Chaire de recherche en **évaluation économique** des **programmes publics**

Robust dynamic panel data models using epsilon-contamination

Badi H **Baltagi**
Georges **Bresson**
Anoop **Chaturvedi**
Guy **Lacroix**

wp eepp 2021-008

www.eepp.ulaval.ca

Robust dynamic panel data models using epsilon-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 epsilon-contamination class of prior distributions for the model parameters as well as for the individual effects. Second, both the base elicited priors and the epsilon-contamination priors use Zellner's (1986) *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.

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 and the Blundell-Bond system GMM (see Arellano and Bond (1991), Blundell and Bond (1998) and the survey by Bun and Sarafidis (2015)). Despite its optimal asymptotic properties, 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 (Dhaene and Jochmans, 2011, 2016) or estimation methods based on the likelihood function of the first differences (Hsiao et al., 2002; 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), Everaert and Pozzi (2007) 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 the aforementioned 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 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), Hsiao and Pesaran (2008), 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 may be invalid, weak or both (see Bun and Sarafidis (2015)). Based on the same identifying assumption, some alternatives have been proposed in the literature (e.g. Ahn and Schmidt (1995) and Hsiao et al. (2002), to mention a few). Maximum likelihood estimators, asymptotically equivalent to the GMM 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

[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 (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) and Wooldridge (2005)).

Hsiao et al. (1999), Hsiao and Pesaran (2008), Koop et al. (2008), Juárez and Steel (2010), Liu et al. (2017), 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 (see Bauwens et al. (2005)). Geweke and Keane (2000) 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.

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. For instance, Su and Yang (2015) 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), Berger (1985) and Berger and Berliner (1984, 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 $\sigma_u^2 \, (= 1/\tau)$.[3]

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

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

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 $(NT \times K_2)$ with $K_2 = Nk_2$, u is $(NT \times 1)$ and I_{NT} is a $(NT \times NT)$ identity matrix. TO better understand the difference between $X_{it}'\beta$ and $W_{it}'b_i$ (or $X\beta$ and Wb for the NT observations), we provide the following examples of Wb which we will use in the dynamic version in Section 3 of the Monte Carlo simulation study.[4] In the random effects world, $Wb = Z_\mu \mu$ with $Z_\mu = I_N \otimes \iota_T$ being of dimension $(NT \times N)$, $\otimes$ is the Kronecker product, ι_T is a $(T \times 1)$ vector of ones and μ is a $(N \times 1)$ vector of idiosyncratic parameters. When $W \equiv Z_\mu$, the random effects are $\mu \sim N\left(0, \sigma_\mu^2 I_N\right)$. For a Chamberlain-type fixed effects world, the individual effects are given by $\mu = \underline{X}\Pi + \varpi$, where $\underline{X}$ is a $(N \times TK_x)$ matrix with $\underline{X}_i = (X_{i1}', ..., X_{iT}')$ and $\Pi = (\pi_1', ..., \pi_T')'$ is a $(TK_x \times 1)$ vector. The model can be rewritten as: $y = X\beta + Z_\mu \underline{X}\Pi + Z_\mu \varpi + u = Z\theta + Z_\mu \varpi + u$ with $Z = [X, Z_\mu \underline{X}]$ and let $Wb \equiv Z_\mu \varpi$ with $\varpi \sim N\left(0, \sigma_\varpi^2 I_N\right)$. For the Hausman-Taylor world, $y = X\beta + V\eta + Z_\mu \mu + u$, where V is a vector of time-invariant variables, and subsets of X and V may be correlated with the individual effects μ, but leaves the correlations unspecified. Then, $y = Z\theta + Wb + u$ with $Z = [X, V]$ and $Wb = Z_\mu \mu$ with $\mu \sim N\left(0, \sigma_\mu^2 I_N\right)$. For the panel data world with common correlated effects (or common trends), $y = X\beta + Wb + u = X\beta + F\Gamma + u$ where the $(NT \times Nm)$ matrix F of the m unobserved factors (or common trends) is a blockdiagonal matrix where each $(T \times m)$ sub block f is replicated N times and Γ is the $(Nm \times 1)$ individual varying coefficients vector.

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)

[3]Here we assume the homoskedasticity of u_{it}. As the GMM or QML estimators are robust to the presence of time-series heteroskedasticity, one could introduce time-series heteroskedasticity as an ARCH processes or HAR-type models. We thank an anonymous referee for this suggestion. However, in that case, the derivation of the marginal likelihoods (or predictive densities) (see section 2.3) is quite involved. An extension along these lines is beyond the scope of the current paper and is more appropriately dealt with in a separate one.

[4]We thank an anonymous referee for this suggestion.

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)).

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):

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

where $\pi_0(\cdot)$ is the base elicited prior, $q(\cdot)$ is the contamination belonging to some suitable class Q of prior distributions, and $0 \leq \varepsilon \leq 1$ 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 $\pi_0(\beta, b, \tau | g_0, h_0)$ is the base prior assumed to be a specific g-prior with

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

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 $\pi_0(\beta, b, \tau | g_0, h_0)$. The probability density function (henceforth pdf) of the base prior $\pi_0(.)$ is given by:

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

The possible class of contamination Q is defined as:

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

with

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

where β_q, b_q, g_q and h_q are unknown. The restrictions $g_q \leq g_0$ and $h_q \leq h_0$ 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).[5] 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:

$$y_{it} = \rho y_{it-1} + X'_{it}\beta + W'_{it}b_i + u_{it} = Z'_{it}\theta + W'_{it}b_i + u_{it} \ , \ i = 1, ..., N \ , \ t = 2, ..., T, \tag{8}$$

where $Z'_{it} = [y_{it-1}, X'_{it}]$ and $\theta' = [\rho, \beta']$ are $(1 \times K_1)$ vectors 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 (2), we have the following:

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

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 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, 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 could be assumed

[5] 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 $\pi_0(\beta, b, \tau | g_0, h_0)$ 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.

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. Unfortunately, the results using Monte Carlo simulations showed biased estimates of ρ, β and residual variances (see Appendices A.1 and 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.

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

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:

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

and

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

with

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

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[6]:

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

with $\widetilde{H} = \frac{\Gamma\left(\frac{N(T-1)}{2}\right)}{\pi^{\left(\frac{N(T-1)}{2}\right)} v(b)^{\left(\frac{N(T-1)}{2}\right)}}$, $R^2_{\theta_0} = \frac{(\widehat{\theta}(b)-\theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b)-\theta_0\iota_{K_1})}{(\widehat{\theta}(b)-\theta_0\iota_{K_1})'\Lambda_Z(\widehat{\theta}(b)-\theta_0\iota_{K_1})+v(b)}$, $\widehat{\theta}\left(b\right) = \Lambda_Z^{-1}Z'y^*$ and $v\left(b\right) = (y^* - Z\widehat{\theta}\left(b\right))'(y^* - Z\widehat{\theta}\left(b\right))$, and where $\Gamma\left(\cdot\right)$ is the Gamma function.

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

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

where

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

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

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

and

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

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

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

and

$$\widehat{g}_q = \min\left(g_0, g^*\right) \text{ with } g^* = \max\left[\left(\frac{(N(T-1)-K_1)}{K_1}\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right)-1\right)^{-1}, 0\right]. \tag{14}$$

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

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

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

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

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

Note that $\widehat{\lambda}_{\theta,g_0}$ depends upon the ratio of the $R^2_{\theta_0}$ and $R^2_{\theta_q}$, but primarily on the sample size $N(T-1)$. Indeed, $\widehat{\lambda}_{\theta,g_0}$ tends to 0 when $R^2_{\theta_0} > R^2_{\theta_q}$ and tends to 1 when $R^2_{\theta_0} < R^2_{\theta_q}$, irrespective of the model fit (*i.e*, the absolute values of $R^2_{\theta_0}$ or $R^2_{\theta_q}$). Only the relative values of $R^2_{\theta_q}$ 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 $\left(\frac{\xi_{0,\theta}M^{-1}_{0,\theta}}{N(T-1)-2}\right)$ and degrees of freedom $(N(T-1))$ with

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

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

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

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

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

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

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

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

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

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 $\widehat{\theta}_{EB}\left(b|g_0\right)$. 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 $\pi_0^*\left(\theta|g_0\right)$ 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 $\widehat{q}^*\left(\theta|g_0\right)$ and to the empirical Bayes estimator $\widehat{\theta}_{EB}\left(b|g_0\right)$. 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.[7]

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:

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

with

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

and where $\widehat{\lambda}_{b,h_0}$ is defined in a similar manner as $\widehat{\lambda}_{\theta,g_0}$.[8] $\pi_0^*\left(b|h_0\right)$ is the pdf of a multivariate t-distribution with mean vector $b_*(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{0,b}M_{0,b}^{-1}}{N(T-1)-2}\right)$ and degrees of freedom $(N(T-1))$ with

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

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

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

[7]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)).

[8]$\widehat{\lambda}_{b,h_0}$ is defined by the following elements $\widehat{h}$, R_{b_0}, $R_{\widehat{b}_q}$, $\widehat{b}\left(\theta\right)$, $\widehat{b}_q$ which are analogous respectively to $\widehat{g}_q$, R_{θ_0}, $R_{\widehat{\theta}_q}$, $\widehat{\theta}\left(b\right)$, $\widehat{\theta}_q$ and where W, Λ_W, $\widetilde{y}$, h_0, K_2, b_0, ι_{K_2} replace Z, Λ_Z, y^*, g_0, K_1, θ_0, ι_{K_1} of the preceeding subsection.

$q^*(b|h_0)$ is the pdf of a multivariate t-distribution with mean vector $\widehat{b}_{EB}(\theta|h_0)$, variance-covariance matrix $\left(\frac{\xi_{1,b}M_{1,b}^{-1}}{N(T-1)-2}\right)$ and degrees of freedom $(N(T-1))$ 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[9] 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[10]. Here, we will use the same individual block resampling bootstrap method. Following Kapetanios (2008), individual block resampling consists of drawing an $(N \times (T-1))$ matrix Y^{BR} whose rows are obtained by resampling those of an $(N \times (T-1))$ 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
2. In the first step, compute the mean of the ML-II posterior density of θ using our initial shrinkage procedure

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

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

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

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:

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

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 are inspired by Chudik and Pesaran (2015a,b).

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

$$\begin{aligned} x_{1,1,it} &= 0.7x_{1,1,it-1} + \delta_i + \zeta_{it} \\ x_{1,2,it} &= 0.7x_{1,2,it-1} + \theta_i + \varsigma_{it} \\ u_{it} &\sim N\left(0, \tau^{-1}\right), \ (\delta_i, \theta_i, \zeta_{it}, \varsigma_{it}) \sim U(-6, 6) \\ \text{and } \rho &= 0.75, \ \beta_{1,1} = \beta_{1,2} = \beta_2 = 1. \end{aligned}$$

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

$$\begin{aligned} \eta_1 &= \eta_2 = 0 \\ x_{2,it} &= 0.7x_{2,it-1} + \kappa_i + \vartheta_{it} \ , \ (\kappa_i, \vartheta_{it}) \sim U(-6, 6) \\ \mu_i &\sim N\left(0, \sigma_\mu^2\right) \ , \ \sigma_\mu^2 = 4\tau^{-1}. \end{aligned}$$

   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:

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

$x_{1,1,it}$ 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:

$$\begin{aligned}
\eta_1 &= \eta_2 = 1; \\
x_{2,it} &= 0.7x_{2,it-1} + \mu_i + \vartheta_{it} \ , \ \vartheta_{it} \sim U(-6,6); \\
V_{1,i} &= 1, \forall i; \\
V_{2,i} &= \mu_i + \delta_i + \theta_i + \xi_i, \ \xi_i \sim U(-6,6); \\
\mu_i &\sim N\left(0, \sigma_\mu^2\right) \text{ and } \sigma_\mu^2 = 4\tau^{-1}.
\end{aligned}$$

$x_{1,1,it}$ 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

$$\begin{aligned}
&\zeta_{x_{it}} \sim U(-3,3) \ , && \eta_{i,\gamma_l} \sim N(0, \sigma_{\gamma_l}^2) \ , && \eta_{i,\gamma_{x_l}} \sim N(0, \sigma_{\gamma_{x_l}}^2) \\
&\sigma_{\gamma_l}^2 = \sigma_{\gamma_{x_l}}^2 = 0.2^2 \ , && \gamma_l = \sqrt{l \times c_\gamma} \ , && \gamma_{x_l} = \sqrt{l \times c_{x,l}} \\
&c_\gamma = (1/m) - \sigma_{\gamma_l}^2 \ , && c_{x,l} = \tfrac{2}{m(m+1)} - \tfrac{2\sigma_{\gamma_{x_l}}^2}{(m+1)} \ , && \text{and } u_{it} \sim N\left(0, \tau^{-1}\right).
\end{aligned}$$

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 common correlated effects, we follow Chudik and Pesaran (2015a,b), and assume that the m common trends f_t in (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 common correlated 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}, $\omega_{2,it}$, ... 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.[11]

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.[12] In the normal regression case, and following Kass and Wasserman (1995), the UIP corresponds to $g_0 = h_0 = 1/N(T-1)$, 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 $\text{Var}(\nu) = \sigma_\mu^2 (I_N \otimes J_{T-1}) + \sigma_u^2 I_{N(T-1)}$, where $J_{T-1} = \iota_{T-1} \iota'_{T-1}$. This model is usually estimated using GMM. It could also be estimated using the quasi-maximum likelihood (QML) estimator (see 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.[13]

Table 1 reports the results of fitting the Bayesian two-stage model with block resampling bootstrap (*2S bootstrap*)[14] along with those from the GMM and QMLE, each in a separate panel for

[11] $\varepsilon = 0.5$ 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).

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

[13] 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 for the QMLE. 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.

[14] 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.

$N = 100, T = 10$ (see Table B.1 in the supplementary material for $N = 200, T = 30$).[15] The true parameter values appear in the first row of the Table. The last column reports the computation time in seconds.[16] 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[17] and RMSEs that are very close. For the coefficient of the lagged dependent variable, ρ, as well as for the β coefficients, the RMSE are the lowest both for the *2S bootstrap* and the QMLE when $N = 100$ and $T = 10$ or when $N = 200$ and $T = 30$. GMM yields higher RMSEs for all coefficients. The *2S bootstrap* and the QMLE have also better RMSEs than the GMM estimator 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 $b_i(\equiv \mu_i)$. Note that the computation times are roughly the same for the *2S bootstrap* and QMLE compared to those of GMMs which explode when N and T increase.

We can also use a simple and efficient way to drastically reduce the computation time of our Bayesian two-stage estimator. The ML-II posterior density of θ in (15) is a two-component finite mixture of multivariate t-distributions whose location parameters and scale matrices are given in (16) to (19). Following McLachlan and Lee (2013), one can generate mixture of multivariate skewed (or non-skewed) t-distributions via an EM Algorithm approach. Thus, generating 1000 (or more) random samples of K_1-dimensional multivariate t observations with location parameters, scale matrices given in (16)-(19) and degrees of freedom $N(T-1)$, allows one to sample a mixture of the two components to get 1000 (or more) random vectors of θ_{ML-II}. The latter can then be used to compute the variances of the K_1 parameters.[18] Using this device reduces the computation time by as much as by $80-90\%$ in all cases, although small discrepancies with the bootstrapped standard errors may occur in specific cases. For instance, for $N = 100$, $T = 10$ (resp. $N = 200$, $T = 30$) and $r = 0.8$, the computation time shrinks from 285 seconds to 19 seconds (resp. 3079 seconds to 289 seconds) — a reduction by a factor of 15 (resp. 10.6) — and the estimated standard errors of the parameters ρ and β by the bootstrap and mixture approaches differ by at most 0.007 (resp. 0.003). (see Tables C.1 to C.3 in the supplementary material). This considerable time-saving device makes our Bayesian two-stage estimator all the more favorable.

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, mainly for the estimation of the residual disturbances (σ_u^2) and the random effects (σ_μ^2). An important point to

[15] Whenever we discuss samples other than ($N = 100$, $T = 10$) for the random effects, the Chamberlain-type fixed effects and the Hausman-Taylor worlds or ($N = 100$, $T = 30$) for the commom correlated effects world, we refer the reader to Tables in section B in the supplementary material.

[16] 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.

[17] 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.

[18] See the supplementary material for more details.

note is the significant bias and unlikely estimated values of σ_u^2 and σ_μ^2 for QMLE.[19] *2S bootstrap* has the lowest RMSE for all the parameters and the variances of specific effects and remainder disturbances. 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.

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 1 once again shows that the results of the *2S bootstrap* are very close to those of the QML estimator. Table B.3 in the supplementary material reports the estimates of the π_t coefficients. *2S bootstrap* has RMSE very close to those of QMLE whatever the sample size except for σ_μ^2 when $N = 100$ and $T = 10$ but this difference diminishes quickly when $N = 200$ and $T = 30$ (see Table B.4 in the supplementary material). Hsiao and Zhou (2018) show that a properly specified QMLE that uses the Chamberlain approach to condition the unobserved effects and initial values on the observed strictly exogenous covariates is asymptotically unbiased if N goes to infinity whether T is fixed or goes to infinity.[20] Note that the computation times of the QMLE are 1.6 to 3.7 times longer than those of the *2S bootstrap*.[21] Here again, the RMSE of *2S bootstrap* and QMLE are close to each other.

[19] When ρ is close to the non-stationarity bounds (*i.e.*, close to one), the QMLE, two-stage QMLE, CCEP and CCEMG estimators give unrealistic values of the variances σ_u^2 and σ_μ^2 or cannot be computed.

[20] If they underline that it does not matter whether the initial values are treated as fixed constants or random variables when N is fixed and T is large, they point out that the treatment of initial values becomes important when N is large. When one treats initial values as fixed constants, the QMLE is inconsistent if T is fixed and $N \to \infty$. It becomes consistent but asymptotically biased of order $\sqrt{N/T^3}$ when both N and T are large. However, if the specific effects are correlated with the exogenous variables, the naive GLS estimator on fixed initial conditions is asymptotically biased of order $\sqrt{N/T}$ as $(N, T) \to \infty$. Hsiao and Zhou (2018) show that using the Chamberlain's approach to condition the effects on observed explanatory variables can reduce the order of asymptotic bias to $\sqrt{N/T^3}$. On the other hand, they show that the QMLE with a properly modeled initial value distribution combined with Chamberlain's approach is consistent and asymptotically unbiased whether T is fixed or goes to infinity as long as $N \to \infty$. For our specification $y_{it} = \rho y_{i,t-1} + z_{it}'\theta + \varpi_i + u_{it}$, the used QMLE estimator with the `xtdpdqml` Stata command iterates the initial observations continuously backward in time as: $y_{i0} = \rho^m y_{i,-m} + \sum_{s=0}^{m-1} \rho^s z_{i,-s}'\theta + \frac{1-\rho^m}{1-\rho}\varpi_i + + \sum_{s=0}^{m-1} \rho^s u_{i,-s}$ with $m \to \infty$. The last right three terms of the previous equation imply a restriction on the covariance between the initial observations and the unit-specific effects: $\phi\sigma_0^2 = \sigma_\varpi^2/(1-\rho)$ with $\sigma_0^2 = (\sigma_\varpi^2 + \sigma_u^2)/(1-\rho)^2$. Feasible starting values for the variance parameters $(\sigma_\varpi^2, \sigma_u^2, \sigma_0^2, \phi)$ need to satisfy the restriction $(\sigma_\varpi^2 - \phi^2\sigma_0^2)T > -\sigma_u^2$ (see Kripfganz (2016)). We use the following starting values $(\sigma_\varpi^2, \sigma_u^2, \sigma_0^2, \phi) = (0.1, 0.2, 0.2, 0.2)$ which allow to get unbiased QMLE.

[21] Again, the mixture approach for calculating the standard errors considerably reduces the computation time. For $N = 200$ and $T = 30$, it decreases from 2708 seconds to 254 seconds — a reduction by a factor of 10.7 — and the differences with the bootstrap approach are marginal (see Table C.4 in the supplementary material).

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 $X'_{2,it}$ and $E_{\overline{X}'_2}$ is the general mean of X'_2. We could estimate the conditional distribution of $\mu_i \mid \overline{X}'_{2,i}, V'_{2i}$ if we knew the elements of the variance-covariance matrix Σ_{js}. We can nevertheless assume that

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

where ϖ_i 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

$$\mu_i = (\overline{x}_{2,i} - E_{\overline{x}_2})\theta_X + f\left[(\overline{x}_{2,i} - E_{\overline{x}_2}) \odot (V_{2i} - E_{V_2})\right]\theta_V + \varpi_i. \tag{25}$$

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.[22] They have proposed a new `xtseqreg` Stata command which implements the standard error correction for two-stage dynamic linear panel data models.[23]

Table 1 compares results of the *2S bootstrap* estimator to those of the two-stage QML sequential approach. Once again, the estimates and the RMSE are very close to one another (see Table B.5 in the supplementary material for other (T, N) simulations). 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 3.9% for $T = 10$ to 1.3% 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 in a static model (see Baltagi et al. (2018)). Note that the *2S bootstrap* estimator yields very small biases of σ_μ^2 but these biases decrease rapidly as the individual span is increased from -1.5% for $N = 100$ to -0.75% for $N = 200$. Finally, note that the computation times of the two-stage QML sequential approach are 1.1 to 2.6 times longer than those of the *2S bootstrap*.[24]

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:

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

[22] 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.

[23] 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_μ^2 but only σ_u^2.

[24] As with previous estimators, the mixture approach for calculating the standard errors considerably reduces the computation time. For $N = 200$ and $T = 10$, it shrinks from 1183 seconds to as litle as 106 seconds — a reduction by a factor of 11 — although the standard errors differ marginally from those of the bootstrap approach (see Table C.5 in the supplementary material).

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:

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

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.[25] 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 2 shows that the results of the *2S bootstrap* estimator are close to those of the CCEP estimator. Moreover, the RMSEs of *2S bootstrap* and CCEP are close each other. Even if one increases the size of N and/or T, all estimators yield essentially the same parameter estimates and associated RMSEs (See Table B.6 in the supplementary material). The computation time is slightly longer with our estimator given the bootstrap procedure and the use of the inverse $(F'F)^{-1}$ where F could be a huge block-diagonal matrix. In fact, the *2S bootstrap* computation times are 1.2 to 1.6 times longer than those of CCEP.[26]

3.1.5. The dynamic homogeneous panel data world with common correlated 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 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 (24)), we can approximate the $((T-1) \times m)$ f

[25] 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.

[26] Once again, the use of the mixture approach for the calculation of the standard errors considerably reduces the computation time. For $N = 100$, $T = 30$ (resp. $N = 50$, $T = 50$), it goes from 1775 to 106 seconds (resp. from 1901 seconds to 113 seconds) — a sizable factor (16.6 and 16.8, respectively) — and differences with the bootstrap approach are minimal (see Table C.6 in the supplementary material).

matrix with a $((T-1)\times K_1)$ f^* matrix of the within time transformation[27] 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: $f_t^* = \left[\overline{y}_t, \overline{y}_{-1,t}, \overline{x}_t, \overline{x}_{-1,t}\right]$.[28] 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 2 shows that the results of the *2S bootstrap* are very close to those of the dynamic CCEP estimator. However, the RMSE is slightly smaller for CCEP than *2S bootstrap*.[29] Once again, the computation time is longer with our estimator as the *2S bootstrap* computation times are 1.4 to 1.6 times longer than those of CCEP.[30]

3.1.6. The dynamic heterogeneous panel data world with common correlated 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.[31]

While the bottom panel of Table 2 gives insights on the distribution of ρ_i and β_{1i} for different sample sizes, the panel of Table 2 located just above 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 2 shows that the results of the *2S bootstrap* estimator are close to those of the CCEMG estimator but the bias of σ_u^2 is slightly larger for CCEMG. The RMSEs results are mixed. *2S bootstrap* gives smaller RMSE for ρ and σ_u^2 than CCEMG, but CCEMG gives a smaller RMSE for $\overline{\beta}_1$. The results

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

[28] 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 $p_T = 0$.

[29] See Table B.6 in the supplementary material for $N = 50$, $T = 50$.

[30] The mixture approach for the calculation of the standard errors once again reduces considerably the computation time. For $N = 100$, $T = 30$ (resp. $N = 50$, $T = 50$), it goes from 1996 seconds to 344 seconds (resp. from 1820 seconds to 177 seconds) — a reduction factor of between 5.8 and 10.3 — and differences with the bootstrap approach are minimal (see Table C.7 in the supplementary material).

[31] 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.

for β_2 depend on the sample size. For $\overline{\beta}_1$ the bias is 2.5% for $N = 100$, $T = 30$, (resp. 5.7% for $N = 50$, $T = 50$)[32] for the *2S bootstrap* estimator as compared to those of the CCEMG estimator (-1%, resp.(-0.8%)). For the residuals' variance σ_u^2, the bias increases with the time dimension for both estimators but to a lesser extent for *2S bootstrap*. However, all the estimators yield roughly the same parameter estimates. Once again, the computation time is longer with our estimator as the *2S bootstrap* computation times are 1.3 to 1.6 times longer than those of CCEMG.[33]

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 the conventional (classical) estimators. Although not reported for the sake of brevity, a thorough analysis of the sensitivity of our estimators to the contamination part of the prior distribution shows that parameter estimates are relatively stable.[34] Likewise, the robustness of our estimators when the remainder disturbances are assumed to follow a right-skewed t-distribution is investigated and shown to behave well in terms of precision and bias relative to classical estimators.[35]

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 or to settle for those 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. This is because our estimator is easily amenable to

[32] See Table B.6 in the supplementary material for $N = 50$, $T = 50$.

[33] The mixture approach once again is highly efficient. For $N = 100$, $T = 30$ (resp. $N = 50$, $T = 50$), computation time decreases from 3059 to 159 seconds (resp. from 1566 seconds to 177 seconds) — a reduction factor of between 8.8 and 19.2 — although small differences between the estimated standard errors remain (see Tables C.8 and C.9 in the supplementary material).

[34] See the supplementary material.

[35] See the supplementary material.

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, space-time), *etc.* We leave these for future work.

Table 1: Dynamic Random Effects, Chamberlain-type Fixed Effects and Hausman-Taylor Models

$\varepsilon = 0.5$, $r = 0.8$, N=100, T=10, Replications=$1,000$

	Dynamic Random Effects World							Chamberlain-type Fixed Effects World						
	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	CPU (sec.)	ρ	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	CPU (sec.)
true	0.75	1	1	1	1	4		0.75	1	1	1	1	162.534	
2S boot coef	0.749	1.005	0.996	1.006	0.994	4.084	285.88	0.748	1.001	1.001	0.998	0.999	162.571	349.40
se	0.002	0.008	0.008	0.024	0.048	0.158		0.002	0.008	0.008	0.023	0.066	31.096	
rmse	0.002	0.010	0.009	0.024	0.048	0.179		0.002	0.008	0.008	0.023	0.066	31.081	
GMM coef	0.748	1.005	0.979	1.015	0.874	4.844	134.53							
se	0.002	0.009	0.009	0.026	0.052	0.215								
rmse	0.003	0.013	0.024	0.031	0.135	0.871								
QMLE coef	0.748	1.004	0.992	1.006	0.993	4.078	471.14	0.750	0.999	0.999	1.004	1.002	162.258	1286.36
se	0.002	0.008	0.008	0.023	0.050	0.172		0.004	0.018	0.018	0.196	0.065	23.352	
rmse	0.002	0.009	0.009	0.009	0.051	0.188		0.004	0.018	0.018	0.018	0.065	23.342	
(†) $\lambda_\theta < 10^{-4}$, $\lambda_\mu = 0.494$								(†) $\lambda_\theta < 10^{-4}$, $\lambda_\mu = 0.496$						

2S boot coef: two-stage with individual block resampling bootstrap. (†): the parameters λ_θ and λ_μ only concern the 2s boot estimator. QMLE: quasi-maximum likelihood estimation. For the Chamberlain-type fixed effects world, the parameters π_t are omitted from the Table, see Table B.3 in the supplementary material.

Dynamic Hausman-Taylor World

	ρ	β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	CPU
True	0.75	1	1	1	1	1	1	4	
2s boot coef	0.748	1.000	1.000	1.001	1.028	1.039	0.994	3.946	311.80
se	0.003	0.011	0.011	0.012	0.230	0.047	0.063	0.681	
rmse	0.003	0.011	0.011	0.012	0.232	0.061	0.064	0.682	
two-stage QML coef	0.749	1.000	0.999	1.000	1.021	1.000	0.993	$n.a$	696.25
se	0.002	0.009	0.009	0.009	0.201	0.070	0.048	$n.a$	
rmse	0.002	0.009	0.009	0.009	0.202	0.070	0.048	$n.a$	
(†) $\lambda_\theta < 10^{-4}$, $\lambda_\mu = 0.499$									

2S boot: two-stage with individual block resampling bootstrap. (†): the parameters λ_θ and λ_μ only concern the 2s boot estimator.
two-stage QMLE: two-stage quasi-maximum likelihood sequential approach with non available (n.a) estimate of σ_μ^2.

Table 2: Dynamic Homogeneous/Heterogeneous Panel Data Models With Common Correlated Effects
$\varepsilon = 0.5$, Replications=$1,000$, N=100, T=30

Dynamic Homogeneous Model with Common Trends							
	ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.75	1	1	1			
2S boot coef	0.751	0.992	0.996	1.004	$< 10^{-4}$	0.459	1775.99
se	0.001	0.012	0.015	0.026			
rmse	0.002	0.014	0.015	0.026			
CCEP coef	0.748	1.001	1.006	0.991			1437.60
se	0.001	0.010	0.012	0.027			
rmse	0.002	0.010	0.013	0.028			
Dynamic Homogeneous Model with Common Correlated Effects							
	ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.75	1	1	1			
2S boot coef	0.745	1.019	1.013	1.028	$< 10^{-4}$	0.481	1996.72
se	0.003	0.018	0.019	0.032			
rmse	0.005	0.027	0.023	0.043			
CCEP coef	0.748	1.000	1.004	1.023			1392.09
se	0.002	0.011	0.012	0.031			
rmse	0.002	0.011	0.013	0.038			
Dynamic Heterogeneous Model with Common Correlated Effects							
	$\overline{\rho}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.750	0.750	1	1			
2S boot coef	0.752	0.769	1.011	1.097	0.002	0.416	1996.33
se	0.013	0.029	0.018	0.061			
rmse	0.013	0.035	0.021	0.115			
CCEMG coef	0.742	0.754	1.013	1.181			1569.35
se	0.014	0.026	0.016	0.134			
rmse	0.016	0.026	0.021	0.226			

2S boot: two-stage with individual block resampling bootstrap.
CCEP: Common Correlated Effects Pooled estimator.
CCEMG: Common Correlated Effects Mean Group estimator. Mean coefficients:

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

.

	$N = 100, T = 30$	
	ρ_i	β_{1i}
min	0.603	0.505
mean	0.750	0.750
sd	0.087	0.144
max	0.897	0.995

References

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

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

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.

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.

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., 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.

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

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.

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.

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

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.

Hsiao, C., Zhou, Q., 2018. Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. Journal of Econometrics 207, 114–128.

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.

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.

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.

McLachlan, G., Lee, S., 2013. EMMIXuskew: An R package for fitting mixtures of multivariate skew t distributions via the EM algorithm. Journal of Statistical Software 55, 1–22.

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., 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.

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.

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.

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*

September 25, 2021

*Corresponding author.
Email addresses: bbaltagi@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 ρ 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 $\pi^*\left(\rho|\beta, b\right)$ can be written as:

$$\begin{aligned}\int\limits_0^{\infty} p(\rho)p\left(\tau\right)p\left(y^{\circ}|y_{-1}, \rho, \tau\right)d\tau &= \int\limits_0^{\infty}\left(\frac{1}{2}\right)\left(\frac{1}{\tau}\right)\left(\frac{\tau}{2\pi}\right)^{\frac{N(T-1)}{2}} \\ &\quad \times \exp\left(-\frac{\tau}{2}\left\{\varphi\left(\beta, b\right) + \Lambda_y \left\{\rho - \widehat{\rho}(\beta, b)\right\}^2\right\}\right)d\tau\end{aligned}$$

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

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

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

$$
\begin{aligned}
A &= [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[\int_{-\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))}^{0} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt + \int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt \right] \\
&= [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[\int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt + \int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt \right] \\
&= [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}} \sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}} \left[I_1 + I_2 \right] \qquad \text{(A.3)}
\end{aligned}
$$

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]

$$
I_1 = \int_{0}^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1+\widehat{\rho}(\beta,b))} \left(1+t^2\right)^{-\frac{N(T-1)}{2}} dt = \frac{1}{2}\int_{0}^{\zeta_1} \eta^{-\frac{1}{2}} (1-\eta)^{\frac{N(T-1)-3}{2}} d\eta \qquad \text{(A.4)}
$$

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

$$
\begin{aligned}
I_1 &= \frac{1}{2}\zeta_1^{\frac{1}{2}} \int_{0}^{1} z^{-\frac{1}{2}} (1-\zeta_1 z)^{\frac{N(T-1)-3}{2}} dz = \frac{1}{2}\zeta_1^{\frac{1}{2}} \frac{\Gamma(1/2)}{\Gamma(3/2)} \times_2 F_1\left(-\frac{N(T-1)-3}{2}; \frac{1}{2}; \frac{3}{2}; \zeta_1\right) \\
&= \zeta_1^{\frac{1}{2}} \times_2 F_1\left(-\frac{N(T-1)-3}{2}; \frac{1}{2}; \frac{3}{2}; \zeta_1\right) \qquad \text{(A.5)}
\end{aligned}
$$

Using the Pfaff's transformation:

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

[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.

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

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

Hence

$$ \begin{aligned} I_1 &= \zeta_1^{\frac{1}{2}}(1-\zeta_1)^{-\frac{1}{2}}\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};\frac{\zeta_1}{\zeta_1-1}\right) \\ &= \sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}\left(1+\widehat{\rho}(\beta,b)\right)\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2\right) \end{aligned} \tag{A.6} $$

Similarly we obtain

$$ I_2 = \int_0^{\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}(1-\widehat{\rho}(\beta,b))}\left(1+t^2\right)^{-\frac{N(T-1)}{2}}dt=\frac{1}{2}\int_0^{\zeta_2}\eta^{-\frac{1}{2}}(1-\eta)^{\frac{N(T-1)-3}{2}}d\eta \tag{A.7} $$

$$ \text{where } \zeta_2=\frac{\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2}{\left[1+\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2\right]} $$

$$ I_2 = \zeta_2^{\frac{1}{2}}\times_2 F_1\left(-\frac{N(T-1)-3}{2};\frac{1}{2};\frac{3}{2};\zeta_2\right)=\zeta_2^{\frac{1}{2}}(1-\zeta_2)^{-\frac{1}{2}}\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};\frac{\zeta_2}{\zeta_2-1}\right) $$

Since

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

Hence

$$ I_2 = \sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}\left(1-\widehat{\rho}(\beta,b)\right)\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2\right) \tag{A.8} $$

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

$$ \begin{aligned} A &= [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}}\sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}}\left[I_1+I_2\right] \\ &= [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}}\sqrt{\frac{\varphi(\beta,b)}{\Lambda_y}}\left[\begin{array}{c}\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}\left(1+\widehat{\rho}(\beta,b)\right)\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2\right) \\ +\sqrt{\frac{\Lambda_y}{\varphi(\beta,b)}}\left(1-\widehat{\rho}(\beta,b)\right)\times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2\right)\end{array}\right] \end{aligned} $$

$$A = [\varphi(\beta,b)]^{-\frac{N(T-1)}{2}} \left[\begin{array}{c} (1+\widehat{\rho}(\beta,b)) \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1+\widehat{\rho}(\beta,b)\right)^2\right) \\ +(1-\widehat{\rho}(\beta,b)) \times_2 F_1\left(\frac{N(T-1)}{2};\frac{1}{2};\frac{3}{2};-\frac{\Lambda_y}{\varphi(\beta,b)}\left(1-\widehat{\rho}(\beta,b)\right)^2\right) \end{array} \right] \quad \text{(A.9)}$$

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

$$\begin{aligned} \pi^*(\rho|\beta,b) &= \frac{1}{A}\left[\varphi(\beta,b)+\Lambda_y\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}} \\ &= \frac{\varphi(\beta,b)^{-\frac{N(T-1)}{2}}}{(AB)} B\left[1+\frac{\Sigma^{-1}}{N(T-1)-1}\{\rho-\widehat{\rho}(\beta,b)\}^2\right]^{-\frac{N(T-1)}{2}} \end{aligned} \quad \text{(A.10)}$$

where[2]

$$\Sigma^{-1} = \frac{\Lambda_y[N(T-1)-1]}{\varphi(\beta,b)} \text{ and } B = \frac{\Gamma\left(\frac{N(T-1)}{2}\right)}{\Gamma\left(\frac{N(T-1)}{2}-\frac{1}{2}\right)}\sqrt{\frac{\Lambda_y}{\pi\varphi(\beta,b)}} \quad \text{(A.11)}$$

Since $N(T-1)$ is large, we can use the following result:

$$\lim_{x\to\infty} x^{b-a}\frac{\Gamma(x+a)}{\Gamma(x+b)} = 1 \to \lim_{x\to\infty}\frac{\Gamma(x+a)}{\Gamma(x+b)} = x^{a-b}$$

with $x = N(T-1)/2$, $a = 0$ and $b = -1/2$ so that B becomes:[3]

$$B = \sqrt{\frac{N(T-1)}{2\pi}\frac{\Lambda_y}{\varphi(\beta,b)}} \quad \text{(A.12)}$$

Then, the conditional posterior density of ρ, given (β, b), $\pi^*(\rho|\beta,b)$ is the pdf of a t-distribution $C^{-1}t_\nu(\widehat{\rho}(\beta,b),\Sigma)$ with shape matrix $\Sigma = \frac{\varphi(\beta,b)}{\Lambda_y[N(T-1)-1]}$, $\nu = N(T-1)-1$ degrees of freedom and

[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

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

[3]The asymptotic expansion of the ratio of two Gamma functions is given by:

$$\frac{\Gamma(x+a)}{\Gamma(x+b)} \sim x^{a-b}\sum_{k=0}^{\infty}\frac{G_k(a,b)}{x^k}$$

where $G_k(a,b) = \binom{a-b}{k}B_k^{a-b+1}(a)$ and $B_j^i(n)$ is a generalized Bernoulli polynomial and $x = N(T-1)/2$, $a = 0$, $b = -1/2$. Using the computational knowledge engine Wolfram alpha (`https://www.wolframalpha.com`) with the command `Series[ Gamma[n/2]/Gamma[(n - 1)/2], {n, [Infinity], 3}]`, we get:

$$\frac{\Gamma\left(\frac{N(T-1)}{2}\right)}{\Gamma\left(\frac{N(T-1)-1}{2}\right)} \sim \sqrt{\frac{N(T-1)}{2}} - \frac{3}{4\sqrt{2N(T-1)}} - \frac{7}{32\sqrt{2(N(T-1))^3}} + \mathcal{O}\left(\frac{1}{(N(T-1))^2}\right) = \sqrt{\frac{N(T-1)}{2}} \text{ as } N(T-1)\to\infty.$$

$C = A.B.\varphi(\beta, b)^{\frac{N(T-1)}{2}} = \sqrt{\frac{N(T-1)}{2\pi}}(I_1 + I_2)$. Using the results of Kotz and Nadarajah (2004)[4], the mean of the posterior density of ρ is:

$$\widehat{\rho} = C^{-1}\widehat{\rho}(\beta, b) \tag{A.13}$$

and the variance of the posterior density of ρ is

$$Var[\widehat{\rho}] = C^{-2}.\frac{\nu}{\nu - 2}\Sigma \tag{A.14}$$

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 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 still slightly biased and the computation time is much longer than that of the two-stage two-step approach. The last panel of Table A.1 shows that less biased results are obtained with initial value of ρ coming from a LSDV estimator of ρ (rather than an OLS estimator of ρ).[5] So, there is a significant cost to pay when using a constrained ρ coefficient because it is not enough to

[4] *"If X has the p-variate t distribution with degrees of freedom ν, mean vector μ, and correlation matrix R, then, for any nonsingular scalar matrix C and for any a, $CX + a$ has the p-variate t distribution with degrees of freedom ν, mean vector $C\mu + a$, and correlation matrix CRC'."* (Kotz and Nadarajah (2004), p.15).

[5] OLS: $y = Z\theta + \nu$ with $Z = [y_{-1}, X]$, $\theta = (\rho, \beta')'$, $\nu = Z_\mu \mu + u$ and LSDV: $y = Z\theta + u$ with $Z = [y_{-1}, X, Z_\mu]$, $\theta = (\rho, \beta', \mu')'$.

generate it initially from a $U(-1,1)$ but it must first be estimated from an LSDV estimator which also greatly increases the computation time and that the bias on σ_μ^2 increases. All this confirms our idea that we are doing well not to impose stationarity constraints like many authors and respect in this sense the philosophy of ε-contamination class using data-driven priors.

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	0.7492	1.0057	0.9964	1.0062	0.9946	4.0849	$< 10^{-4}$	0.4940	285.88
	se	0.0025	0.0084	0.0083	0.0243	0.0481	0.1584			
	rmse	0.0023	0.0101	0.0090	0.0247	0.0484	0.1797			
2S-3S boot	coef	0.2511	1.1934	1.1705	1.2323	5.2118	160.5033	0.0520	0.3836	616.37
$U(-1,1)$	se	0.3843	0.2437	0.2418	0.2661	1.4235	50.1455			
	rmse	0.5060	0.2258	0.2033	0.2612	4.4456	164.3330			
2S-3S boot	coef	0.8020	1.0119	0.9812	1.0033	1.0751	4.4848	0.0560	0.4824	620.38
$U(-1,1)$ init OLS	se	0.0080	0.0294	0.0291	0.0308	0.0504	0.3067			
	rmse	0.0524	0.0301	0.0330	0.0284	0.0904	0.5735			
2S-3S boot	coef	0.7226	1.0384	1.0085	1.0362	0.9934	5.1754	0.0578	0.4926	937.83
$U(-1,1)$ init LSDV	se	0.0086	0.0265	0.0262	0.0263	0.0479	0.5296			
	rmse	0.0289	0.0457	0.0255	0.0441	0.0484	1.2891			

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

2S-3S boot $U(-1,1)$: 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 $U(-1,1)$ 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.

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

B. Some extra Monte Carlo simulation results

1. Dynamic random effects world with $N = 200, T = 30, \rho = 0.75$.
2. Dynamic random effects world with $N = 100, T = 10, \rho = 0.98$.
3. Chamberlain-type fixed effects world with $N = 100, T = 10, \rho = 0.75$ and the π_t parameters.
4. Chamberlain-type fixed effects world with $N = 200, T = 30, \rho = 0.75$ and the π_t parameters.
5. Hausman-Taylor world with $N = 100, T = 30, \rho = 0.75$ and $N = 200, T = 10, \rho = 0.75$.
6. Dynamic homogeneous/heterogeneous models with common correlated effets with $N = 50, T = 50, \rho = 0.75$ and $\overline{\rho} = 0.75$.

Table B.1: Dynamic Random Effects World

$\varepsilon = 0.5,\ r = 0.8, N = 200, T = 30$, 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	0.7498	1.0009	1.0003	1.0004	0.9991	4.1245	$< 10^{-4}$	0.4975	3079.78
se	0.0009	0.0030	0.0030	0.0030	0.0187	0.0548			
rmse	0.0007	0.0031	0.0030	0.0030	0.0187	0.1360			
GMM coef	0.7510	0.9942	0.9880	0.9894	0.9675	4.2862			5075.67
se	0.0007	0.0026	0.0026	0.0026	0.0229	0.0736			
rmse	0.0019	0.0080	0.0133	0.0119	0.0397	0.2955			
QMLE coef	0.7497	1.0002	0.9996	0.9997	0.9989	4.1227			2486.84
se	0.0007	0.0029	0.0030	0.0031	0.0188	0.0561			
rmse	0.0007	0.0029	0.0029	0.0029	0.0188	0.1349			

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

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

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	0.9801	1.0037	0.9946	1.0022	0.9944	4.0514	$< 10^{-4}$	0.4939	321.40
	se	0.0003	0.0081	0.0083	0.0083	0.0486	0.1502			
	rmse	0.0002	0.0090	0.0096	0.0086	0.0489	0.1587			
	hpdi_lower	0.9794	0.9805	0.9718	0.9793	0.9015	3.7619			
	hpdi_upper	0.9809	1.0252	1.0166	1.0251	1.0896	4.3364			
GMM	coef	0.9799	1.0053	0.9790	1.0051	0.8761	4.8407			132.04
	se	0.0003	0.0092	0.0090	0.0094	0.0523	0.2144			
	rmse	0.0004	0.0134	0.0245	0.0136	0.1345	0.8676			
QMLE	coef	0.9798	1.0034	0.9920	1.0025	12.1494	2.6910			1093.13
	se	0.0003	0.0127	0.0130	0.0119	352.7469	44.061			
	rmse	0.2298	0.0131	0.0131	0.0131	352.7467	44.058			

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	0.7489	1.0016	1.0014	0.9984	0.9997	162.5717	$< 10^{-4}$	0.4959	349.40
se	0.0022	0.0079	0.0080	0.0230	0.0667	31.0964			
rmse	0.0023	0.0083	0.0083	0.0236	0.0667	31.0808			
	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	0.1733	0.2125	0.2657	0.3331	0.4129	0.5129	0.6435	0.8012	1.0036
se	0.1044	0.1039	0.1040	0.1039	0.1060	0.1052	0.1047	0.1053	0.1046
rmse	0.0765	0.0784	0.0803	0.0762	0.0815	0.0769	0.0818	0.0780	0.0794

	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			
QMLE coef	0.7502	0.9990	0.9991	1.0044	1.0021	162.2586			1286.36
se	0.0048	0.0183	0.0184	0.1966	0.0651	23.3529			
rmse	0.0048	0.0183	0.0184	0.0183	0.0651	23.3429			
	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	0.1694	0.2084	0.2614	0.3274	0.4087	0.5092	0.6391	0.7968	0.9976
se	0.0780	0.0797	0.0802	0.0771	0.0803	0.0813	0.0824	0.0794	0.0844
rmse	0.0780	0.0797	0.0802	0.0771	0.0803	0.0813	0.0824	0.0795	0.0844

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	0.7496	1.0008	1.0007	0.9995	0.9971	203.2024	$< 10^{-4}$	0.4984	2708.19	
se	0.0009	0.0030	0.0030	0.0090	0.0265	29.3042				
rmse	0.0009	0.0032	0.0031	0.0094	0.0266	29.2960				
		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		0.0021	0.0020	0.0017	0.0025	0.0007	0.0058	0.0073	0.0086	0.0134
se		0.0763	0.0750	0.0754	0.0763	0.0765	0.0770	0.0768	0.0765	0.0756
rmse		0.0589	0.0553	0.0571	0.0589	0.0577	0.0583	0.0578	0.0583	0.0551
	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	0.0218	0.0203	0.0201	0.0254	0.0364	0.0412	0.0598	0.0704	0.0854	0.1078
se	0.0766	0.0760	0.0767	0.0743	0.0754	0.0765	0.0760	0.0760	0.0760	0.0773
rmse	0.0579	0.0554	0.0577	0.0544	0.0529	0.0583	0.0586	0.0584	0.0577	0.0565
	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	0.1376	0.1666	0.2082	0.2601	0.3286	0.4128	0.5147	0.6407	0.8010	1.0005
se	0.0760	0.0755	0.0762	0.0769	0.0761	0.0760	0.0755	0.0757	0.0757	0.0763
rmse	0.0551	0.0585	0.0600	0.0582	0.0594	0.0569	0.0591	0.0585	0.0561	0.0582

(*) *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	0.9999	0.9995	1.9166	204.3702			4335.38	
se	0.0008	0.0030	0.0029	0.0094	0.2765	20.5173				
rmse	0.0008	0.0030	0.0030	0.0030	0.9574	20.4968				
		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	0.0012	0.0016	0.0032	0.0018	0.0065	0.0031	0.0088	0.0119
se		0.0558	0.0528	0.0536	0.0557	0.0542	0.0553	0.0553	0.0567	0.0529
rmse		0.0558	0.0528	0.0535	0.0556	0.0542	0.0552	0.0554	0.0566	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	0.0361	0.0426	0.0587	0.0709	0.0848	0.1068
se	0.0546	0.0517	0.0553	0.0543	0.0502	0.0560	0.0544	0.0544	0.0559	0.0532
rmse	0.0549	0.0517	0.0552	0.0543	0.0502	0.0560	0.0545	0.0544	0.0558	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	0.1365	0.1652	0.2082	0.2598	0.3272	0.4122	0.5127	0.6404	0.7997	0.9993
se	0.0516	0.0561	0.0572	0.0560	0.0562	0.0544	0.0561	0.0551	0.0542	0.0558
rmse	0.0516	0.0561	0.0572	0.0560	0.0561	0.0544	0.0561	0.0551	0.0542	0.0558

Table B.5: 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	0.7496	1.0003	1.0006	1.0005	1.0196	1.0130	0.9984	3.9421	$< 10^{-4}$	0.5005	966.48
$T = 30$	se	0.0015	0.0058	0.0058	0.0060	0.2092	0.0251	0.0383	0.6015			
	rmse	0.0016	0.0058	0.0058	0.0060	0.2102	0.0283	0.0383	0.6039			
	two-stage QML coef	0.7500	0.9999	1.0001	1.0001	1.0191	0.9987	0.9978	$n.a$			2528.13
	se	0.0011	0.0043	0.0042	0.0043	0.1972	0.0463	0.0273	$n.a$			
	rmse	0.0011	0.0043	0.0043	0.0043	0.1981	0.0463	0.0274	$n.a$			
$N = 200$	2S boot coef	0.7484	1.0010	1.0009	1.0005	1.0057	1.0378	0.9972	3.9724	$< 10^{-4}$	0.4999	1183.16
$T = 10$	se	0.0022	0.0078	0.0077	0.0085	0.1611	0.0324	0.0466	0.4585			
	rmse	0.0027	0.0078	0.0078	0.0085	0.1612	0.0498	0.0467	0.4591			
	two-stage QML coef	0.7500	1.0001	1.0001	0.9998	0.9986	1.0013	0.9965	$n.a$			1242.89
	se	0.0020	0.0066	0.0063	0.0066	0.1427	0.0487	0.0355	$n.a$			
	rmse	0.0020	0.0066	0.0066	0.0066	0.1427	0.0487	0.0357	$n.a$			

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 B.6: Dynamic Homogeneous/Heterogeneous Panel Data Models with Common Correlated Effects
$\varepsilon = 0.5$, Replications=$1,000$, $N = 50$, $T = 50$

Dynamic Homogeneous Panel Data model with Common Trends							
	ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.75	1	1	1			
2S boot coef	0.7501	1.0032	1.0026	1.0116	$< 10^{-4}$	0.4613	1901.78
se	0.0026	0.0152	0.0188	0.0492			
rmse	0.0026	0.0156	0.0190	0.0505			
CCEP coef	0.7486	1.0007	1.0037	0.9807			1139.69
se	0.0020	0.0115	0.0141	0.0280			
rmse	0.0024	0.0116	0.0145	0.0340			
Dynamic Homogeneous Panel Data Model with Common Correlated Effects							
	ρ	β_1	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.75	1	1	1			
2S boot coef	0.7474	1.0146	1.0068	1.0235	$< 10^{-4}$	0.4865	1820.55
se	0.0031	0.0174	0.0196	0.0606			
rmse	0.0040	0.0227	0.0208	0.0650			
CCEP coef	0.7493	1.0011	1.0022	1.0148			1100.73
se	0.0019	0.0116	0.0142	0.0315			
rmse	0.0020	0.0117	0.0143	0.0348			
Dynamic Heterogenous Panel Data Model with Common Correlated Effects							
	$\overline{\rho}$	$\overline{\beta}_1$	β_2	σ_u^2	λ_θ	λ_μ	CPU (secs.)
true	0.7501	0.7503	1	1			
2S boot coef	0.7591	0.7936	1.0101	1.1808	$< 10^{-4}$	0.4407	1566.68
se	0.0217	0.0973	0.0643	0.0474			
rmse	0.0219	0.1572	0.0925	0.0963			
CCEMG coef	0.7568	0.7776	1.0070	1.2473			1205.11
se	0.0175	0.0383	0.0206	0.5980			
rmse	0.0188	0.0471	0.0218	0.6468			

2S boot: two-stage with individual block resampling bootstrap.
CCEP: Common Correlated Effects Pooled estimator.
CCEMG: Common Correlated Effects Mean Group estimator. Mean coefficients:

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

.

	$N = 50, T = 50$	
	ρ_i	β_{1i}
min	0.6059	0.5095
mean	0.7501	0.7503
sd	0.0863	0.1442
max	0.8942	0.9905

C. A simple and efficient way to drastically reduce the computation time of our Bayesian two-stage estimator.

The first stage of the Gaussian dynamic linear mixed model (eq.(10) in the main text) is given by

$$y = Z\theta + Wb + u \ , \ u \sim N(0, \tau^{-1} I_{NT}) \tag{C.15}$$

where y is $(N(T-1) \times 1)$. Z is $(N(T-1) \times (K_x + 1))$, W is $(N(T-1) \times K_2)$ with $K_2 = Nk_2$ and u is $(N(T-1) \times 1)$. The mean of the ML-II posterior density of θ is:

$$\begin{aligned} \widehat{\theta}_{ML-II} &= E\left[\widehat{\pi}^*\left(\theta|g_0\right)\right] = \widehat{\lambda}_{\theta,g_0} E\left[\pi_0^*\left(\theta|g_0\right)\right] + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) E\left[\widehat{q}^*\left(\theta|g_0\right)\right] \\ &= \widehat{\lambda}_{\theta,g_0} \theta_*(b|g_0) + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) \widehat{\theta}_{EB}\left(b|g_0\right). \end{aligned} \tag{C.16}$$

Baltagi et al. (2018) have shown that the ML-II posterior variance-covariance matrix of θ is given by

$$\begin{aligned} Var\left(\widehat{\theta}_{ML-II}\right) &= \widehat{\lambda}_{\theta,g_0} Var\left[\pi_0^*\left(\theta \mid g_0\right)\right] + \left(1 - \widehat{\lambda}_{\theta,g_0}\right) Var\left[\widehat{q}^*\left(\theta \mid g_0\right)\right] \\ &+ \widehat{\lambda}_{\theta,g_0}\left(1 - \widehat{\lambda}_{\theta,g_0}\right)\left(\theta_*(b \mid g_0) - \widehat{\theta}_{EB}\left(b \mid g_0\right)\right)\left(\theta_*(b \mid g_0) - \widehat{\theta}_{EB}\left(b \mid g_0\right)\right)' \end{aligned} \tag{C.17}$$

Many have raised concerns about the unbiasedness of the posterior variance-covariance matrix of $\widehat{\theta}_{ML-II}$. Indeed, it will be biased towards zero as $\widehat{\lambda}_{\theta,g_0} \to 0$ 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)). To solve this problem, Baltagi et al. (2018) have proposed two different strategies to approximate it, each with different desirable properties: MCMC with multivariate t-distributions or individual block resampling bootstrap. They have shown that fortunately, one needs as few as 20 bootstrap samples to achieve acceptable results. They also showed that the bootstrap method had some advantages over the MCMC method, especially in terms of computation time.

Computation times can be improved by using the Choleski decomposition for matrices inversion[6] for all the tested worlds (RE, Chamberlain, Hausman-Taylor, CCE). Additionally, multivariate normal random vectors in the common correlated effects (CCE) models could make use of the sparse matrices.[7] Yet, the efficiency gains would be relatively modest given the number of bootstrap draws that need to be generated.

An alternative approach arises if we exploit the intrinsic features of the distributions of the Bayes estimate $\theta_*(b|g_0)$ for the prior distribution $\pi_0\left(\theta, \tau\right)$ and the empirical Bayes estimate $\widehat{\theta}_{EB}\left(b|g_0\right)$ for the contaminated prior distribution $q\left(\theta, \tau\right)$.
We have shown that $\pi_0^*\left(\theta|g_0\right)$ is the pdf of a multivariate t-distribution $t_{K_1}\left(\theta_*(b|g_0), \Sigma_{\theta_*(b|g_0)}, N(T-1)\right)$

[6]The `R` function `chol2inv(chol(X))` instead of the standard function `solve(X)` for a symmetric definite positive X matrix.

[7]Using the `rmvn.sparse` command in the `sparseMVN` `R` package.

and $\widehat{q}^{*}(\theta)$ is the pdf of a multivariate t-distribution $t_{K_1}\left(\widehat{\theta}_{EB}(b|g_0), \Sigma_{\widehat{\theta}_{EB}(b|g_0)}, N(T-1)\right)$ where the mean vectors $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$ are given by

$$\theta_*(b|g_0) = \frac{\widehat{\theta}(b) + g_0\theta_0\iota_{K_1}}{g_0+1}\ ,\ \widehat{\theta}_{EB}(b|g_0) = \frac{\widehat{\theta}(b) + \widehat{g}_q\widehat{\theta}_q\iota_{K_1}}{\widehat{g}_q+1}.$$

and the variance-covariance matrices $\Sigma_{\theta_*(b|g_0)}$ and $\Sigma_{\widehat{\theta}_{EB}(b|g_0)}$ are given by

$$\begin{aligned}
\Sigma_{\theta_*(b|g_0)} &= \left(\frac{\xi_{0,\theta}M_{0,\theta}^{-1}}{N(T-1)-2}\right) \text{ with } M_{0,\theta} = \frac{(g_0+1)}{v(b)}\Lambda_Z \text{ and } \xi_{0,\theta} = 1 + \left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\theta_0}}{1-R^2_{\theta_0}}\right). \\
\Sigma_{\widehat{\theta}_{EB}(b|g_0)} &= \left(\frac{\xi_{q,\theta}M_{q,\theta}^{-1}}{N(T-1)-2}\right) \text{ with } M_{q,\theta} = \left(\frac{(\widehat{g}_q+1)}{v(b)}\right)\Lambda_Z \text{ and } \xi_{q,\theta} = 1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\theta}_q}}{1-R^2_{\widehat{\theta}_q}}\right).
\end{aligned}$$

Thus the ML-II posterior density of θ in (C.16) is a two-component finite mixture of multivariate t-distributions. Its pdf is given by

$$\pi\left(\tilde{\theta}_{ML-II}\right) = \sum_{h=1}^{2} \varrho_h \pi_h\left(\widehat{\theta}_{ML-II}, m_h, \Sigma_h, \nu_h\right). \tag{C.18}$$

where $\pi_h\left(\widehat{\theta}_{ML-II}, m_h, \Sigma_h, \nu_h\right)$ denotes the h-th pdf of the mixture model with location parameter m_h, scale matrix Σ_h and degrees of freedom ν_h. The mixing proportions satisfy $\varrho_h \geq 0$ $(h = 1, 2)$ and $\sum_{h=1}^{2} \varrho_h = 1$. In our case, $\nu_h = N(T-1)$, $\forall h$, $m_1 = \theta_*(b|g_0)$, $m_2 = \widehat{\theta}_{EB}(b|g_0)$, $\Sigma_1 = \Sigma_{\theta_*(b|g_0)}$, $\Sigma_2 = \Sigma_{\widehat{\theta}_{EB}(b|g_0)}$, $\varrho_1 = \widehat{\lambda}_{\theta,g_0}$ and $\varrho_2 = 1 - \widehat{\lambda}_{\theta,g_0}$.

Derivations of the location parameter and the scale matrix of a mixture of multivariate t-distributions is a very difficult task (see for instance Walker and Saw (1978), Peel and McLachlan (2000), Kotz and Nadarajah (2004), McLachlan and Peel (2004), Mengersen et al. (2011) among others). Parameter estimates of the mixture of t-distributions is generally obtained via an EM algorithm. McLachlan and Lee (2013) have proposed a `EMMIXuskew` `R` package for generating and fitting mixture of multivariate skewed (and non-skewed) t distributions via the EM Algorithm. Based on the command `rfmmst` of this package, and given the parameters of the two components defined above, one can generate 1000 (or more) random samples of K_1-dimensional multivariate t observations with location parameter m_h, scale matrix Σ_h and degrees of freedom ν_h for $h = 1, 2$, and hence sample from the mixture of these two components to generate as many random vectors of $\tilde{\theta}_{ML-II}$. The variances of the K_1 parameters can then be computed over these 1000 (or more) random samples.

After extensive experimentation, it was found that the estimated variances were slightly underestimated compared to those obtained with the bootstrap method. We therefore propose to correct the variances with the following multiplicative factor: $\sqrt{k_2}\left(1+\sqrt{\hat{r}}\right)^2$. In the RE, Chamberlain and Hausman-Taylor worlds, $\hat{r} = \hat{\sigma}_\mu^2/\left(\hat{\sigma}_\mu^2 + \hat{\sigma}_u^2\right)$ is the fraction of the variance $\left(\hat{\sigma}_\mu^2 + \hat{\sigma}_u^2\right)$ due to the

specific effects μ_i, with $\hat{\sigma}_u^2 = \hat{\tau}^{-1}$ and k_2 is the number of covariates in W in eq.(C.15).[8] In the common correlated effects worlds (CCE) worlds,[9] $\hat{\sigma}_\mu^2 = Var\left[\Gamma\right]$.

For the random effects world, and when the true $r = 0.8$ (see Table C.1)[10], the difference between the estimated standard errors of the parameters ρ and β from the bootstrap and mixture approaches is at most 0.007, and decreases when N and T increase. The more r decreases, the more the difference between the estimated standard errors (bootstrap *versus* mixture) decreases until reaching as little as 10^{-3} (see Tables C.2 and C.3). We also note that whatever the sample sizes N, T or the value of r, the bias of the GMM estimator of $\hat{\sigma}_\mu^2$, and to a lesser extent that of $\hat{\sigma}_u^2$, is bigger than those of our Bayesian two-stage estimator (*B2S*) and the QMLE estimator.

In the Chamberlain-type fixed effects worlds, the differences between the standard errors from the bootstrap and the mixture approaches are slightly larger than those of the random effects (see Table C.4). Note that the standard errors of β_{11} and β_{12} — from the mixture of t-distributions — are approximately 30% larger than those computed by the bootstrap approach. Yet the more important difference concerns the standard error of β_2, which is twice as large. However, as these standard errors are small, the 95% confidence intervals are nevertheless very similar (for instance, for β_2: $[0.95; 1.04]$ for `se_boot` and $[0.90; 1.08]$ for `se_mixt` for $N = 100$ and $T = 10$). Increasing N and T greatly reduces the differences between the two approaches. Thus, for $N = 200$ and $T = 30$, the 95% confidence intervals of β_2 are $[0.98; 1.01]$ for `se_boot` and $[0.96; 1.03]$ for `se_mixt` and the differences are marginal.

For the Hausman-Taylor world, the mixture approach leads to similar standard errors for ρ, β_{11} and β_{12} but generates a small bias for β_2, and yet more important downward biases for the standard errors of the parameters of the time-invariant variables η_1 (the constant) and to a lesser extent for η_2 (see Table C.5). As N increases and especially when T increases, these biases reduce but that of the constant η_1 always remains bigger.

For the common correlated effects world with common trends, we find few differences between standard errors arising from the bootstrap and the mixture approaches and these differences taper off as N increases (see Table C.6). We find the same type of results for the common correlated effects world with unobserved common factors (see Table C.7).

In the case of the dynamic heterogeneous panel data world with common correlated unobserved effects, the correction factor needs to be modified slightly to take into account the average over all

[8] In other words, $k_2 = 1$ for the RE, Chamberlain and Hausman-Taylor worlds. $k_2 = m$ for the common trends world and the common correlated effects world.

[9] Γ is given in sections 3.1.4 and 3.1.6 in the main text.

[10] Table C.1 reports the results of fitting the Bayesian two-stage (*B2S*) model with block resampling bootstrap (*se_boot*) and mixture of t-distributions (*se_mixt*) for estimating the standard errors 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 on the first line of the Table. The last column reports the computation time in seconds.

individuals. The corrected variance of $\tilde{\theta}_{ML-II}$ is computed as:[11]

$$\begin{aligned} Var\left[\tilde{\theta}_{ML-II}\right] &= \frac{1}{N}\sum_{i=1}^{N} Var\left[\tilde{\theta}_{i,ML-II}\right]_{cor} \\ & \text{with } Var\left[\tilde{\theta}_{i,ML-II}\right]_{cor} = Var\left[\tilde{\theta}_{i,ML-II}\right]\sqrt{mN}\left(1+\sqrt{\hat{r}_i}\right)^2 \\ & \text{and } \hat{r}_i = Var\left[\hat{\gamma}_i\right]/\left(Var\left[\hat{\gamma}_i\right]+Var\left[\hat{u}_i\right]\right). \end{aligned}$$

This additional correction factor, $\sqrt{N}$, is somewhat reminescent of the results from Theorem 3 of Chudik and Pesaran (2015), which shows that the convergence rate of the CCEMG estimator $\hat{\theta}_{CCEMG}$ of θ is $\sqrt{N}$ due to the heterogeneity of the coefficients. Moreover, Chudik and Pesaran (2015) show that the ratio $N/T \to \kappa_1$, for some constant $0 < \kappa_1 < \infty$, is required for the derivation of the asymptotic distribution of $\hat{\theta}_{CCEMG}$ due to the time series bias and that it is unsuitable for panels with T being small relative to N.[12] Standard errors computed using the mixture approach are close to those from the bootstrap one except for the standard error of the coefficient ρ of the lagged dependent variable which appears to be slightly downward biased (see Table C.8). When the sample size increases ($N = 200$, $T = 30$ and $T = 50$) (see Table C.9), the previous results are confirmed but the gain in computation time for the mixture approach is reduced from a factor of 10 to a factor of 6, which is still considerable in a dynamic heterogeneous panel data world with common correlated effects.

As confirmed in Tables C.1 to C.9, the standard errors calculated with the bootstrap method and those calculated with the mixture of t distributions diverge slightly in some cases. Assuredly, the most impressive result concerns the gain in computation speed irrespective of the tested worlds. Yet, the small differences in the standard errors is a small price to pay to obtain ε-contamination estimators in a timely fashion. The computation time of the two-stage method using the mixture model to estimate the parameter variances is 10 times faster than the one using the bootstrap method. This also applies to approaches (GMM, QMLE, 2SQMLE, CCEP or CCEMG). With a larger sample size ($N = 200$ and $T = 50$), a final confirmation is provided in Table C.10 for the random effects world.[13] The differences between the estimates for the Bayesian two-stage (*B2S*) estimator with `se_boot`, or with `se_mixt`, and those obtained with QMLE are marginal when the sample size is significantly increased. The difference in computation times is impressive (464 seconds for `se_mixt`, 3911 seconds for `se_boot` and 6150 seconds for QMLE for $1,000$ replications). The advantage of our Bayesian two-stage estimator is pretty obvious. And because it has little computing time, it should be valuable for applied econometricians.

[11] γ_i is given in sections 3.1.4 and 3.1.6 in the main text.

[12] In their simulation study, Chudik and Pesaran (2015) use $0.2 \leq N/T \leq 5$. They also use a jackknife bias correction and a recursive mean adjustment correction of the CCEMG estimator.

[13] We do not report the GMM estimation since it takes more than $50,000$ seconds!

Table C.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			
$N = 100$	B2S coef	0.7492	1.0057	0.9964	1.0062	0.9946	4.0849	$< 10^{-4}$	0.4940	
$T = 10$	se_boot	0.0025	0.0084	0.0083	0.0243	0.0481	0.1584			285.88
	se_mixt	0.0024	0.0097	0.0093	0.0168	0.0481	0.1569			19.96
	rmse	0.0023	0.0101	0.0090	0.0247	0.0484	0.1797			
	GMM coef	0.7489	1.0052	0.9793	1.0153	0.8746	4.8449			134.53
	se	0.0027	0.0093	0.0092	0.0263	0.0523	0.2152			
	rmse	0.0038	0.0137	0.0243	0.0315	0.1359	0.8718			
	QMLE coef	0.7485	1.0041	0.9927	1.0061	0.9931	4.0782			471.14
	se	0.0023	0.0088	0.0086	0.0236	0.0507	0.1720			
	rmse	0.0027	0.0097	0.0097	0.0097	0.0511	0.1889			
$N = 200$	B2S coef	0.7498	1.0009	1.0003	1.0004	0.9991	4.1245	$< 10^{-4}$	0.4975	
$T = 30$	se_boot	0.0009	0.0030	0.0030	0.0030	0.0187	0.0548			3079.78
	se_mixt	0.0004	0.0037	0.0033	0.0068	0.0181	0.0569			289.18
	rmse	0.0007	0.0031	0.0030	0.0030	0.0187	0.1360			
	GMM coef	0.7510	0.9942	0.9880	0.9894	0.9675	4.2862			5075.67
	se	0.0007	0.0026	0.0026	0.0026	0.0229	0.0736			
	rmse	0.0019	0.0080	0.0133	0.0119	0.0397	0.2955			
	QMLE coef	0.7497	1.0002	0.9996	0.9997	0.9989	4.1227			2486.84
	se	0.0007	0.0029	0.0030	0.0031	0.0188	0.0561			
	rmse	0.0007	0.0029	0.0029	0.0029	0.0188	0.1349			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\rho, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table C.2: Dynamic Random Effects World

$\varepsilon = 0.5$, $r = 0.5$, 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	1			
$N = 100$	B2S coef	0.7490	1.0026	0.9981	1.0020	0.9940	1.0378	$< 10^{-4}$	0.4991	
$T = 10$	se_boot	0.0017	0.0079	0.0080	0.0080	0.0481	0.0795			279.60
	se_mixt	0.0019	0.0079	0.0078	0.0080	0.0481	0.0787			19.84
	rmse	0.0020	0.0085	0.0083	0.0086	0.0485	0.0880			
	GMM coef	0.7497	1.0037	0.9874	1.0035	0.8737	1.4006			124.57
	se	0.0017	0.0074	0.0072	0.0076	0.0522	0.0907			
	rmse	0.0025	0.0103	0.0158	0.0105	0.1366	0.4108			
	QMLE coef	0.7494	1.0036	0.9917	1.0028	0.9951	0.9975			653.15
	se	0.0015	0.0069	0.0067	0.0068	0.0508	0.0768			
	rmse	0.0016	0.0078	0.0078	0.0078	0.0510	0.0768			
$N = 200$	B2S coef	0.7496	1.0006	1.0003	1.0003	0.9991	1.0333	$< 10^{-4}$	0.4997	
$T = 30$	se_boot	0.0007	0.0029	0.0029	0.0029	0.0187	0.0268			2350.81
	se_mixt	0.0019	0.0079	0.0078	0.0080	0.0481	0.0787			205.64
	rmse	0.0008	0.0030	0.0030	0.0030	0.0187	0.0427			
	GMM coef	0.7497	1.0037	0.9874	1.0035	0.8737	1.4006			4605.65
	se	0.0017	0.0074	0.0072	0.0076	0.0522	0.0907			
	rmse	0.0025	0.0103	0.0158	0.0105	0.1366	0.4108			
	QMLE coef	0.7496	1.0003	0.9994	0.9996	0.9990	1.0307			2545.70
	se	0.0006	0.0026	0.0027	0.0028	0.0188	0.0272			
	rmse	0.0006	0.0027	0.0027	0.0027	0.0188	0.0411			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\rho, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table C.3: Dynamic Random Effects World

$\varepsilon = 0.5$, $r = 0.2$, 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	0.25			
$N = 100$	B2S coef	0.7490	1.0017	0.9994	1.0016	0.9939	0.2779	$< 10^{-4}$	0.4985	
$T = 10$	se_boot	0.0016	0.0078	0.0078	0.0080	0.0481	0.0459			290.17
	se_mixt	0.0016	0.0068	0.0067	0.0069	0.0481	0.0452			18.89
	rmse	0.0020	0.0083	0.0081	0.0085	0.0485	0.0537			
	GMM coef	0.7499	1.0021	0.9931	1.0019	0.8734	0.5491			122.24
	se	0.0016	0.0063	0.0062	0.0064	0.0522	0.0536			
	rmse	0.0023	0.0091	0.0112	0.0093	0.1370	0.3039			
	QMLE coef	0.7497	1.0024	0.9943	1.0020	0.9959	0.2449			1101.57
	se	0.0014	0.0061	0.0059	0.0061	0.0508	0.0404			
	rmse	0.0014	0.0066	0.0066	0.0066	0.0509	0.0407			
$N = 200$	B2S coef	0.7496	1.0005	1.0004	1.0004	0.9991	0.2607	$< 10^{-4}$	0.4973	
$T = 30$	se_boot	0.0006	0.0029	0.0029	0.0029	0.0187	0.0139			2358.91
	se_mixt	0.0007	0.0029	0.0029	0.0028	0.0187	0.0138			207.69
	rmse	0.0008	0.0030	0.0030	0.0030	0.0187	0.0175			
	GMM coef	0.7499	1.0001	0.9981	0.9985	0.9640	0.3387			4619.38
	se	0.0004	0.0018	0.0019	0.0018	0.0226	0.0199			
	rmse	0.0008	0.0033	0.0038	0.0036	0.0425	0.0909			
	QMLE coef	0.7497	1.0004	0.9994	0.9996	0.9991	0.2572			2455.69
	se	0.0005	0.0023	0.0023	0.0024	0.0188	0.0137			
	rmse	0.0005	0.0023	0.0023	0.0023	0.0188	0.0155			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors of $\theta = (\rho, \beta')'$ computed with mixture of t-distributions of $\theta_*(b|g_0)$ and $\widehat{\theta}_{EB}(b|g_0)$.

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table C.4: 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			
$N = 100$	B2S coef	0.7489	1.0016	1.0014	0.9984	0.9997	162.5717	$< 10^{-4}$	0.4959	
$T = 10$	se_boot	0.0022	0.0079	0.0080	0.0230	0.0667	31.0964			349.40
	se_mixt	0.0029	0.0116	0.0116	0.0466	0.0450	23.0263			35.40
Replications=$1,000$	rmse	0.0023	0.0083	0.0083	0.0236	0.0667	31.0808			
	QMLE coef	0.7502	0.9990	0.9991	1.0044	1.0021	162.2586			1286.36
	se	0.0048	0.0183	0.0184	0.1966	0.0651	23.3529			
	rmse	0.0048	0.0183	0.0184	0.0183	0.0651	23.3429			
	true	0.75	1	1	1	1	204.3142			
$N = 200$	B2S coef	0.7496	1.0008	1.0007	0.9995	0.9971	203.2024	$< 10^{-4}$	0.4984	
$T = 30$	se_boot	0.0009	0.0030	0.0030	0.0090	0.0265	29.3042			2708.19
	se_mixt	0.0012	0.0047	0.0047	0.0191	0.0188	20.5737			254.19
Replications=$500^{(*)}$	rmse	0.0009	0.0032	0.0031	0.0094	0.0266	29.2960			
	QMLE coef	0.7499	0.9999	0.9999	0.9995	1.9166	204.3702			4335.38
	se	0.0008	0.0030	0.0029	0.0094	0.2765	20.5173			
	rmse	0.0008	0.0030	0.0030	0.0030	0.9574	20.4968			

B2S : Bayesian two-stage estimation. The parameters π_t are omitted from the table.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

QMLE: quasi-maximum likelihood estimation. The parameters π_t are omitted from the table.

(*) *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 C.5: Dynamic Hausman-Taylor World

$r = 0.8$, $\varepsilon = 0.5$, 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			
$N = 100$	B2S coef	0.7483	1.0007	1.0002	1.0011	1.0281	1.0394	0.9945	3.9468	$< 10^{-4}$	0.4999	
$T = 10$	se_boot	0.0031	0.0109	0.0109	0.0121	0.2305	0.0472	0.0639	0.6808			301.06
	se_mixt	0.0026	0.0100	0.0100	0.0159	0.0692	0.0221	0.0447	0.6620			25.53
	rmse	0.0035	0.0109	0.0109	0.0121	0.2322	0.0615	0.0641	0.6826			
	2SQML coef	0.7499	1.0000	0.9997	1.0003	1.0219	1.0003	0.9935	$n.a$			688.99
	se	0.0028	0.0092	0.0090	0.0091	0.2012	0.0703	0.0480	$n.a$			
	rmse	0.0028	0.0092	0.0092	0.0092	0.2023	0.0703	0.0484	$n.a$			
$N = 100$	B2S coef	0.7496	1.0003	1.0006	1.0005	1.0196	1.0130	0.9984	3.9421	$< 10^{-4}$	0.5005	
$T = 30$	se_boot	0.0015	0.0058	0.0058	0.0060	0.2092	0.0251	0.0383	0.6015			966.48
	se_mixt	0.0014	0.0056	0.0056	0.0078	0.0388	0.0124	0.0267	0.5944			104.60
	rmse	0.0016	0.0058	0.0058	0.0060	0.2102	0.0283	0.0383	0.6039			
	2SQML coef	0.7500	0.9999	1.0001	1.0001	1.0191	0.9987	0.9978	$n.a$			2528.13
	se	0.0011	0.0043	0.0042	0.0043	0.1972	0.0463	0.0273	$n.a$			
	rmse	0.0011	0.0043	0.0043	0.0043	0.1981	0.0463	0.0274	$n.a$			
$N = 200$	B2S coef	0.7484	1.0010	1.0009	1.0005	1.0057	1.0378	0.9972	3.9724	$< 10^{-4}$	0.4999	
$T = 10$	se_boot	0.0022	0.0078	0.0077	0.0085	0.1611	0.0324	0.0466	0.4585			1183.16
	se_mixt	0.0018	0.0069	0.0068	0.0113	0.0466	0.0151	0.0333	0.4519			106.52
	rmse	0.0027	0.0078	0.0078	0.0085	0.1612	0.0498	0.0467	0.4591			
	2SQML coef	0.7500	1.0001	1.0001	0.9998	0.9986	1.0013	0.9965	$n.a$			1242.89
	se	0.0020	0.0066	0.0063	0.0066	0.1427	0.0487	0.0355	$n.a$			
	rmse	0.0020	0.0066	0.0066	0.0066	0.1427	0.0487	0.0357	$n.a$			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

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

Table C.6: 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			
$N = 100$	B2S coef	0.7518	0.9923	0.9961	1.0048	$< 10^{-4}$	0.4594	
$T = 30$	se_boot	0.0014	0.0122	0.0154	0.0262			1775.99
	se_mixt	0.0010	0.0140	0.0185	0.0262			106.80
	rmse	0.0023	0.0144	0.0159	0.0267			
	CCEP coef	0.7487	1.0016	1.0063	0.9910			1437.60
	se	0.0017	0.0103	0.0122	0.0270			
	rmse	0.0021	0.0104	0.0137	0.0285			
$N = 50$	B2S coef	0.7501	1.0032	1.0026	1.0116	$< 10^{-4}$	0.4613	
$T = 50$	se_boot	0.0026	0.0152	0.0188	0.0492			1901.78
	se_mixt	0.0024	0.0161	0.0206	0.0316			113.04
	rmse	0.0026	0.0156	0.0190	0.0505			
	CCEP coef	0.7486	1.0007	1.0037	0.9807			1139.69
	se	0.0020	0.0115	0.0141	0.0280			
	rmse	0.0024	0.0116	0.0145	0.0340			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

CCEP: Common Correlated Effects Pooled estimator.

Table C.7: 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			
$N = 100$	B2S coef	0.7459	1.0199	1.0131	1.0289	$< 10^{-4}$	0.4809	
$T = 30$	se_boot	0.0039	0.0184	0.0198	0.0322			1996.72
	se_mixt	0.0033	0.0219	0.0289	0.0348			344.04
	rmse	0.0056	0.0270	0.0237	0.0432			
	CCEP coef	0.7487	1.0003	1.0044	1.0229			1392.09
	se	0.0020	0.0107	0.0124	0.0313			
	rmse	0.0023	0.0107	0.0131	0.0388			
$N = 50$	B2S coef	0.7474	1.0146	1.0068	1.0235	$< 10^{-4}$	0.4865	
$T = 50$	se_boot	0.0031	0.0174	0.0196	0.0606			1820.55
	se_mixt	0.0024	0.0169	0.0212	0.0281			177.83
	rmse	0.0040	0.0227	0.0208	0.0650			
	CCEP coef	0.7493	1.0011	1.0022	1.0148			1100.73
	se	0.0019	0.0116	0.0142	0.0315			
	rmse	0.0020	0.0117	0.0143	0.0348			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

CCEP: Common Correlated Effects Pooled estimator.

Table C.8: 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			
$N = 100$	B2S coef	0.7529	0.7690	1.0112	1.0977	0.0020	0.4160	
$T = 30$	se_boot	0.0135	0.0296	0.0184	0.0616			3059.25
	se_mixt	0.0097	0.0444	0.0531	0.0611			159.42
	rmse	0.0137	0.0353	0.0215	0.1155			
	CCEMG coef	0.7427	0.7543	1.0136	1.1818			1569.35
	se	0.0142	0.0261	0.0167	0.1347			
	rmse	0.0160	0.0265	0.0216	0.2263			
	true	0.7501	0.7503	1	1			
$N = 50$	B2S coef	0.7591	0.7936	1.0101	1.1808	0.0010	0.4407	
$T = 50$	se_boot	0.0217	0.0973	0.0643	0.0474			1566.68
	se_mixt	0.0087	0.0415	0.0500	0.1172			177.83
	rmse	0.0219	0.1572	0.0925	0.0963			
	CCEMG coef	0.7568	0.7776	1.0070	1.2473			1205.11
	se	0.0175	0.0383	0.0206	0.5980			
	rmse	0.0188	0.0471	0.0218	0.6468			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

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

Table C.9: 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	0.75	0.7496	1	1			
$N = 200$	B2S coef	0.7491	0.7616	1.0119	1.0752	0.0010	0.4107	
$T = 30$	se_Boot	0.0123	0.0224	0.0146	0.0506			6156.63
	se_mixt	0.0083	0.0382	0.0459	0.0501			776.54
	rmse	0.0123	0.0254	0.0188	0.0906			
	CCEMG coef	0.7377	0.7476	1.0147	1.1652			4367.99
	se	0.0126	0.0191	0.0113	0.0370			
	rmse	0.0175	0.0192	0.0185	0.1693			
$N = 200$	B2S coef	0.7532	0.7933	1.0179	1.1659	$< 10^{-4}$	0.4402	
$T = 50$	se_Boot	0.0157	0.0438	0.0256	0.1224			7046.55
	se_mixt	0.0063	0.0305	0.0369	0.1223			1159.68
	rmse	0.0160	0.0618	0.0312	0.2061			
	CCEMG coef	0.7524	0.7716	1.0113	1.1767			5260.12
	se	0.0148	0.0290	0.0127	0.2279			
	rmse	0.0150	0.0364	0.0170	0.2883			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

CCEMG: Common Correlated Effects Mean Group estimator.

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

	$N = 200, T = 30$		$N = 200, T = 50$	
	ρ_i	β_{1i}	ρ_i	β_{1i}
min	0.6015	0.5052	0.6015	0.5024
mean	0.7500	0.7496	0.7500	0.7496
sd	0.0866	0.1442	0.0866	0.1442
max	0.8985	0.9975	0.8985	0.9975

Table C.10: 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			
$N = 200$	B2S coef	0.7501	1.0011	1.0012	0.9990	1.0005	4.1091	$< 10^{-4}$	0.4999	
$T = 50$	se_boot	0.0007	0.0022	0.0023	0.0023	0.0147	0.0439			3910.89
	se_mixt	0.0007	0.0028	0.0028	0.0027	0.0147	0.0435			464.95
	rmse	0.0005	0.0025	0.0025	0.0024	0.0147	0.1176			
	QMLE coef	0.7499	1.0002	1.0003	0.9992	1.0002	4.1178			6150.10
	se	0.0005	0.0022	0.0022	0.0022	0.0147	0.0448			
	rmse	0.0005	0.0022	0.0022	0.0022	0.0147	0.1261			

B2S : Bayesian two-stage estimation.

se_boot: standard errors computed with individual block resampling bootstrap.

se_mixt: standard errors computed with mixture of t-distributions.

QMLE: quasi-maximum likelihood estimation.

D. Sensitivity to ε-contamination values

Tables D.1 and D.2 investigate the sensitivity of the *2S bootstrap* estimator for the random effects world and for the heterogeneous panel data world with common correlated effects[14] with respect to ε, the contamination part of the prior distributions, which varies between 0 and 90%. As shown in Table D.1 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 $R^2_{b_0}/\left(1-R^2_{b_0}\right) = R^2_{\widehat{b}_q}/\left(1-R^2_{\widehat{b}_q}\right)$ (see the definition of $\widehat{\lambda}_{b,h_0}$ in section 2.3.2 in the main text). 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 D.1 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^2_μ which has a fairly large upward bias (11.4%) as well as a large RMSE.[15]

Table D.2 shows similar results. All the parameter estimates are insensitive to ε ($\varepsilon \neq 0$) for the heterogeneous panel data world with common correlated 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^2_u which has also a fairly large upward bias (23.8%) 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 be why, even though the choice of $\varepsilon = 0.5$ is somewhat arbitrary, the adjustment compensates for it not being optimal (see Berger (1985)).

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

[15] 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.

Table D.1: 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	0.7489	1.0052	0.9793	1.0153	0.8746	4.8449			134.53
	se	0.0027	0.0093	0.0092	0.0263	0.0523	0.2152			
	rmse	0.0038	0.0137	0.0243	0.0315	0.1359	0.8718			
	QMLE coef	0.7485	1.0041	0.9927	1.0061	0.9931	4.0782			471.14
	se	0.0023	0.0088	0.0086	0.0236	0.0507	0.1720			
	rmse	0.0027	0.0097	0.0097	0.0097	0.0511	0.1889			
$\varepsilon = 0$	2S boot coef	0.7383	1.0168	0.9871	1.0116	0.9906	4.4573	1.0000	1.0000	317.02
	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	0.7497	1.0045	0.9955	1.0031	0.9948	4.0745	$< 10^{-4}$	1.0000	326.04
	se	0.0021	0.0083	0.0084	0.0084	0.0481	0.1565			
	rmse	0.0018	0.0094	0.0092	0.0090	0.0484	0.1732			
$\varepsilon = 0.1$	2S boot coef	0.7497	1.0045	0.9955	1.0031	0.9948	4.0745	$< 10^{-4}$	0.8978	322.75
	se	0.0021	0.0083	0.0084	0.0084	0.0481	0.1565			
	rmse	0.0018	0.0094	0.0092	0.0090	0.0484	0.1732			
$\varepsilon = 0.5$	2S boot coef	0.7497	1.0045	0.9955	1.0031	0.9948	4.0745	$< 10^{-4}$	0.4940	327.98
	se	0.0021	0.0083	0.0084	0.0084	0.0481	0.1565			
	rmse	0.0018	0.0094	0.0092	0.0090	0.0484	0.1732			
$\varepsilon = 0.9$	2S boot coef	0.7497	1.0045	0.9955	1.0031	0.9948	4.0745	$< 10^{-4}$	0.0979	310.67
	se	0.0021	0.0083	0.0084	0.0084	0.0481	0.1565			
	rmse	0.0018	0.0094	0.0092	0.0090	0.0484	0.1732			

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

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table D.2: 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	0.7427	0.7543	1.0136	1.1818			1569.35
	se	0.0142	0.0261	0.0167	0.1347			
	rmse	0.0160	0.0265	0.0216	0.2263			
$\varepsilon = 0$	2S boot coef	0.7521	0.7678	1.0108	1.2380	1.0000	1.0000	3105.32
	se	0.0135	0.0298	0.0188	0.4510			
	rmse	0.0136	0.0348	0.0217	0.5098			
$\varepsilon = 10^{-17}$	2S boot coef	0.7526	0.7688	1.0110	1.0977	0.4122	0.9890	3171.07
	se	0.0097	0.0444	0.0532	0.0611			
	rmse	0.0135	0.0349	0.0212	0.1152			
$\varepsilon = 0.1$	2S boot coef	0.7529	0.7690	1.0112	1.0977	0.0020	0.8158	3095.86
	se	0.0135	0.0296	0.0184	0.0616			
	rmse	0.0137	0.0353	0.0215	0.1155			
$\varepsilon = 0.5$	2S boot coef	0.7529	0.7690	1.0112	1.0977	0.0020	0.4160	3059.25
	se	0.0135	0.0296	0.0184	0.0616			
	rmse	0.0137	0.0353	0.0215	0.1155			
$\varepsilon = 0.9$	2S boot coef	0.7529	0.7690	1.0112	1.0977	0.0020	0.0789	2988.05
	se	0.0135	0.0296	0.0184	0.0616			
	rmse	0.0137	0.0353	0.0215	0.1155			

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 $\overline{\beta}_1 = (1/N)\sum_{i=1}^{N} \beta_{1i}$.

E. Departure from normality

Tables E.1 and E.2 investigate the robustness of the estimators to a non-normal framework for the random effects world and for the heterogeneous panel data world with common correlated 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)).[16] 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 E.1). 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 E.2 investigates the robustness of the CCEMG and *2S bootstrap* estimators to the right-skewed t-distribution for the heterogeneous panel data world with common correlated 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 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.

[16] 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.

Table E.1: 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			
2S boot coef	0.7445	1.0105	1.0020	1.0101	6.8157	4.4287	$< 10^{-4}$	0.4986	288.47
se	0.0051	0.0206	0.0207	0.0218	4.2378	0.7319			
rmse	0.0076	0.0235	0.0224	0.0249	4.2399	0.8479			
GMM coef	0.7509	1.0037	0.9703	1.0019	6.0394	6.2868			124.22
se	0.0036	0.0153	0.0146	0.0155	4.4155	1.0284			
rmse	0.0055	0.0217	0.0361	0.0217	4.5177	2.5072			
QMLE coef	0.7489	1.0075	0.9808	1.0061	6.8424	3.9816			468.18
se	0.0041	0.0170	0.0176	0.0176	4.4516	0.4327			
rmse	0.0042	0.0186	0.0186	0.0186	4.4524	0.4329			

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

GMM: Arellano-Bond GMM estimation.

QMLE: quasi-maximum likelihood estimation.

Table E.2: 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			
2S boot coef	0.6930	0.8788	1.2188	7.0220	0.0110	0.4202	2371.93
se	0.0222	0.1292	0.1018	2.5438			
rmse	0.0613	0.1825	0.2413	2.5428			
CCEMG coef	0.6300	0.8064	1.2112	7.3744			1569.28
se	0.0319	0.1364	0.1119	1.6597			
rmse	0.1242	0.1477	0.2390	6.5867			

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

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.

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

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

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

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

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

Kotz, S., Nadarajah, S., 2004. Multivariate t-distributions and their applications. Cambridge University Press.

McLachlan, G., Lee, S., 2013. EMMIXuskew: An R package for fitting mixtures of multivariate skew t distributions via the EM algorithm. Journal of Statistical Software 55, 1–22.

McLachlan, G., Peel, D., 2004. Finite Mixture Models. John Wiley & Sons.

Mengersen, K., Robert, C., Titterington, M., 2011. Mixtures: Estimation and Applications. Wiley Series in Probability and Statistics, John Wiley & Sons.

Peel, D., McLachlan, G., 2000. Robust mixture modelling using the t distribution. Statistics and Computing 10, 339–348.

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

Shrivastava, A., Chaturvedi, A., Bhatti, M.I., 2019. Robust Bayesian analysis of a multivariate dynamic model. Physica A: Statistical Mechanics and its Applications 528, 1–16.

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

Walker, G., Saw, J., 1978. The distribution of linear combinations of t-variables. Journal of the American Statistical Association 73, 876–878.

Robust dynamic panel data models using epsilon-contamination

Badi H. **Baltagi,** Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, U.S.A.
Georges **Bresson,** Département de science économique, Université Pair-II, France
Anoop **Chaturvedi,** Department of Statistics, University of Allahabad, India
Guy **Lacroix,** Département d'économique, Université Laval

This paper is written in honor of M. Hashem Pesaran for his many contributions to econometrics. In particular, heterogeneous panel data, Bayesian estimation of dynamic panel data models, random coefficient models for panel data and cross-section dependence in panels.

We would like to thank Jean-Michel Etienne for help and support with Stata Mata codes. We also thank an anonymous referee and Cheng Hsiao for useful comments and suggestions. Many thanks to the participants of the 2020 Econometric Society/Bocconi University Virtual World Congress for fruitful comments. Anoop Chaturvedi gratefully acknowledges the Department of Economics, Université Paris II for his visiting professorship and facilities to carry out this work. The usual disclaimers apply disclaimers apply

Chaire de recherche en **évaluation économique** des **programmes publics**

UNIVERSITÉ LAVAL

Secrétariat du Conseil du trésor Québec