise to only* 500 *replications, not because of our estimator under R but because of the size of the simulated database and its reading under Stata. Indeed, for* $1,000$ *replications and* $T = 30$, *one must read, under Stata,* $(1 + 4 + 30) \times 1,000 = 35,000$ *variables of size* $(NT, 1)$*! Even with a* 64*-bit computer and Stata (MP and S), only* $32,767$ *variables can be read. On the other hand, with our code under R, there is no limitation of that order.*

Table B.4 **- Cont'd:** Dynamic Chamberlain-type Fixed Effects World

$N = 200$, $T = 30$, $\varepsilon = 0.5$, Replications=500

	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)	
true	0.75	1	1	1	1	204.3142				
QMLE coef	0.7499	0.9999	0.9998	0.9995	0.9995	204.4599			4059.55	
se	0.0022	0.0090	0.0086	0.0095	0.0191	20.8648				
rmse	0.0022	0.0090	0.0090	0.0090	0.0191	20.8444				
		π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
true		0.0019	0.0024	0.0030	0.0038	0.0047	0.0059	0.0074	0.0092	0.0115
QMLE coef		0.0028	0.0013	0.0016	0.0032	0.0018	0.0065	0.0032	0.0088	0.0120
se		0.0558	0.0528	0.0535	0.0557	0.0542	0.0553	0.0553	0.0567	0.0529
rmse		0.0558	0.0528	0.0535	0.0556	0.0543	0.0552	0.0554	0.0567	0.0529
	π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	π_{16}	π_{17}	π_{18}	π_{19}	π_{20}
true	0.0144	0.0180	0.0225	0.0281	0.0352	0.0440	0.0550	0.0687	0.0859	0.1074
QMLE coef	0.0205	0.0204	0.0211	0.0261	0.0362	0.0426	0.0587	0.0709	0.0848	0.1068
se	0.0546	0.0517	0.0552	0.0543	0.0503	0.0560	0.0544	0.0543	0.0559	0.0532
rmse	0.0549	0.0517	0.0551	0.0543	0.0502	0.0560	0.0545	0.0543	0.0559	0.0531
	π_{21}	π_{22}	π_{33}	π_{24}	π_{25}	π_{26}	π_{27}	π_{28}	π_{29}	π_{30}
true	0.1342	0.1678	0.2097	0.2621	0.3277	0.4096	0.5120	0.6400	0.8000	1.0000
QMLE coef	0.1366	0.1653	0.2082	0.2598	0.3272	0.4123	0.5128	0.6406	0.7998	0.9994
se	0.0517	0.0563	0.0573	0.0560	0.0562	0.0544	0.0562	0.0554	0.0546	0.0565
rmse	0.0517	0.0563	0.0573	0.0560	0.0561	0.0544	0.0562	0.0553	0.0545	0.0564

Table B.5: Dynamic Random Effects World, robustness to ε-contamination

$N = 100$, $T = 10$, replications=$1,000$

		ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.75	1	1	1	1	4			
	GMM coef	0.7554	0.9912	0.9135	0.9886	0.8837	4.6186			149.15
	se	0.0067	0.0273	0.0266	0.0279	0.0541	0.3857			
	rmse	0.0115	0.0392	0.0941	0.0403	0.1282	0.7289			
$\varepsilon = 0$	2S boot coef	0.7383	1.0168	0.9871	1.0116	0.9906	4.4573	1.0000	1.0000	546.59
	se	0.0059	0.0248	0.0252	0.0250	0.0479	0.2310			
	rmse	0.0127	0.0296	0.0274	0.0275	0.0488	0.5123			
$\varepsilon = 10^{-17}$	2S boot coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	1.0000	532.84
	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			
$\varepsilon = 0.1$	2S boot coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	0.8979	542.62
	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			
$\varepsilon = 0.5$	2S boot coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	0.4943	521.89
	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			
$\varepsilon = 0.9$	2S boot coef	0.7522	1.0252	0.9952	1.0207	0.9976	4.0433	$< 10^{-4}$	0.0980	543.019
	se	0.0060	0.0249	0.0253	0.0252	0.0481	0.2025			
	rmse	0.0055	0.0351	0.0248	0.0324	0.0482	0.2070			

2S boot: two stage with individual block resampling bootstrap.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table B.6: Dynamic heterogeneous panel data model with common correlated effects, robustness to ε-contamination

$N = 100$, $T = 30$, replications=$1,000$

		$\overline{\rho}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
	true	0.7501	0.7498	1	1			
	CCEMG coef	0.7205	0.7826	1.0856	1.0409			1618.05
	se	0.0221	0.0554	0.0453	0.0353			
	rmse	0.0369	0.0632	0.0968	0.0540			
$\varepsilon = 0$	2S boot coef	0.7338	0.8047	1.0746	1.1454	1.0000	1.0000	4694.98
	se	0.0192	0.0659	0.0508	0.2734			
	rmse	0.0252	0.0858	0.0903	0.3095			
$\varepsilon = 10^{-17}$	2S boot coef	0.7345	0.8071	1.0753	1.0536	0.1997	0.9809	4664.89
	se	0.0195	0.0671	0.0531	0.0392			
	rmse	0.0250	0.0882	0.0921	0.0664			
$\varepsilon = 0.1$	2S boot coef	0.7347	0.8072	1.0759	1.0536	0.0004	0.7847	4550.70
	se	0.0194	0.0670	0.0531	0.0392			
	rmse	0.0248	0.0882	0.0926	0.0664			
$\varepsilon = 0.5$	2S boot coef	0.7347	0.8071	1.0759	1.0536	0.0011	0.3925	4695.86
	se	0.0194	0.0671	0.0531	0.0392			
	rmse	0.0248	0.0882	0.0926	0.0664			
$\varepsilon = 0.9$	2S boot coef	0.7347	0.8071	1.0759	1.0536	0.0010	0.0738	4566.65
	se	0.0194	0.0671	0.0531	0.0392			
	rmse	0.0248	0.0882	0.0926	0.0664			

2S boot: two stage with individual block resampling bootstrap.

CCEMG: Common Correlated Effects Mean Group estimator.

mean coefficients: $\overline{\rho} = (1/N)\sum_{i=1}^N \rho_i$ and $\overline{\beta}_1 = (1/N)\sum_{i=1}^N \beta_{1i}$.

Table B.7: Dynamic Random Effects World, robustness to ε-contamination, departure from normality: the skewed t-distribution

$N = 100$, $T = 10$, $\varepsilon = 0.5$, replications=1,000

	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_θ	λ_μ	Computation Time (secs.)
true	0.75	1	1	1	7.0052	4			
GMM coef	0.7588	0.9871	0.8901	0.9824	6.0976	5.9883			149.15
se	0.0104	0.0449	0.0414	0.0456	4.5903	1.1453			
rmse	0.0196	0.0683	0.1262	0.0698	4.6769	2.2943			
2S boot coef	0.7141	1.0711	1.0382	1.0714	6.7171	5.8258	0.0004	0.4984	599.38
se	0.0169	0.0625	0.0625	0.0651	4.0849	1.8050			
rmse	0.0402	0.0956	0.0756	0.0985	4.0931	2.5668			
QMLE coef	0.7471	1.0215	0.9414	1.0172	6.8577	4.0629	460.36		
se	0.0158	0.0637	0.0625	0.0631	4.8017	0.7293			
rmse	0.0160	0.0672	0.0673	0.0673	4.8016	0.7317			

2S boot: two stage with individual block resampling bootstrap.
GMM: Arellano-Bond GMM estimation.
QMLE: quasi-maximum likelihood estimation.

Table B.8: Dynamic heterogeneous panel data model with common correlated effects, departure from normality: the skewed t-distribution

$N = 100$, $T = 30$, $\varepsilon = 0.5$, replications=1,000

	$\overline{\rho}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	Computation Time (secs.)
true	0.7501	0.7498	1	6.9824			
CCEMG coef	0.6300	0.8064	1.2112	7.3744			1624.07
se	0.0319	0.1364	0.1119	1.6597			
rmse	0.1242	0.1477	0.2390	6.5867			
2S boot coef	0.6930	0.8788	1.2188	7.0220	0.0110	0.4202	1544.04
se	0.0222	0.1292	0.1018	2.5438			
rmse	0.0613	0.1825	0.2413	2.5428			

2S boot: two stage with individual block resampling bootstrap.
CCEMG: Common Correlated Effects Mean Group estimator.
mean coefficients: $\overline{\rho} = (1/N)\sum_{i=1}^N \rho_i$ and $\overline{\beta}_1 = (1/N)\sum_{i=1}^N \beta_{1i}$.

References

References

Baltagi, B.H., Bresson, G., Chaturvedi, A., Lacroix, G., 2018. Robust linear static panel data models using ε-contamination. Journal of Econometrics 202, 108–123.

Fairweather, W.R., 1972. A method of obtaining an exact confidence interval for the common mean of several normal populations. Journal of the Royal Statistical Society: Series C (Applied Statistics) 21, 229–233.

Shrivastava, A., Chaturvedi, A., Bhatti, M.I., 2019. Robust Bayesian analysis of a multivariate dynamic model. Physica A: Statistical Mechanics and its Applications, *(forthcoming)* .

Singh, A., Chaturvedi, A., 2012. Robust Bayesian analysis of autoregressive fixed effects panel data model. Working Paper. Department of Statistics, University of Allahabad (India).

Welch, B.L., 1947. The generalization of Student's problem when several different population variances are involved. Biometrika 34, 28–35.

Witkovskỳ, V., 2004. Matlab algorithm TDIST: The distribution of a linear combination of Student's random variables, in: Antoch, J. (Ed.), COMPSTAT 2004, Proceedings in Computational Statistics, Physica-Verlag Springer. pp. 1971–1978.