Robust linear static panel data models using ε-contamination

Badi H. Baltagia
Georges Bresson
Anoop Chaturvedi
Guy Lacroix

Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

www.crrep.ca

Abstract

The paper develops a general Bayesian framework for robust linear static panel data models using ε-contamination. A two-step approach is employed to derive the conditional type-II maximum likelihood (ML-II) posterior distribution of the coeffcients and individual effects. The ML-II posterior densities are weighted averages of the Bayes estimator under a base prior and the data-dependent empirical Bayes estimator. Two-stage and three stage hierarchy estimators are developed and their finite sample performance is investigated through a series of Monte Carlo experiments. These include standard random effects as well as Mundlak-type, Chamberlain-type and Hausman-Taylor-type models. The simulation results underscore the relatively good performance of the three-stage hierarchy estimator. Within a single theoretical framework, our Bayesian approach encompasses a variety of specifications while conventional methods require separate estimators for each case.

Baltagi : Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, U.S.A. and Department of Economics, Leicester University, Leicester, UK - bbaltagi@maxwell.syr.edu

Bresson: Université Paris II / Sorbonne Universités, France - bresson-georges@orange.fr

Chaturvedi: University of Allahabad, India - anoopchaturv@gmail.com

Lacroix: Départment d'économique, Université Laval, Québec, Canada - Guy.Lacroix@ecn.ulaval.ca

1. Introduction

The choice of which classic panel data estimator to use in a linear static regression framework depends upon the hypothesized correlation between the individual effects and the regressors. One usually assumes either of two polar situations. The random effects model assumes that the regressors and the individual effects are uncorrelated. On the other hand, the fixed effects model assumes that all the regressors are correlated with the individual effects (see *e.g.* Mundlak (1978); Chamberlain (1982)). An intermediate situation arises when both a subset of time-varying and time-invariant regressors are assumed correlated with the individual effects, in which case the instrumental variables estimator of Hausman and Taylor (1981) is an appropriate alternative.

To a Bayesian analyst, the distinction between fixed, random and mixed models boils down to the specification of the number of stages in a given hierarchical model. While intuitively more attractive, the Bayesian approach nevertheless rests upon hypothesized prior distributions (and possibly on their hyperparameters). The choice of specific distributions is often made out of convenience rather than being based upon relevant subjective information.[1] 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 studies the robustness of Bayesian panel data models to possible misspecification of the prior distribution in the spirit of the works of Good (1965), Dempster (1977), Rubin (1977), Hill (1980), Berger (1985), Berger and Berliner (1984) and Berger and Berliner (1986) to mention a few. In particular, it is concerned with what they call the posterior robustness which is different from the robustness à la White (1980). The idea is to acknowledge the prior uncertainty by specifying a class Γ of possible prior distributions and then investigating the robustness of the posterior distribution as the prior varies over Γ. Several classes of priors have been proposed in the literature but the most commonly used one is the ε-contamination class. As mentioned by Berger (1985), the ε-contamination class of priors is particularly attractive to work with when investigating posterior robustness. The ε-contamination class combines the elicited prior, termed the base prior, with a contamination class of priors. This approach implements the Type-II maximum likelihood (ML-II) procedure for selecting the appropriate prior distribution. The primary benefit of using such a contamination class of prior distributions is that the resulting estimator performs well even if the elicited base prior distribution differs from the prior distribution that is consistent with the data.

The objective of our paper is to propose a robust Bayesian approach for linear static 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. The paper contributes to the 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, while Section 3 derives the Type-II maximum likeli-

---

[1]For instance, conventional proper priors in the normal linear model have been based on the conjugate Normal-Gamma family essentially because all the marginal likelihoods have closed-form solutions. Likewise, statisticians customarily assume that the variance-covariance matrix of the slope parameters follow a Wishart distribution because it is convenient from an analytical point of view.

hood posterior mean and the variance-covariance matrix of the coefficients in a two-stage hierarchy model. Section 4 introduces a three-stage hierarchy with generalized hyper-g priors on the variance-covariance matrix of the individual effects. The predictive densities corresponding to the base priors and the ε-contaminated priors turn out to be Gaussian and Appell hypergeometric functions, respectively. The main difference between the two-stage and the three-stage hierarchy models pertains to the definition of the Bayes estimators, the empirical Bayes estimators and the weights of the ML-II posterior means. Section 5 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 three-stage hierarchy estimator as compared to the standard frequentist estimation methods. Two applications on an earnings equation and a crime model, available in the supplementary appendix, illustrate and confirm the simulation results. Section 6 gives our conclusion.

2. The general setup

Let the Gaussian linear mixed model be written as:

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

where X'_{it} is a $(1 \times K_1)$ vector of explanatory variables including the intercept, and β is a $(K_1 \times 1)$ vector of parameters. 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. Finally, u_{it} is a remainder term assumed to be normally distributed, *i.e.* $u_{it} \sim N\left(0, \tau^{-1}\right)$. The distribution of u_{it} is parametrized in terms of its precision τ rather than its variance $\sigma_u^2 \left(= 1/\tau\right)$. In the statistics literature, the elements of β do not differ across i and are referred to as *fixed effects* whereas the b_i's are referred to as *random effects*.[2] This terminology differs from the one used in econometrics. In the latter, the b_i's are treated either as random variables, and hence referred to as random effects, or as constant but unknown parameters and hence referred to as fixed effects. In line with the econometrics terminology, whenever the b_i's are assumed to be correlated (uncorrelated) with all the X'_{it}s, they will be termed fixed (random) effects.[3]

In the Bayesian context, following the seminal papers of Lindley and Smith (1972) and Smith (1973), several authors have proposed a very general three-stage hierarchy framework to handle such models (see, *e.g.*, Chib and Carlin (1999); Koop (2003); Chib (2008); Greenberg (2008); Zheng et al. (2008); Rendon (2013)):

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

where y is $(NT \times 1)$, X is $(NT \times K_1)$, W is $(NT \times K_2)$, u is $(NT \times 1)$ and I_{NT} is a $(NT \times NT)$ identity matrix. The parameters depend upon hyperparameters which themselves follow random

---

[2]See Lindley and Smith (1972), Smith (1973), Laird and Ware (1982), Chib and Carlin (1999), Koop (2003), Chib (2008) and Greenberg (2008) to mention a few.

[3]When we write *fixed effects* in italics, we refer to the terminology of the statistical or Bayesian literature. Conversely, when we write fixed effects (in normal characters), we refer to the terminology of panel data econometrics.

distributions. The second stage (also called *fixed effects model* in the Bayesian literature) updates the distribution of the parameters. The third stage (also called *random effects model* in the Bayesian literature) updates the distribution of the hyperparameters. As stated by Smith (1973) (page 67) *"for the Bayesian model the distinction between fixed, random and mixed models, reduces to the distinction between different prior assignments in the second and third stages of the hierarchy"*. In other words, the *fixed effects model* is a model that does not have a third stage. 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[4] and the precision τ. As is well-known, Bayesian models may be 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 as arising from the design matrix of observables in the sample. Since the calculation of marginal likelihoods using a mixture of g-priors involves only a one-dimensional integral, this approach provides an attractive computational solution that made the original g-priors popular while insuring robustness to misspecification of g (see Zellner (1986) and Fernández et al. (2001) to mention a few). To guard against mispecifying the distributions of the priors, many suggest considering classes of priors (see Berger (1985)).

3. The robust linear static model in the two-stage hierarchy

Following Berger (1985), Berger and Berliner (1984, 1986), Zellner (1986), Moreno and Pericchi (1993), Chaturvedi (1996), and Chaturvedi and Singh (2012) among others, we consider the ε-contamination class of prior distributions for (β, b, τ):

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

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

$$\left\{ \begin{array}{ccl} \beta & \sim & N\left(\beta_0 \iota_{K_1}, (\tau g_0 \Lambda_X)^{-1}\right) \text{ with } \Lambda_X = X'X \\ b & \sim & N\left(b_0 \iota_{K_2}, (\tau h_0 \Lambda_W)^{-1}\right) \text{ with } \Lambda_W = W'W, \end{array} \right. \tag{4}$$

where ι_{K_1} is a $(K_1 \times 1)$ vector of ones. Furthermore, β_0, b_0, g_0 and h_0 are known scalar hyperparameters of the base prior $\pi_0\left(\beta, b, \tau \mid g_0, h_0\right)$. The probability density function (henceforth pdf) of

---

[4]Note that in (2), the prior distribution of β and b are assumed to be independent, so Var$[\theta]$ is block-diagonal with $\theta = (\beta', b')'$. The third stage can be extended by adding hyperparameters on the prior mean coefficients β_0 and b_0 and on the variance-covariance matrix of the β coefficients: $\beta_0 \sim N\left(\beta_{00}, \Lambda_{\beta_0}\right)$, $b_0 \sim N\left(b_{00}, \Lambda_{b_0}\right)$ and $\Lambda_\beta^{-1} \sim Wish\left(\nu_\beta, R_\beta\right)$ (see for instance, Koop (2003); Greenberg (2008); Hsiao and Pesaran (2008); Bresson and Hsiao (2011)).

the base prior $\pi_0(.)$ is given by:

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

The possible class of contamination Q is defined as:

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

with

$$\left\{ \begin{array}{ccl} \beta & \sim & N\left(\beta_q \iota_{K_1}, \left(\tau g_q \Lambda_X\right)^{-1}\right) \\ b & \sim & N\left(b_q \iota_{K_2}, \left(\tau h_q \Lambda_W\right)^{-1}\right), \end{array} \right. \tag{7}$$

where β_q, b_q, g_q and h_q are unknown. The restrictions $g_q \leq g_0$ 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 and two estimation strategies are possible:

1. a one-step estimation of the ML-II posterior distribution[5] of β, b and τ;
2. or a two-step approach as follows[6]:
   (a) Let $y^* = (y - Wb)$. Derive the conditional ML-II posterior distribution of β given the specific effects b.
   (b) Let $\widetilde{y} = (y - X\beta)$. Derive the conditional ML-II posterior distribution of b given the coefficients β.

We use the two-step approach because it simplifies the derivation of the predictive densities (or marginal likelihoods). In the one-step approach the pdf of y and the pdf of the base prior $\pi_0(\beta, b, \tau \mid g_0, h_0)$ need to be combined to get the predictive density. It thus leads to a complicated expression whose integration with respect to (β, b, τ) may be involved. Using a two-step approach we can integrate first with respect to (β, τ) given b and then, conditional on β, we can next integrate with respect to (b, τ). Thus, the marginal likelihoods (or predictive densities) corresponding to the base priors are:

$$m\left(y^* \mid \pi_0, b, g_0\right) = \int\limits_0^\infty \int\limits_{\mathbb{R}^{K_1}} \pi_0\left(\beta, \tau \mid g_0\right) \times p\left(y^* \mid X, b, \tau\right) \, d\beta \, d\tau$$

and

$$m\left(\widetilde{y} \mid \pi_0, \beta, h_0\right) = \int\limits_0^\infty \int\limits_{\mathbb{R}^{K_2}} \pi_0\left(b, \tau \mid h_0\right) \times p\left(\widetilde{y} \mid W, \beta, \tau\right) \, db \, d\tau,$$

---

[5] "*We consider the most commonly used method of selecting a hopefully robust prior in* Γ, *namely choice of that prior* π *which maximizes the marginal likelihood* $m\left(y \mid \pi\right)$ *over* Γ. *This process is called Type II maximum likelihood by Good (1965)*" (Berger and Berliner, 1986, page 463).

[6] We will see that the mean of the ML-II posterior of β (resp. of b) depends on draws from multivariate t-distributions. The two steps can thus be viewed as a Gibbs sampler.

with

$$\pi_0\left(\beta, \tau \mid g_0\right) = \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} \left|\Lambda_X\right|^{1/2} \exp\left(-\frac{\tau g_0}{2}\left(\beta - \beta_0 \iota_{K_1}\right)' \Lambda_X \left(\beta - \beta_0 \iota_{K_1}\right)\right),$$

$$\pi_0\left(b, \tau \mid h_0\right) = \left(\frac{\tau h_0}{2\pi}\right)^{\frac{K_2}{2}} \tau^{-1} \left|\Lambda_W\right|^{1/2} \exp\left(-\frac{\tau h_0}{2}\left(b - b_0 \iota_{K_2}\right)' \Lambda_W \left(b - b_0 \iota_{K_2}\right)\right).$$

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

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[7]:

$$m\left(y^* \mid \pi_0, b, g_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\beta, \tau \mid g_0\right) \times p\left(y^* \mid X, b, \tau\right) \, d\beta \, d\tau \tag{8}$$

$$= \widetilde{H} \left(\frac{g_0}{g_0+1}\right)^{K_1/2} \left(1 + \left(\frac{g_0}{g_0+1}\right)\left(\frac{R_{\beta_0}^2}{1 - R_{\beta_0}^2}\right)\right)^{-\frac{NT}{2}}$$

with $\widetilde{H} = \frac{\Gamma\left(\frac{NT}{2}\right)}{\pi\left(\frac{NT}{2}\right) v(b)\left(\frac{NT}{2}\right)}$, $R_{\beta_0}^2 = \frac{(\widehat{\beta}(b) - \beta_0 \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_0 \iota_{K_1})}{(\widehat{\beta}(b) - \beta_0 \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_0 \iota_{K_1}) + v(b)}$, $\widehat{\beta}(b) = \Lambda_X^{-1} X' y^*$ and $v(b) = (y^* - X\widehat{\beta}(b))'(y^* - X\widehat{\beta}(b))$, and where $\Gamma(\cdot)$ is the Gamma function.

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

$$m\left(y^* \mid q, b, g_0\right) = \widetilde{H}\left(\frac{g_q}{g_q+1}\right)^{\frac{K_1}{2}} \left(1 + \left(\frac{g_q}{g_q+1}\right)\left(\frac{R_{\beta_q}^2}{1 - R_{\beta_q}^2}\right)\right)^{-\frac{NT}{2}}, \tag{9}$$

where

$$R_{\beta_q}^2 = \frac{(\widehat{\beta}(b) - \beta_q \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_q \iota_{K_1})}{(\widehat{\beta}(b) - \beta_q \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_q \iota_{K_1}) + v(b)}.$$

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

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

and

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

---

[7]Derivation of all the following expressions can be found in the supplementary appendix.

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

$$\widehat{\beta}_q = \left(\iota'_{K_1} \Lambda_X \iota_{K_1}\right)^{-1} \iota'_{K_1} \Lambda_X \widehat{\beta}\left(b\right) \tag{10}$$

and

$$
\begin{aligned}
\widehat{g}_q &= \min\left(g_0, g^*\right), \\
\text{with } g^* &= \max\left[\left(\frac{(NT - K_1)}{K_1} \frac{(\widehat{\beta}\left(b\right) - \widehat{\beta}_q \iota_{K_1})' \Lambda_X (\widehat{\beta}\left(b\right) - \widehat{\beta}_q \iota_{K_1})}{v\left(b\right)} - 1\right)^{-1}, 0\right] \\
&= \max\left[\left(\frac{(NT - K_1)}{K_1}\left(\frac{R^2_{\widehat{\beta}_q}}{1 - R^2_{\widehat{\beta}_q}}\right) - 1\right)^{-1}, 0\right].
\end{aligned}
\tag{11}
$$

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

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

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

$$
\begin{aligned}
\widehat{\pi}^*\left(\beta \mid g_0\right) &= \int_0^\infty \widehat{\pi}^*\left(\beta, \tau \mid g_0\right) d\tau \\
&= \widehat{\lambda}_{\beta, g_0} \int_0^\infty \pi_0^*\left(\beta, \tau \mid g_0\right) d\tau + \left(1 - \widehat{\lambda}_{\beta, g_0}\right) \int_0^\infty q^*\left(\beta, \tau \mid g_0\right) d\tau \\
&= \widehat{\lambda}_{\beta, g_0} \pi_0^*\left(\beta \mid g_0\right) + \left(1 - \widehat{\lambda}_{\beta, g_0}\right) \widehat{q}^*\left(\beta \mid g_0\right)
\end{aligned}
\tag{12}
$$

with

$$\widehat{\lambda}_{\beta, g_0} = \left[1 + \frac{\varepsilon}{1 - \varepsilon}\left(\frac{\frac{\widehat{g}_q}{\widehat{g}_q + 1}}{\frac{g_0}{g_0 + 1}}\right)^{K_1/2}\left(\frac{1 + \left(\frac{g_0}{g_0 + 1}\right)\left(\frac{R^2_{\beta_0}}{1 - R^2_{\beta_0}}\right)}{1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q + 1}\right)\left(\frac{R^2_{\widehat{\beta}_q}}{1 - R^2_{\widehat{\beta}_q}}\right)}\right)^{\frac{NT}{2}}\right]^{-1}.$$

Note that $\widehat{\lambda}_{\beta, g_0}$ depends upon the ratio of the $R^2_{\beta_0}$ and $R^2_{\widehat{\beta}_q}$, but primarily on the sample size NT. Indeed, $\widehat{\lambda}_{\beta, g_0}$ tends to 0 when $R^2_{\beta_0} > R^2_{\widehat{\beta}_q}$ and tends to 1 when $R^2_{\beta_0} < R^2_{\widehat{\beta}_q}$, irrespective of the model fit (*i.e*, the absolute values of $R^2_{\beta_0}$ or $R^2_{\widehat{\beta}_q}$). Only the relative values of $R^2_{\widehat{\beta}_q}$ and $R^2_{\beta_0}$ matter.

It can be shown that $\pi_0^*\left(\beta \mid g_0\right)$ is the pdf (see the supplementary appendix) of a multivariate t-distribution with mean vector $\beta_*(b \mid g_0)$, variance-covariance matrix $\left(\frac{\xi_{0,\beta} M_{0,\beta}^{-1}}{NT - 2}\right)$ and degrees of

6

freedom (NT) with

$$M_{0,\beta} = \frac{(g_0 + 1)}{v(b)} \Lambda_X \text{ and } \xi_{0,\beta} = 1 + \left(\frac{g_0}{g_0 + 1}\right)\left(\frac{R^2_{\beta_0}}{1 - R^2_{\beta_0}}\right). \tag{13}$$

$\beta_*(b \mid g_0)$ is the Bayes estimate of β for the prior distribution $\pi_0(\beta, \tau)$:

$$\beta_*(b \mid g_0) = \frac{\widehat{\beta}(b) + g_0\beta_0\iota_{K_1}}{g_0 + 1}. \tag{14}$$

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

$$\xi_{q,\beta} = 1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q + 1}\right)\left(\frac{R^2_{\widehat{\beta}_q}}{1 - R^2_{\widehat{\beta}_q}}\right) \text{ and } M_{q,\beta} = \left(\frac{(\widehat{g}_q + 1)}{v(b)}\right)\Lambda_X, \tag{15}$$

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

$$\widehat{\beta}_{EB}(b \mid g_0) = \frac{\widehat{\beta}(b) + \widehat{g}_q\widehat{\beta}_q\iota_{K_1}}{\widehat{g}_q + 1}. \tag{16}$$

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

$$
\begin{aligned}
\widehat{\beta}_{ML-II} &= E\left[\widehat{\pi}^*(\beta \mid g_0)\right] \\
&= \widehat{\lambda}_{\beta,g_0}E\left[\pi_0^*(\beta \mid g_0)\right] + \left(1 - \widehat{\lambda}_{\beta,g_0}\right)E\left[\widehat{q}^*(\beta \mid g_0)\right] \\
&= \widehat{\lambda}_{\beta,g_0}\beta_*(b \mid g_0) + \left(1 - \widehat{\lambda}_{\beta,g_0}\right)\widehat{\beta}_{EB}(b \mid g_0).
\end{aligned} \tag{17}
$$

The ML-II posterior density of β, given b and g_0 is a shrinkage estimator. It is a weighted average of the Bayes estimator $\beta_*(b \mid g_0)$ under base prior g_0 and the data-dependent empirical Bayes estimator $\widehat{\beta}_{EB}(b \mid g_0)$. If the base prior is consistent with the data, the weight $\widehat{\lambda}_{\beta,g_0} \to 1$ and the ML-II posterior density of β gives more weight to the posterior $\pi_0^*(\beta \mid g_0)$ derived from the elicited prior. In this case $\widehat{\beta}_{ML-II}$ is close to the Bayes estimator $\beta_*(b \mid g_0)$. Conversely, if the base prior is not consistent with the data, the weight $\widehat{\lambda}_{\beta,g_0} \to 0$ and the ML-II posterior density of β is then close to the posterior $\widehat{q}^*(\beta \mid g_0)$ and to the empirical Bayes estimator $\widehat{\beta}_{EB}(b \mid g_0)$. The ability of the ε-contamination model to extract more information from the data is what makes it superior to the classical Bayes estimator based on a single base prior. Following Berger (1985)(page 207), we derive the analytical ML-II posterior variance-covariance matrix of β in the supplementary appendix.

3.2. The second step of the robust Bayesian estimator

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

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

with

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

where

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

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

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

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

and

$$\widehat{h}_q = \min(h_0, h^*)$$

$$\text{with } h^* = \max\left[\left(\frac{(NT - K_2)}{K_2}\frac{(\widehat{b}(\beta) - \widehat{b}_q\iota_{K_2})'\Lambda_W(\widehat{b}(\beta) - \widehat{b}_q\iota_{K_2})}{v(\beta)} - 1\right)^{-1}, 0\right]$$

$$= \max\left[\left(\frac{(NT - K_2)}{K_2}\left(\frac{R^2_{\widehat{b}_q}}{1 - R^2_{\widehat{b}_q}}\right) - 1\right)^{-1}, 0\right].$$

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

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

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

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

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

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

and where $\widehat{b}_{EB}\left(\beta \mid h_0\right)$ is the empirical Bayes estimator of b for the contaminated prior distribution $q\left(b, \tau \mid h_0\right)$:

$$\widehat{b}_{EB}\left(\beta \mid h_0\right) = \frac{\widehat{\beta}(b) + \widehat{h}_q \widehat{b}_q \iota_{K_2}}{\widehat{h}_q + 1}.$$

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

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

The ML-II posterior variance-covariance matrix of b can be derived in a similar fashion[8] to that of $\widehat{\beta}_{ML-II}$.

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

Many have raised concerns about the unbiasedness of the posterior variance-covariance matrices of $\widehat{\beta}_{ML-II}$ and $\widehat{b}_{ML-II}$. Indeed, they will both be biased towards zero as $\widehat{\lambda}_{\beta,g_0}$ 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 $\widehat{\beta}_{ML-II}$ 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. In what follows, we propose two different strategies to approximate these, each with different desirable properties.

3.3.1. MCMC with multivariate **t**-distributions

Recall that the ML-II posterior densities of β and b, $\widehat{\pi}^*(\beta|g_0)$ and $\widehat{\pi}^*(b|h_0)$, are both multivariate t-distributions. We thus propose to use the following algorithm to approximate the variance matrices once the model has been estimated:

1. Loop over D draws
2. In the first step of the model, draw a set of K_1 values from the multivariate t-distributions $\pi_0^*\left(\beta \mid g_0\right)$ and $\widehat{q}^*\left(\beta\right)$ to get a $\left(K_1 \times 1\right)$ vector of β_d as

$$\widehat{\pi}^*\left(\beta \mid g_0\right) = \widehat{\lambda}_{\beta,g_0} \pi_0^*\left(\beta \mid g_0\right) + \left(1 - \widehat{\lambda}_{\beta,g_0}\right) \widehat{q}^*\left(\beta \mid g_0\right)$$

3. In the second step of the model, draw a set of N values from the multivariate t-distributions $\pi_0^*\left(b \mid h_0\right)$ and from $q^*\left(b \mid h_0\right)$ to get a $\left(N \times 1\right)$ vector of b_d as

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

4. Once the D draws are completed, use $D^*(= D - D_{burn})$ draws to compute

$$\widehat{\beta}_{ML-II} = E\left[\beta^{(D^*)}\right] , \qquad \widehat{\sigma}_{\beta_{ML-II}} = \sqrt{diag\left(Var\left[\beta^{(D^*)}\right]\right)}$$

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

where D_{burn} are the burn in draws.

---

[8]See the supplementary appendix.

9

3.3.2. Block resampling bootstrap

As an alternative to MCMC simulations, we follow Laird and Louis (1987) and directly bootstrap the data. In the context of panel data, various strategies may be used, and chief among them is the so-called individual block resampling bootstrap (see, *e.g.* Bellman et al. (1989); Andersson and Karlsson (2001); Kapetanios (2008)). Thus for an $(N \times T)$ matrix Y, individual block resampling consists in drawing an $(N \times T)$ matrix Y^{BR} whose rows are obtained by resampling those of Y with replacement. Conditionally on Y, the rows of Y^{BR} are independent and identically distributed.[9] We use such a procedure for all the dependent Y and explanatory X variables. Fortunately, we need as few as $BR = 20$ bootstrap samples to achieve acceptable results.[10] 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 β using our initial shrinkage procedure

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

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

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

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

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

4. The robust linear static model in the three-stage hierarchy

As stressed earlier, the Bayesian literature introduces a third stage in the hierarchical model in order to discriminate between *fixed effects* and *random effects*. Hyperparameters can be defined for the mean and the variance-covariance of b (and sometimes β). Our goal in this paper is to consider a contamination class of priors to account for uncertainty pertaining to the base prior $\pi_0(\beta, b, \tau)$, *i.e.*, uncertainty about the prior means of the base prior. Consequently, assuming hyper priors for the means β_0 and b_0 of the base prior is tantamount to assuming the mean of the base prior to be unknown, which is contrary to our initial assumption. Following Chib and Carlin (1999), Chib (2008), Greenberg (2008), Zheng et al. (2008) among others, hyperparameters only concern

---

[9]Note that this assumes homoskedasticity as defined in (2). We do not account for heteroskedasticity in the block bootstrap.

[10]For convenience, the number of bootstrap samples BR is relatively small compared to the sample size N. Increasing the number of bootstrap samples does not change the results but increases the computation time considerably. Note also that this approach is many times less computationally intensive than both the 2S with MCMC on the multivariate t-distributions and the full Bayesian estimator.

the variance-covariance matrix of the b coefficients. Because we use g-priors in the second stage for β and b, g_0 is kept fixed and assumed known. We need only define mixtures of g-priors on the precision matrix of b, or equivalently on h_0.

Maruyama and George (2011, 2014) have proposed the following generalized hyper-g prior :

$$p\left(g\right) = \frac{g^{c-1}\left(1+g\right)^{-(c+d)}}{B\left(c,d\right)} \ , \ c > 0, \ d > 0, \tag{18}$$

where $B\left(\cdot\right)$ is the Beta function. This Beta-prime (or Pearson Type-VI) hyper prior for g is a generalization of the so-called Pareto type-II hyper-g. Using the generalized hyper-g prior specification, the three-stage hierarchy of the model can be defined as:

$$\text{First stage}: y \sim N\left(X\beta + Wb, \Sigma\right), \ \Sigma = \tau^{-1}I_{NT} \tag{19}$$

$$\text{Second stage}: \beta \sim N\left(\beta_0\iota_{K_1}, \left(\tau g_0\Lambda_X\right)^{-1}\right) \ , \ b \sim N\left(b_0\iota_{K_2}, \left(\tau h_0\Lambda_W\right)^{-1}\right)$$

$$\text{Third stage}: h_0 \sim \beta'(c,d) \ \rightarrow p\left(h_0\right) = \frac{h_0^{c-1}\left(1+h_0\right)^{-(c+d)}}{B\left(c,d\right)} \ , \ c > 0, \ d > 0.$$

We thus use hyperparameters only on the variance $\left(\tau h_0\Lambda_W\right)^{-1}$, *i.e.* only on h_0. Furthermore, we do not introduce an ε-contamination class of prior distributions for this hyperparameter of the third stage of the hierarchy, for example, $p\left(h_0\right) = \left(1 - \varepsilon_a\right)\pi_0\left(h_0\right) + \varepsilon_a q\left(h_0\right)$, as our objective is to account for the uncertainty about the prior means of the base prior $\pi_0\left(\beta, b, \tau\right)$. The third stage does depend on the priors, so the specification uses Gaussian distributions for the parameters at the second stage and a Beta-prime distribution for the priors at the third stage. Moreover, Berger (Berger, 1985, page 232) has stressed that the choice of a specific functional form for the third stage matters little.[11] Therefore we restrict the ε-contamination class of prior distributions to the first stage prior only (the second stage of the hierarchy, *i.e.*, for (β, b, τ)).

The first step of the robust Bayesian estimator in the three-stage hierarchy is strictly similar to the one in the two-stage hierarchy. But the three-stage hierarchy differs from the two-stage hierarchy in that it introduces a generalized hyper-g prior on h_0. The unconditional predictive density corresponding to the base prior is then given by

$$
\begin{aligned}
m\left(\widetilde{y} \mid \pi_0, \beta\right) &= \int_0^\infty m\left(\widetilde{y} \mid \pi_0, \beta, h_0\right)p(h_0)dh_0 \\
&= \frac{\widetilde{H}}{B\left(c,d\right)}\int_0^1 \left(\varphi\right)^{\frac{K_2}{2}+c-1}\left(1-\varphi\right)^{d-1}\left(1+\varphi\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right)^{-\frac{NT}{2}}d\varphi
\end{aligned}
$$

which can be written as:

$$m\left(\widetilde{y} \mid \pi_0, \beta\right) = \frac{B(d, \frac{K_2}{2}+c)}{B\left(c,d\right)}\widetilde{H} \times_2 F_1\left(\frac{NT}{2}; \frac{K_2}{2}+c; \frac{K_2}{2}+c+d; -\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right), \tag{20}$$

---

[11]Sinha and Jayaraman (2010a,b) studied a ML-II contaminated class of priors at the third stage of hierarchical priors using normal, lognormal and inverse Gaussian distributions to investigate the robustness of Bayes estimates with respect to possible misspecification at the third stage. Their results confirmed Berger (1985)'s assertion that the form of the second stage prior (the third stage of the hierarchy) does not affect the Bayes decision.

where $_2F_1(.)$ is the Gaussian hypergeometric function (see Abramovitz and Stegun (1970) and the supplementary appendix). As shown by Liang et al. (2008), numerical overflow is problematic for moderate to large NT and large $R_{b_0}^2$. As the Laplace approximation involves an integral with respect to a normal kernel, we can follow the suggestion of Liang et al. (2008) and develop an expansion[12] after a change of variable given by $\log\left(\frac{h_0}{h_0+1}\right)$.

Similar to the conditional predictive density corresponding to the contaminated prior on β (see equation (9)), the unconditional predictive density corresponding to the contaminated prior on b can be shown to give:

$$
m\left(\widetilde{y}\mid\widehat{q},\beta\right)=\frac{\widetilde{H}}{B\left(c,d\right)}\left\{
\begin{array}{c}
\frac{2\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c}}{K_2+2c} \\
\times F_1\left(\frac{K_2}{2}+c;1-d;\frac{NT}{2};\frac{K_2}{2}+c+1;\frac{h^*}{h^*+1};-\frac{h^*}{h^*+1}\left(\frac{R_{\widehat{b}_q}^2}{1-R_{\widehat{b}_q}^2}\right)\right) \\
+\left\{
\begin{array}{c}
\left[\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{\widehat{b}_q}^2}{1-R_{\widehat{b}_q}^2}\right)\right)\right]^{-\frac{NT}{2}} \\
\times\left[
\begin{array}{c}
B\left(c,d\right)-\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\
\times_2F_1\left(c;d-1;c+1;\frac{h^*}{h^*+1}\right)
\end{array}
\right]
\end{array}
\right\}
\end{array}
\right\},\quad(21)
$$

where $F_1(\cdot)$ is the Appell hypergeometric function (see Appell (1882); Slater (1966); Abramovitz and Stegun (1970) and the supplementary appendix). $m\left(\widetilde{y}\mid\widehat{q},\beta\right)$ can also be approximated using the same clever transformation as in Liang et al. (2008)(see the supplementary appendix).

We have shown earlier that the posterior density of (b,τ) for the base prior $\pi_0\left(b,\tau\mid h_0\right)$ in the two-stage hierarchy model is given by:

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

with

$$
\widehat{\lambda}_{b,h_0}=\frac{(1-\varepsilon)\,m\left(\widetilde{y}\mid\pi_0,\beta,h_0\right)}{(1-\varepsilon)\,m\left(\widetilde{y}\mid\pi_0,\beta,h_0\right)+\varepsilon m\left(\widetilde{y}\mid\widehat{q},\beta,h_0\right)}.
$$

Hence, we can write

$$
\widehat{\lambda}_b=\int_0^\infty\widehat{\lambda}_{b,h_0}p(h_0)dh_0=\left[1+\left(\frac{\varepsilon}{1-\varepsilon}\right)\cdot\frac{m\left(\widetilde{y}\mid\widehat{q},\beta\right)}{m\left(\widetilde{y}\mid\pi_0,\beta\right)}\right]^{-1}.\quad(22)
$$

Therefore, under the base prior, the Bayes estimator of b in the three-stage hierarchy model is given by:

$$
b_*\left(\beta\right)=\int_0^\infty b_*\left(\beta\mid h_0\right)p(h_0)dh_0=\frac{1}{c+d}\left[d\cdot\widehat{b}\left(\beta\right)+c\cdot b_0\iota_{K_2}\right].
$$

---

[12]See the supplementary appendix. For the Monte Carlo simulation study and the empirical applications, instead of using Gaussian hypergeometric integrals $_2F_1$ and Appel integrals F_1 with Laplace approximations, we prefer to solve the integrals numerically with adaptive quadrature methods.

Likewise, under the contamination class of priors, the empirical Bayes estimator of b for the three-stage hierarchy model is given by

$$\widehat{b}_{EB}(\beta) = \int_0^\infty \widehat{b}_{EB}(\beta \mid h_0) \, p(h_0) dh_0 \tag{23}$$

$$= \frac{1}{B(c,d)} \left[\begin{array}{c} \widehat{b}(\beta) \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1\left(c; -d; c+1; \frac{h^*}{h^*+1}\right) \\ + \widehat{b}_q \iota_{K_2} \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1\left(c+1; 1-d; c+2; \frac{h^*}{h^*+1}\right) \\ + \left\{ \widehat{b}(\beta)\left(\frac{1}{h^*+1}\right) + \widehat{b}_q \iota_{K_2}\left(\frac{h^*}{h^*+1}\right) \right\} \\ \times \left[\begin{array}{c} B(c,d) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2 F_1\left(c; d-1; c+1; \frac{h^*}{h^*+1}\right) \end{array} \right] \end{array} \right]$$

and the ML-II posterior density of b is given by:

$$\widehat{\pi}^*(b) = \int_0^\infty \widehat{\pi}^*(b, \tau) \, d\tau = \widehat{\lambda}_b \int_0^\infty \pi_0^*(b, \tau) \, d\tau + \left(1 - \widehat{\lambda}_b\right) \int_0^\infty q^*(b, \tau) \, d\tau$$

$$= \widehat{\lambda}_b \pi_0^*(b) + \left(1 - \widehat{\lambda}_b\right) \widehat{q}^*(b).$$

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

$$M_{0,b} = \frac{(h_0 + 1)}{v(\beta)} \Lambda_W \text{ and } \xi_{0,b} = 1 + \left(\frac{h_0}{h_0 + 1}\right) \left(\frac{R_{b_0}^2}{1 - R_{b_0}^2}\right).$$

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

$$\xi_{q,b} = 1 + \left(\frac{\widehat{h}_q}{\widehat{h}_q + 1}\right) \left(\frac{R_{\widehat{b}_q}^2}{1 - R_{\widehat{b}_q}^2}\right) \text{ and } M_{q,b} = \left(\frac{(\widehat{h}_q + 1)}{v(\beta)}\right) \Lambda_W.$$

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

$$\begin{aligned} \widehat{b}_{ML-II} &= E\left[\widehat{\pi}^*(b)\right] = \widehat{\lambda}_b E\left[\pi_0^*(b)\right] + \left(1 - \widehat{\lambda}_b\right) E\left[\widehat{q}^*(b)\right] \\ &= \widehat{\lambda}_b b_*(\beta) + \left(1 - \widehat{\lambda}_b\right) \widehat{b}_{EB}(\beta). \end{aligned} \tag{24}$$

The ML-II posterior variance-covariance matrix of b can be derived as in the two-stage hierarchy model. The main difference with the latter relates to the definition of the Bayes estimator $b_*(\beta)$, the empirical Bayes estimator $\widehat{b}_{EB}(\beta)$ and the weights $\widehat{\lambda}_b$ (as compared to $b_*(\beta \mid h_0)$, $\widehat{b}_{EB}(\beta \mid h_0)$ and $\widehat{\lambda}_{b,h_0}$). Since the variance-covariance matrix of both β and b are likely underestimated (see the

supplementary appendix) for the same reasons as those raised previously, they are computed using both MCMC with multivariate t-distributions and block resampling bootstrap.

As stressed earlier, the two-stage and three-stage hierarchical models correspond to *fixed* and *random* effects models, respectively. In the classical framework, these models refer to entirely different assumptions about the data generating process. Bayesian analysis with non-hierarchical priors is analogous to the fixed effects model in panel data econometrics while introducing hierarchical priors leads to the random effects model in panel data econometrics. These specifications can be tested using the Chib method of marginal likelihood (see Chib (1995) and Koop (2003)). In particular, the maximum value of this marginal likelihood determines the choice between the two-stage (FE) and the three-stage (RE) estimators. Adaptation of the Chib method to our ML-II estimator is outside the scope of our study. Our estimator needs to be compared to classical frequentist estimators as well as the standard full hierarchical model in order to assess its relative efficiency. In the next section, we perform a series of simulations to unearth the relative merit of each estimator.

5. A Monte Carlo simulation study

5.1. The DGP of the Monte Carlo study

Following Baltagi et al. (2003, 2009) and Baltagi and Bresson (2012), consider the static linear model:

$$
\begin{aligned}
y_{it} &= x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + Z_{1,i}\eta_1 + Z_{2,i}\eta_2 + \mu_i + u_{it}, \\
\text{for } i &= 1,...,N \;,\; t = 1,...,T, \text{ with}
\end{aligned}
$$

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

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

$$
\begin{aligned}
\eta_1 &= \eta_2 = 0 \\
x_{2,it} &= 0.7x_{2,it-1} + \kappa_i + \vartheta_{it} \;,\; (\kappa_i,\vartheta_{it}) \sim U(-2,2) \\
\mu_i &\sim N\left(0,\sigma_\mu^2\right) \;,\; \rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \tau^{-1}} = 0.3,\; 0.8.
\end{aligned}
$$

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 Mundlak-type fixed effects (FE) world, we assume that:

$$
\begin{aligned}
\eta_1 &= \eta_2 = 0; \\
x_{2,it} &= \delta_{2,i} + \omega_{2,it} \;,\; \delta_{2,i} \sim N(m_{\delta_2},\sigma_{\delta_2}^2),\; \omega_{2,it} \sim N(m_{\omega_2},\sigma_{\omega_2}^2); \\
m_{\delta_2} &= m_{\omega_2} = 1,\; \sigma_{\delta_2}^2 = 8,\; \sigma_{\omega_2}^2 = 2; \\
\mu_i &= \overline{x}_{2,i}\pi + \nu_i,\; \nu_i \sim N(0,\sigma_\nu^2),\; \overline{x}_{2,i} = \frac{1}{T}\sum_{t=1}^{T} x_{2,it}; \\
\sigma_\nu^2 &= 1,\; \pi = 0.8.
\end{aligned}
$$

14

$x_{1,1,it}$ and $x_{1,2,it}$ are assumed to be exogenous but $x_{2,it}$ is correlated with the μ_i and we assume a constant correlation coefficient, $\pi = 0.8$.

3. 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.

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

$$
\begin{aligned}
\eta_1 &= \eta_2 = 1; \\
x_{2,it} &= 0.7x_{2,it-1} + \mu_i + \vartheta_{it} \ , \ \vartheta_{it} \sim U(-2, 2); \\
Z_{1,i} &= 1, \ \forall i; \\
Z_{2,i} &= \mu_i + \delta_i + \theta_i + \xi_i, \ \xi_i \sim U(-2, 2); \\
\mu_i &\sim N\left(0, \sigma_\mu^2\right) \ , \text{ and } \rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \tau^{-1}} = 0.3, \ 0.8.
\end{aligned}
$$

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

For each set-up, we vary the size of the sample and the duration of the panel. We choose several (N, T) pairs with $N = 100, 500$ and $T = 5, 10$. We set the initial values of $x_{1,1,it}$ and $x_{1,2,it}$ to zero. We next generate $x_{1,1,it}, x_{1,2,it}, u_{it}, \zeta_{it}, u_{it}, \varsigma_{it}, \omega_{2,it}$ 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 (2S) and for the three-stage hierarchy (3S) are estimated with $\varepsilon = 0.5$, though we also investigate their robustness to various values of ε.[13]

We must set the hyperparameters values $\beta_0, b_0, g_0, h_0, \tau$ for the initial distributions of $\beta \sim N\left(\beta_0\iota_{K_1}, (\tau g_0 \Lambda_X)^{-1}\right)$ and $b \sim N\left(b_0\iota_{K_2}, (\tau h_0 \Lambda_W)^{-1}\right)$. While we can choose arbitrary values for β_0, b_0 and τ, the literature generally recommends using the unit information prior (UIP) to set the g-priors.[14] In the normal regression case, and following Kass and Wasserman (1995), the UIP corresponds to $g_0 = h_0 = 1/NT$, leading to Bayes factors that behave like the Bayesian Information Criterion (BIC).

For the three-stage hierarchy (3S), we need to choose the coefficients (c, d) of the generalized hyper-g priors. Following Liang et al. (2008) we set $c = d = 1$ for the Beta-prime distribution. In that case, the density is shaped as a hyperbola. In order to have the same shape under the UIP

---

[13]$\varepsilon = 0.5$ is an arbitrary value. We implicitly assume 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). We could have chosen any value for ε.

[14]We chose: $\beta_0 = 0, b_0 = 0$ and $\tau = 1$.

principle (*i.e.*, h_0 close to $1/NT$), we chose $c = 0.1$ and $d = 1$. For the three-stage hierarchy (3S), we solve the integrals numerically with adaptive quadrature methods (Davis and Rabinowitz, 1984; Press et al., 2007).

Most simulations are based upon $N = 100$, $T = 5$, $\rho = 0.8$ and $\varepsilon = 0.5$, although we investigate the robustness of our estimators to different parameter sets. For the 2S and 3S robust estimators, we use $BR = 20$ samples in the block resampling bootstrap and $D = 1,000$ with $D_{burn} = 500$ when drawing from the multivariate t-distributions. Finally, the Full Bayesian (FB) estimators are not derived in the paper for the sake of brevity but are presented in the supplementary appendix. When estimating the FB models, we also use $D=1,000$ draws with $D_{burn} = 500$.

For each experiment, we run $R = 1,000$ replications and we compute the means, the standard errors and the root mean squared errors (RMSEs) of the coefficients.

5.2. The results of the Monte Carlo study

In its most general form, the Gaussian linear mixed model can be written as (see equation (1)):

$$y = X\beta + Wb + u. \tag{25}$$

In what follows, we show how the usual classical estimators can be rewritten as a transformation of this model. We then estimate the classical and Bayesian counterparts and compare their properties.

5.2.1. The random effects world

Rewrite the general model (25) as follows:

$$y = X\beta + Z_\mu\mu + u,$$

where $u \sim N(0, \Sigma)$, $\Sigma = \tau^{-1}I_{NT}$, $Z_\mu = I_N \otimes \iota_T$ is $(NT \times N)$, $\otimes$ 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, $\mu \sim N\left(0, \sigma_\mu^2 I_N\right)$, are associated with the error term $\nu = Z_\mu\mu + u$ with $\text{Var}(\nu) = \sigma_\mu^2\left(I_N \otimes J_T\right) + \sigma_u^2 I_{NT}$, where $J_T = \iota_T\iota_T'$. This model is customarily estimated using Feasible Generalized Least Squares (FGLS)(see Hsiao (2003) or Baltagi (2013)).

Table 1 reports the results of fitting the classical FGLS model along with those from the 2S, 3S and Full Bayesian (FB) models, each in a separate panel. The true parameter values appear on the first line of the table. The last column reports the computation time in seconds.[15] Note that the computation time increases significantly as we move from the FGLS to the 2S-3S (bootstrap), to the 2S-3S (t-dist) and finally to the FB models. The performance of the robust estimators with block resampling is nevertheless within an acceptable range.

The first noteworthy feature of the table is that all the estimators yield essentially the same parameter estimates, standard errors[16] and RMSEs. The main difference between the 2S and the 3S lies in the estimated value of the weight $\widehat{\lambda}_\mu$, which tends towards zero in the 3S. This implies that the empirical Bayes estimator $\hat{b}_{EB}(\beta)$ (or equivalently $\hat{\mu}_{EB}(\beta)$) accounts for almost 100% of the weight in estimating the individual specific effects μ_i while the empirical Bayes estimator $\hat{\beta}_{EB}(b)$

---

[15]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.

[16]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.

(or equivalently $\hat{\beta}_{EB}(\mu)$) accounts for about 95% of the weight in estimating the β coefficients. For the two-stage hierarchical robust Bayesian estimator, the base prior for the specific effects b (or equivalently μ) is "more consistent" with the data as compared to the base prior for β since the weight $\hat{\lambda}_b$ (or equivalently $\hat{\lambda}_\mu$) is around 36% while the weight $\hat{\lambda}_\beta$ is around 6%. For the 3S robust Bayesian estimator, to the contrary, the base prior for the specific effects b (or equivalently μ) is inconsistent with the data as the ML-II posterior density of b is close to the posterior $\hat{q}^*(b \mid h_0)$ and to the empirical Bayes estimator $\hat{b}_{EB}(\beta \mid h_0)$.

Using either the individual block resampling bootstrap or the multivariate t-distributions yields essentially the same standard errors for both the 2S and 3S models.[17] The table also reports the numerical standard error (nse), often referred to as the Monte Carlo error, for the Full Bayesian estimator and the 3S with multivariate t-distributions. The nse is equal to the difference between the mean of the sampled values and the true posterior mean. As a rule of thumb, as many simulations as necessary should be conducted to ensure that the Monte Carlo error of each parameter of interest is less than approximately 10% of the sample standard error (see Brooks and Gelman (1998); Gilks et al. (1997); Koop (2003)). As shown in the table, the estimated nse for the 2S and 3S with multivariate t-distributions and the Full Bayesian estimators easily satisfy this criterion. Moreover, Table A1 in the supplementary appendix underlines the very good behavior of the 3S *bootstrap* compared to FGLS for several values of N and T, and for $\rho = 0.3$ and $\rho = 0.8$.

5.2.2. The Mundlak-type fixed effects world

In the fixed effects world, we allow the individual effects μ and the covariates X to be correlated. This is usually accounted for through a Mundlak-type (see Mundlak (1978)) or a Chamberlain-type specification (see Chamberlain (1982)). For the Mundlak-type specification, the individual effects are defined as: $\mu = (Z'_\mu X/T)\pi + \varpi$, $\varpi \sim N(0, \sigma^2_\varpi I_N)$ where π is a $(K_1 \times 1)$ vector of parameters to be estimated. The model can be rewritten as $y = X\beta + PX\pi + Z_\mu \varpi + u$, where $P = \left(I_N \otimes \frac{J_T}{T}\right)$ is the between-transformation (see Baltagi (2013)). We can concatenate $[X, PX]$ into a single matrix of observables and let $Wb \equiv Z_\mu \varpi$.

For the Mundlak world, we compare the standard FGLS estimator on the transformed model to our robust 2S, 3S and FB estimators of that same specification. In our case, as $\mu_i = \overline{x}_{2,i}\pi + \nu_i$, the transformed model is given by: $y = x_{1,1}\beta_{1,1} + x_{1,2}\beta_{1,2} + x_2\beta_2 + Px_2\pi + Z_\mu\nu + u$. In this specification, $X = [x_{1,1}, x_{1,2}, x_2, Px_2]$, $W = Z_\mu$ and $b = \nu$.

The simulation results are reported in Table 2. As with the fixed effects model, all the estimators yield essentially the same parameter estimates, standard errors and RMSEs, except for the *2S t-dist* estimator. Indeed, relative to the *3S t-dist* estimator, the former overestimates the standard errors of the slope parameters as well as the variance of the individual effects, σ^2_μ. This result is also found when investigating the Chamberlain-type fixed effects world and the Hausman-Taylor world (see Tables 3 and 4).

Finally, note that the *nse* statistics underlines the accuracy of the posterior estimates of the *3S t-dist* and *FB* estimators. Furthermore, increasing the size of N and/or T ($N = 500$, $T = 5$) or ($N = 100$, $T = 10$) leads to very stable results (see Table A2 in the supplementary appendix).

---

[17]Recall that we use only $BR = 20$ individual block bootstrap resamples. 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 the same results.

5.2.3. The Chamberlain-type fixed effects world

For the Chamberlain-type specification, the individual effects are given by $\mu = \underline{X}\Pi + \varpi$, where $\underline{X}$ is a $(N \times TK_1)$ matrix with $\underline{X}_i = (X'_{i1}, ..., X'_{iT})$ and $\Pi = (\pi'_1, ..., \pi'_T)'$ is a $(TK_1 \times 1)$ vector. Here π_t is a $(K_1 \times 1)$ vector of parameters to be estimated. The model can be rewritten as: $y = X\beta + Z_\mu \underline{X}\Pi + Z_\mu \varpi + u$. We can concatenate $[X, Z_\mu \underline{X}]$ into a single matrix of observables and let $Wb \equiv Z_\mu \varpi$.

For the Chamberlain world, we compare the Minimum Chi-Square (MCS) estimator (see Chamberlain (1982); Hsiao (2003); Baltagi et al. (2009)) to our estimators. These are based on the transformed model: $y_{it} = x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + \sum_{t=1}^{T} x_{2,it}\pi_t + \nu_i + u_{it}$ or $y = x_{1,1}\beta_{1,1} + x_{1,2}\beta_{1,2} + x_2\beta_2 + \underline{x_2}\Pi + Z_\mu\nu + u$. In that specification, $X = [x_{1,1}, x_{1,2}, x_2, \underline{x_2}]$, $W = Z_\mu$ and $b = \nu$.

Table 3 once again shows that the results of the *3S bootstrap*, the *3S t-dist* and the *FB* are very close to those of the classical MCS estimator. This is true of the slope coefficients β_{11}, β_{12}, β_2, the π_t coefficients for $t = 1, ..., 5$ (not reported), their standard errors and their RMSEs', as well as of the two variances σ_ϵ^2 and σ_μ^2. Increasing the size of N and/or T ($N = 500$, $T = 5$) or ($N = 100$, $T = 10$) does not change the results qualitatively, if anything it improves marginally the 3S estimator with individual block resampling bootstrap (see Table A2 in the supplementary appendix).[18]

5.2.4. The Hausman-Taylor world

The Hausman-Taylor model (henceforth HT, see Hausman and Taylor (1981)) posits that $y = X\beta + Z\eta + Z_\mu\mu + u$, where Z is a vector of time-invariant variables, and that subsets of X (*e.g.*, $X'_{2,i}$) and Z (*e.g.*, Z'_{2i}) may be correlated with the individual effects μ, but leave the correlations unspecified. Hausman and Taylor (1981) proposed a two-step IV estimator. For our general model (2): $y = X\beta + Wb + u$, we assume that ($\overline{X'_{2,i}}$, Z'_{2i} and μ_i) are jointly normally distributed:

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

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

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

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

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

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

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}} \right) \theta_X + f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}} \right) \odot \left(Z'_{2i} - E_{Z'_2} \right) \right] \theta_Z + \varpi_i, \tag{26}$$

---

[18]We only report the results for MCS and the 3S with individual block resampling bootstrap for the sake of brevity.

where $\odot$ is the Hadamard product and $f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(Z'_{2i} - E_{Z'_2}\right)\right]$ can be a nonlinear function of $\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(Z'_{2i} - E_{Z'_2}\right)$. The first term on the right-hand side of equation (26) corresponds to the Mundlak transformation while the middle term captures the correlation between Z'_{2i} and μ_i. The individual effects, μ, are a function of PX and $(f[PX \odot Z])$, *i.e.*, a function of the column-by-column Hadamard product of PX and Z. We can once again concatenate $[X, PX, f[PX \odot Z]]$ into a single matrix of observables and let $Wb \equiv Z_\mu \varpi$.

For our model, $y_{it} = x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + Z_{1,i}\eta_1 + Z_{2,i}\eta_2 + \mu_i + u_{it}$ or $y = X_1\beta_1 + x_2\beta_2 + Z_1\eta_1 + Z_2\eta_2 + Z_\mu\mu + u$. Then, we assume that

$$\mu_i = (\overline{x_{2,i}} - E_{\overline{x_2}})\theta_X + f[(\overline{x_{2,i}} - E_{\overline{x_2}}) \odot (Z_{2i} - E_{Z_2})]\theta_Z + \nu_i. \tag{27}$$

We propose adopting the following strategy: If the correlation between μ_i and Z_{2i} is quite large (> 0.2), use $f[.] = (\overline{x_{2,i}} - E_{\overline{x_2}})^2 \odot (Z_{2i} - E_{Z_2})^s$ 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 Z_{2i} *a priori*. We can use a proxy of μ_i defined by the OLS estimation of μ: $\widehat{\mu} = (Z'_\mu Z_\mu)^{-1} Z'_\mu \widehat{y}$ where $\widehat{y}$ are the fitted values of the pooling regression $y = X_1\beta_1 + x_2\beta_2 + Z_1\eta_1 + Z_2\eta_2 + \zeta$. Then, we compute the correlation between $\widehat{\mu}$ and Z_2. In our simulation study, it turns out the correlations between μ and Z_2 are large: 0.97 and 0.70 when $\rho = 0.8$, and $\rho = 0.3$, respectively. Hence, we choose $s = 1$. In this specification, $X = [x_{1,1}, x_{1,2}, x_2, Z_1, Z_2, Px_2, f[Px_2 \odot Z_2]]$, $W = Z_\mu$ and $b = \nu$.

Table 4 compares results of the classical IV estimator to those of the robust hierarchical (2S and 3S) and Full Bayesian estimators. Once again, the results are very close to one another. This is true for the slope coefficients β_{11}, β_{12} of the exogenous time-varying variables $x_{1,1,it}$, $x_{i,1,2,it}$, for the slope coefficients β_2 of the endogenous time-varying variable $x_{i,2,it}$, for the coefficient η_1 of the time-invariant variable $Z_{1,i}$ (uncorrelated with μ_i) and for the remainder variance σ_ϵ^2. On the other hand, the 3S *bootstrap*, the *3S t-dist* and the *FB* estimators yield slightly upward biased estimates of η_2, the coefficient associated with the time-invariant variable $Z_{2,i}$ which is itself correlated with μ_i. While the biases are relatively small (4.53%, 4.54% and 5.99%, respectively), they are insensitive to the sample size, N, but tend to taper off with T (see Table A3 in the supplementary appendix).[19] Interestingly, the standard errors of that same coefficient are considerably smaller when using the Bayesian estimators, irrespective of N and T. Consequently, the RMSEs of the Bayesian estimators are smaller and the 95% confidence intervals are narrower and entirely nested within those obtained with the IV procedure of Hausman-Taylor. The slight bias is thus entirely offset by increased precision.

Finally, note that both the IV and the Bayesian estimators yield biased estimates of σ_μ^2. While the former overestimates the true value and the latter does the opposite, only the IV estimator behaves better when N and/or T increase.

It must be noted that the full Bayesian (FB) estimator sometimes yield smaller RMSEs for β_{11}, β_{12} and β_2 than the classical (FGLS, Mundlak, MCS, IV-HT) estimators and the 3S estimators in the Chamberlain world and more specifically in the Hausman-Taylor world (see Tables 3 and 4). Since $N(= 100)$ is relatively small, the estimated standard errors are likely underestimated. As N gets larger (e.g., $N = 500$, see Table A4 in the supplementary appendix), the coefficients, SEs and RMSEs of the FB become similar to those of the MCS, IV and 3S *bootstrap* estimators. Note

---

[19]Table A3 only reports the results for the IV procedure and the 3S with individual block resampling bootstrap to save space.

that the computational burden of the full Bayesian method is measured in hours and even days compared to the ε-contamination approach.[20]

5.2.5. Sensitivity to ε-contamination values

Table 5 investigates the sensitivity of the Hausman-Taylor estimator [21] with respect to ε, the contamination part of prior distributions, which varies between 0 and 90%. As shown, most parameter estimates are relatively insensitive to ε. The only noteworthy change concerns the estimated value of λ_μ in the 2S *bootstrap* estimator. It more or less corresponds to $(1 - \varepsilon)$. This particular relation may occur whenever $\widehat{h}/\left(\widehat{h} + 1\right) = h_0/\left(h_0 + 1\right)$ 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 3.2). For the 3S *bootstrap*, the estimated standard errors of β_2 are much larger than those obtained from usual HT method. But for larger N or larger T, this does not occur (see Table A3 and A4 in the supplementary appendix). The observed stability of the coefficients estimates stems from the fact that the base prior is not consistent with the data as the weight $\widehat{\lambda}_\beta \to 0$. The ML-II posterior mean of β is thus close to the posterior $\widehat{q}^*\left(\beta \mid g_0\right)$ and to the empirical Bayes estimator $\widehat{\beta}_{EB}\left(\mu \mid g_0\right)$. Hence, the numerical value of the ε-contamination, for $\varepsilon \neq 0$, does not seem to play an important role in our simulated worlds.

As suggested by a referee, we conducted further simulations with $\varepsilon = 0$. Table 5 reports the results for the Hausman-Taylor model.[22] Under the null, $H_0 : \varepsilon = 0$, it follows that the weights $\widehat{\lambda}_{\beta,g_0} = 1$ and $\widehat{\lambda}_{b,h_0} = 1$ so that the restricted ML-II estimator of β is given by $\widehat{\beta}_{rest} = \beta_*(b \mid g_0)$. Under $H_1 : \varepsilon \neq 0$ the unrestricted estimator is $\widehat{\beta}_{un}\left(\equiv \widehat{\beta}_{ML-II}\right) = \widehat{\lambda}_{\beta,g_0}\beta_*(b \mid g_0) + \left(1 - \widehat{\lambda}_{\beta,g_0}\right)\beta_{EB}(b \mid g_0)$. The restricted ML-II estimator $\beta_*(b \mid g_0)$ is the Bayes estimator under the base prior g_0. It differs from the full Bayesian estimator (FB) (described in the supplementary appendix) in that the latter is based on other priors. The use of Zellner's g-priors leads to more diffuse priors than those of the FB. In other words, the FB is likely more sensitive to the choice of priors than our ML-II estimator even when $\varepsilon = 0$. The full Bayesian estimator (FB) does not guarantee the posterior robustness of our ML-II estimator. Additionally, 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. Applying Theorem 1 of Magnus and Durbin (1999), we can derive the conditions under which $MSE\left(\widehat{\beta}_{rest}\right) - MSE\left(\widehat{\beta}_{un}\right)$ is positive or negative semidefinite (see the supplementary appendix).[23] As shown in Table 5 for $\varepsilon = 0$,

---

[20]Recall that in our simulation study, we only use $1,000$ draws and 500 burn-in draws for the full Bayesian estimator (with $N = 500$, $T = 5$ and $1,000$ replications). Had we used more draws (as is typically done in MCMC analyses), for instance $10,000$ draws and $5,000$ burn-in draws, then the computing time would have been approximately 4 days. The computation time of our 3S *bootstrap* estimators would have taken only 3 hours.

[21]This exercise could be conducted for the other worlds (RE, FE) but we only report the results for the Hausman-Taylor world for the sake of brevity.

[22]Note that when $\varepsilon = 0$, the unconditional predictive density corresponding to the base prior, $m\left(\widetilde{y} \mid \pi_0, \beta\right)$, depends on a Gaussian hypergeometric function which can not be numerically evaluated with the Laplace approximation nor with adaptive quadrature methods if Zellner's g-priors are too large (*e.g.*, $g_0 > 1$).

[23]The difference of MSE is given by:

$$MSE\left(\widehat{\beta}_{rest}\right) - MSE\left(\widehat{\beta}_{un}\right) = QQ' - D.$$

Hence

$$MSE\left(\widehat{\beta}_{rest}\right) \leq (\text{resp.}\geq) \; MSE\left(\widehat{\beta}_{un}\right) \; \text{iff} \; Q'D^-Q \leq (\text{resp.}\geq) \; 1,$$

the β's (of the 2S and 3S *bootstrap*) are slightly downward-biased, and even more so for the 2S *bootstrap* η's and the 3S *bootstrap* η_1. Moreover, the estimated variance of the specific effects, σ_μ^2, are downward biased in both the 2S *bootstrap* ($\sigma_\mu^2 = 3.53$) and the 3S *bootstrap* ($\sigma_\mu^2 = 3.38$) estimators. Irrespective of the value of ε, the 3S *bootstrap* standard errors of β_{11} and β_{12} are close to those of the Hausman-Taylor IV estimator. However, when $\varepsilon = 0$ the 3S *bootstrap* standard errors of η_2 is clearly downward-biased compared to those of the 2S and the 3S *bootstrap* estimators when $\varepsilon \neq 0$. This also holds true for the RMSE of this coefficient.

The restricted ML-II estimator ($\varepsilon = 0$) depends on the value of Zellner's g-priors, g_0 (and h_0 for b). Whatever the value of ε, we have chosen g_0 and h_0 so that the prior is as diffuse as possible. A good strategy is to follow the unit information prior (UIP) principle suggested by Kass and Wasserman (1995) which corresponds to $g_0 = h_0 = 1/NT$. This leads to Bayes factors that behave like the Bayesian Information Criterion (BIC). In that case, the difference in the MSEs of the parameter estimates computed with $\varepsilon \in\,]0,1[$ and those with $\varepsilon = 0$ could be small since in both cases the elicited prior is diffuse. From Table 5, and Tables A4 and A5 in the supplementary appendix, the results differ from the full Bayesian method. In Table A5, we investigate the sensitivity of the 3S *bootstrap* estimator when both ε and Zellner's g-priors change. When $\varepsilon = 0.5$, increasing Zellner's g-priors from $g_0 = h_0 = 1/NT$ to $g_0 = h_0 = 0.1$ has little impact on the results. On the other hand, when $\varepsilon = 0$ (with $g_0 = h_0 = 1/NT$), the β's and η_1 are biased and the estimated variance of the specific effects, σ_μ^2, is downward-biased ($\sigma_\mu^2 = 3.38$) as indicated previously. Increasing Zellner's g-priors ($g_0 = h_0 = 0.1$) when $\varepsilon = 0$, worsens the biases on all the parameter estimates, and especially for the two η's and σ_μ^2 (3.30 instead of 4). Moreover, the RMSEs of all the coefficients increase significantly, thus emphasizing the differences in the MSEs when we move from $\varepsilon = 0$ to $\varepsilon \neq 0$.

Table 5 reports the results for $\varepsilon = 0.1, 0.5, 0.9$. The RMSEs are almost the same across the 3S *bootstrap* specifications. Recall that this estimator is data driven and implicitly adjusts the weights to the different values of ε-contamination. This may by why, even though the choice of $\varepsilon = 0.5$ is somewhat arbitrary, the adjustment compensates for it not being optimal (see Berger (1985)). A possible guide for the choice of parameters of the 3S estimator is as follows:

- First, choose g_0 and h_0 so that the prior is as diffuse as possible. A good approach is to follow the unit information prior (UIP) principle suggested by Kass and Wasserman (1995) which corresponds to $g_0 = h_0 = 1/NT$.

- Second, choose any ε value lying in $\varepsilon \in\,]0,1[$.

---

with

$$
\begin{aligned}
\text{with } Q \;&=\; \left(1 - \lambda_{\beta,g_0}\right)\left(\beta_*(b \mid g_0) - \beta_{EB}(b \mid g_0)\right) \\
\text{and } D \;&=\; \left(1 - \lambda_{\beta,g_0}\right)\left[Var\left[q^*(\beta \mid g_0)\right] - Var\left[\pi_0^*(\beta \mid g_0)\right]\right] \\
&\quad + \lambda_{\beta,g_0}\left(1 - \lambda_{\beta,g_0}\right)\left(\beta_*(b \mid g_0) - \beta_{EB}(b \mid g_0)\right)\left(\beta_*(b \mid g_0) - \beta_{EB}(b \mid g_0)\right)',
\end{aligned}
$$

where D^- denotes the generalized inverse of D. If $MSE\left(\widehat{\beta}_{rest}\right)$ and $MSE\left(\widehat{\beta}_{un}\right)$ are two positive semidefinite matrices, the notation $MSE\left(\widehat{\beta}_{rest}\right) \leq MSE\left(\widehat{\beta}_{un}\right)$ means that $MSE\left(\widehat{\beta}_{un}\right) - MSE\left(\widehat{\beta}_{rest}\right)$ is positive semidefinite. In contrast, the difference of MSE between our ML-II estimator and the FB cannot be defined analytically but only be numerically evaluated with our simulation studies.

21

5.2.6. Departure from normality

Table 6 investigates the robustness of the estimators to a non-normal framework. The remainder disturbances, u_{it}, are now assumed to follow a right-skewed t-distribution instead of the usual normal distribution. The parameters of the right-skewed t-distribution are mean $= 0$, degrees of freedom $= 3$, and shape $= 2$ (see Fernández and Steel (1998)). Our *3S bootstrap* estimator yields similar slope parameter estimates to those of the classical estimators, *i.e.* Mundlak FGLS, the MCS or the IV. The main difference concerns the estimates of the remainder variance σ_u^2 and the individual effects variance σ_μ^2. Indeed, in the Mundlak-type fixed effects world, the estimate of σ_u^2 computed from the 3S *bootstrap* is closer to the theoretical value than that of the FGLS estimate. In the Chamberlain-type fixed effects world, the estimate of σ_μ^2 (resp. σ_u^2) computed from the 3S *bootstrap* estimator is closer (resp. is further) to the theoretical value than that of the standard MCS estimator. Finally, the 3S *bootstrap* estimates of σ_μ^2 and σ_u^2 in the Hausman-Taylor world are both closer to their true values than those of the HT estimator.

Yet, the more interesting result concerns the precision and the bias of the estimate of η_2 in the Hausman-Taylor world. Notice first that the standard error of the 3S *bootstrap* estimate is once again much smaller than what obtains using the standard HT estimator (0.2039 *versus* 0.3478). Consequently, the 95% confidence interval of the 3S *bootstrap* estimator is much narrower and entirely nested within the one obtained with the IV procedure of Hausman-Taylor ($[0.6189; 1.418]$ *versus* $[0.3180; 1.6814]$). Second, it is worth emphasizing that in this non-Gaussian framework, the bias of η_2 has decreased significantly (1.88% *versus* 4.53%) and that the estimate of λ_β, while still relatively small, is now more important.

Following the suggestions of the referees, we further investigated the consequences of relaxing the normality of the remainder terms. We first simulated the model using a t-distribution that exhibited more right-skewness, *i.e.* with mean zero, five degrees of freedom and shape parameter of 3. The simulation results show that increasing the skewness amplifies the gains of our ε-contamination methodology over the standard Hausman and Taylor (1981) estimator (see Table A6 of the supplementary appendix). Second, we assumed instead that the remainder terms followed a χ^2 distribution with 2 degrees of freedom. The results in Table A7 of the supplementary appendix clearly indicate that our approach is preferable to the Hausman-Taylor estimator.

5.2.7. Sensitivity to endogenous covariates

As shown above, our robust 3S *bootstrap* estimator exhibits interesting properties in Gaussian and non-Gaussian frameworks. As a final check on its properties, it is worth investigating its relative behavior in the context of endogenous covariates.[24] We focus on the Hausman-Taylor and the Chamberlain worlds with $N = 100$ and $T = 5$. For the Hausman-Taylor world, we assume that

$$x_{2,it} = 0.7 x_{2,it-1} + \mu_i + \vartheta_{it} \text{ with } \vartheta_{it} \equiv u_{it} \text{ and } u_{it} \sim N(0, 8),$$

as opposed to assuming $\vartheta_{it} \sim U(-2, 2)$ above. In addition to being correlated with μ_i and $Z_{2,i}$, $x_{2,it}$ is now an endogenous covariate since it is correlated with y_{it} through u_{it}. We have chosen a large variance for u_{it} so as to amplify the potential impact of the endogeneity problem on the results.

For the Chamberlain world, we assume that

$$x_{2,it} = \rho_{x_2} x_{2,it-1} + u_{it} \text{ , } u_{it} \sim N(0, \sigma_u^2), \sigma_u^2 = 8 \text{ and } \rho_{x_2} = 0.7.$$

---

[24]We are grateful to Richard Blundell for this suggestion and for helpful comments.

The covariance between $x_{2,it}$ and u_{it} is given by $E\left[x_{2,it}u_{it}\right] \simeq \sigma_u^2 \sum_{j=0}^{t-1} \rho_{x_2}^{2j} = \sigma_u^2 \left(1 - \rho_{x_2}^{2t}\right) / \left(1 - \rho_{x_2}^2\right)$. It ranges from σ_u^2 to approximately $\sigma_u^2 / \left(1 - \rho_{x_2}^2\right) \simeq 1.96\sigma_u^2$ as t goes from 1 to T. The specific effect $\mu_i = \sum_{t=1}^T x_{2,i,t}\pi_t + \nu_t \simeq \sum_{t=1}^T \left(\sum_{j=0}^{t-1} \rho_{x_2}^j u_{it-j}\right)\pi_t + \nu_t$ is also correlated with u_{it} and the covariance is given by $E\left[\mu_i u_{it}\right] = \sigma_u^2 \sum_{t=1}^T \pi_t = \sigma_u^2 \sum_{t=1}^T (0.8)^{T-t} = 5\sigma_u^2 \left(1 - 0.8^T\right)$.

Table 7 presents the results for the Chamberlain world in the upper panel and those of the Hausman-Taylor world in the bottom one. As for the Chamberlain world, the endogeneity has little if any impact on both the MCS and the robust Bayesian estimators. The parameter estimates are very close to those found in Table 3. The correlation between $x_{2,it}$ and u_{it} increases the standard error (and the RMSE) of $x_{2,it}$ but less so for the other parameter estimates. Note that the Bayesian estimates of σ_μ^2 are closer to the true value than what obtains from the MCS estimator.

In the Hausman-Taylor world, as opposed to the Chamberlain world, the classical IV estimator is strongly affected by the endogeneity problem. Oddly, the estimate of the coefficient associated with $x_{2,it}$ and its standard error are little impacted. Rather, the estimates of the intercept and of Z_{2i} lose all statistical significance. Likewise, the estimated variance of the specific effects σ_μ^2 is strongly upward biased ($+34.67\%$) while the estimated variance of the remainder terms σ_u^2 is unbiased. Lastly, the standard errors of the exogenous variables $x_{1,1,it}$ and $x_{1,2,it}$ are now twice as large as those found in Table 4 but nevertheless lead to coefficients which are still statistically significant.

The robust estimators (3S *bootstrap* and 3S *t-dist*) and the full Bayesian estimator (FB) behave better, especially in terms of precision. The standard errors of the intercept and the coefficient of Z_{2i} are smaller and lead to coefficients that are statistically significant.[25] The slope parameters are all slightly biased (except β_2), and in particular the one associated with Z_{2i}. Nevertheless, the null hypothesis that $\beta_{1,1} = \beta_{1,2} = \beta_2 = \eta_1 = \eta_2 = 1$ cannot be rejected. Furthermore, as the standard errors of these coefficients are relatively small, the RMSEs are quite acceptable and often smaller than those of the usual IV estimator. In fact, the IV estimator only outperforms the Bayesian estimators in that it provides a slightly less biased estimate of the variance of the specific effects, σ_μ^2. The relative (negative) biases of the 3S *bootstrap*, 3S *t-dist* and FB amount to 46.34%, 48.11%, 47.04%, respectively. This is somewhat more than the (positive) bias of the Hausman-Taylor IV estimator which amounts to 35%. But more interestingly, the intercept η_1 and the slope parameter η_2 of the time-invariant variable Z_{2i} are strongly significant as compared to the classical IV estimator. Our 3S robust Bayesian estimator is thus more robust to endogeneity problems than the Hausman-Taylor estimator.

One may wonder why Chamberlain's MCS estimator is more robust to an endogeneity problem than the Hausman-Taylor IV estimator. Recall that the MCS, sometimes called the optimal minimum distance estimator, is qualitatively similar to the GMM estimator and whose properties lead to more efficient estimates than the 2SLS instrumental variable of the Hausman-Taylor estimator. Notwithstanding this, our robust 3S estimators are well-behaved in the face of endogeneity problems as evidenced above.

---

[25]The values of the coefficients c and d of the generalized hyper-g priors appear to play an important role in the precision of the intercept. Recall that we have chosen $c = d = 1$ for the Beta-prime distribution to increase the precision. With $c = 0.1$ and $d = 1$, the intercept is not significantly different from zero.

6. Conclusion

To our knowledge, our paper is the first to analyze the static linear panel data model using an ε-contamination approach within a two-stage and three-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 such as random effects, Mundlak-type and Chamberlain-type fixed effects and Hausman-Taylor models. The frequentist approach, on the other hand, requires separate estimators for each model, respectively FGLS, MCS and IV.

Following Chaturvedi and Singh (2012), we estimate the Type-II maximum likelihood (ML-II) posterior distribution of the slope coefficients and the individual effects using a two-step procedure. If the base prior is consistent with the data, the ML-II posterior density gives more weight to the conditional posterior density derived from the elicited prior. Conversely, if the base prior is not consistent with the data, the ML-II posterior density is closer to the conditional posterior density derived from the ε-contaminated prior.

The finite sample performance of the two-stage and three-stage hierarchical models are investigated using extensive Monte Carlo experiments. The experimental design includes a random effects world, a Mundlak-type fixed effects world, a Chamberlain-type fixed effects world and a Hausman-Taylor-type world. The simulation results underscore the relatively superior performance of the three-stage hierarchy estimator, irrespective of the data generating process considered. The biases and the RMSEs are close and often smaller than those of the conventional (classical) estimators. In the two-stage hierarchy, their exists a trade-off between the Bayes estimators and the empirical Bayes estimators. In the three-stage hierarchy, this trade-off vanishes and only the empirical Bayes estimator matters in the estimation of the coefficients and the individual effects.

We also investigate the sensitivity of the estimators to the contamination part of the prior distribution. It turns out most parameter estimates are relatively stable. Next we investigate the robustness of the estimators when the remainder disturbances are assumed to follow either a right-skewed t-distribution or a χ^2 distribution. Our robust estimators globally outperform the classical estimators both in terms of precision and bias. Finally, we investigate the consequences of introducing an endogenous covariate. Our results show that both the MCS and our robust 3S estimators perform well. On the other hand, our 3S estimator clearly outperforms the Hausman-Taylor IV estimator both in terms of precision and bias.

The robust Bayesian approach we propose is arguably a relevant all-in-one panel data framework. In future work we intend to broaden its scope by addressing issues such as heteroskedasticity, autocorrelation of residuals, general IV, dynamic and spatial models.

References

Abramovitz, M., Stegun, I.A., 1970. Handbook of Mathematical Functions. Dover Publications, Inc. New York.

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

Appell, P., 1882. Sur les fonctions hypergéométriques de deux variables. Journal de Mathématiques Pures et Appliquées 3, 173–216.

Baltagi, B.H., 2013. Econometric Analysis of Panel Data. Fifth ed., Wiley, Chichester, UK.

Baltagi, B.H., Bresson, G., 2012. A robust Hausman-Taylor estimator, in: Baltagi, B.H., Hill, C., Newey, W., White, H. (Eds.), Advances in Econometrics: Essays in Honor of Jerry Hausman. Emerald Group Publishing Limited. volume 29, pp. 175–214.

Baltagi, B.H., Bresson, G., Pirotte, A., 2003. Fixed effects, random effects or Hausman-Taylor? a pretest estimator. Economics Letters 79, 361–369.

Baltagi, B.H., Bresson, G., Pirotte, A., 2009. Testing the fixed effects restrictions? A Monte Carlo study of chamberlain's minimum chi-square test. Statistics and Probability Letters 79, 1358–1362.

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

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

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

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

Bresson, G., Hsiao, C., 2011. A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris. Advances in Statistical Analysis 95, 501–529.

Brooks, S.P., Gelman, A., 1998. Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–455.

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

Chaturvedi, A., 1996. Robust Bayesian analysis of the linear regression. Journal of Statistical Planning and Inference 50, 175–186.

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

Chib, S., 1995. Marginal likelihood from the Gibbs sampler. Journal of the American Statistical Association 90, 1313–1321.

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.

Davis, P.J., Rabinowitz, P., 1984. Methods of Numerical Integration. 2nd ed., Academic Press, New York.

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

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

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

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.

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

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

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.

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

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.

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

Laird, N.M., Louis, T.A., 1987. Empirical Bayes confidence intervals based on bootstrap samples. Journal of the American Statistical Association, 82, 739–750.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–74.

Liang, F., Paulo, R., Molina, G., Clyde, M.A., Berger, J.O., 2008. Mixtures of g priors for Bayesian variable selection. Journal of the American Statistical Association 103, 410–423.

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

Magnus, J.R., Durbin, J., 1999. Estimation of regression coefficients of interest when other regression coefficients are of no interest. Econometrica 67, 639–643.

Maruyama, Y., George, E.I., 2011. Fully Bayes factors with a generalized g-prior. Annals of Statistics 39, 2740–2765.

Maruyama, Y., George, E.I., 2014. Posterior odds with a generalized hyper-g prior. Econometric Reviews 33, 251–269.

Moreno, E., Pericchi, L.R., 1993. Bayesian robustness for hierarchical ε-contamination models. Journal of Statistical Planning and Inference 37, 159–167.

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

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes: The Art of Scientific Computing. 3rd ed., Cambridge University Press, New York.

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

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

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

Sinha, P., Jayaraman, P., 2010a. Bayes reliability measures of lognormal and inverse Gaussian distributions under ML-II ε-contaminated class of prior distributions. Defence Science Journal 60, 442–450.

Sinha, P., Jayaraman, P., 2010b. Robustness of Bayes decisions for normal and lognormal distributions under hierarchical priors. MPRA Working Paper No. 22416. Faculty of Management Studies, University of Delhi.

Slater, L.J., 1966. Generalized Hypergeometric Functions. Cambridge University Press, Cambridge, UK.

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

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

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.

Table 1: Random Effects World (FGLS, Robust 2S, 3S and Full Bayesian)

$N = 100$, $T = 5$, $\rho = 0.8$, $\varepsilon = 0.5$, Replications$=1,000$

	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_β	λ_μ	Computation Time (secs.)
True	1	1	1	1	4			
FGLS coef	0.9986	0.9992	1.0005	1.0023	3.9701			47.703
se	0.0337	0.0334	0.0341					
rmse	0.0340	0.0342	0.0333					
2S bootstrap coef	0.9977	0.9983	0.9997	0.9975	3.9405	0.0643	0.3574	351.275
se	0.0336	0.0333	0.0344					
rmse	0.0352	0.0351	0.0350					
3S bootstrap coef	0.9972	0.9982	0.9994	0.9974	3.9367	0.0660	$< 10^{-4}$	558.113
se	0.0327	0.0327	0.0337					
rmse	0.0346	0.0347	0.0341					
2S t-dist. coef	0.9915	0.9990	1.0010	0.9962	4.0150	0.0580	0.3708	1,980.579
se	0.0410	0.0413	0.0402					
rmse	0.0418	0.0413	0.0402					
nse	0.0015	0.0015	0.0015					
3S t-dist. coef	0.9918	0.9988	1.0009	0.9959	3.9902	0.0480	$< 10^{-4}$	2,640.001
se	0.0373	0.0375	0.0377					
rmse	0.0382	0.0375	0.0376					
nse	0.0016	0.0016	0.0017					
FB coef	0.9987	0.9992	1.0004	0.9986	3.9101			7,778.272
se	0.0340	0.0342	0.0333					
rmse	0.0341	0.0342	0.0333					
nse	0.0017	0.0016	0.0017					

2S bootstrap (3S bootstrap) : 2S (or 3S) with individual block resampling bootstrap.

2S (3S) t-dist: 2S (3S) with multivariate t-distributions. FB: Full Bayesian.

Table 2: Mundlak-type Fixed Effects World (FGLS, Robust 2S, 3S and Full Bayesian)

$N = 100$, $T = 5$, $\rho = 0.8$, $\varepsilon = 0.5$, replications=1,000

	β_{11}	β_{12}	β_2	π	σ_u^2	σ_μ^2	λ_β	λ_μ	Computation Time (secs.)
True	1	1	1	0.8	1	6.3334			
Mundlak coef	1.0014	1.0010	1.0012	0.7992	0.9981	6.3431			51.184
se	0.0231	0.0232	0.0354	0.0473					
rmse	0.0236	0.0242	0.0348	0.0467					
2S bootstrap coef	1.0000	0.9995	1.0010	0.8019	0.9928	6.3905	0.0008	0.4136	369.346
se	0.0289	0.0293	0.0344	0.0482					
rmse	0.0310	0.0311	0.0353	0.0487					
3S bootstrap coef	1.0002	1.0005	1.0010	0.8017	0.9918	6.3690	0.0006	$< 10^{-4}$	576.094
se	0.0261	0.0263	0.0345	0.0463					
rmse	0.0271	0.0289	0.0356	0.0473					
2S t-dist coef	0.9953	1.0003	1.0011	0.8376	0.9917	7.7819	0.0090	0.4050	2,040.113
se	0.0394	0.0404	0.0348	0.2083					
rmse	0.0396	0.0404	0.0348	0.2116					
nse	0.0014	0.0014	0.0015	0.0042					
3S t-dist coef	0.9967	1.0004	1.0011	0.8030	0.9918	6.4045	0.0060	$< 10^{-4}$	2,796.030
se	0.0304	0.0316	0.0349	0.0492					
rmse	0.0306	0.0316	0.0349	0.0493					
nse	0.0015	0.0015	0.0015	0.0030					
FB coef	1.0014	1.0011	1.0012	0.7992	0.9938	6.3182			8,416.163
se	0.0237	0.0245	0.0348	0.0467					
rmse	0.0237	0.0245	0.0348	0.0467					
nse	0.0012	0.0013	0.0017	0.0025					

2S bootstrap (3S bootstrap) : 2S (or 3S) with individual block resampling bootstrap.

2S (3S) t-dist: 2S (3S) with multivariate t-distributions. FB: Full Bayesian.

Table 3: Chamberlain-type Fixed Effects World (MCS, Robust 2S, 3S and Full Bayesian)

$N = 100$, $T = 5$, $\rho = 0.8$, $\varepsilon = 0.5$, replications=1,000

	β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_β	λ_μ	Computation Time (secs.)
True	1	1	1	1	96.2228			
MCS coef	1.0025	0.9989	1.0018	1.0038	94.6900			17.007
se	0.0255	0.0256	0.0346					
rmse	0.0286	0.0267	0.0372					
2S bootstrap coef	0.9934	0.9961	1.0009	0.9935	95.6672	$< 10^{-4}$	0.4136	343.361
se	0.0301	0.0304	0.0345					
rmse	0.0337	0.0299	0.0368					
3S bootstrap coef	0.9946	0.9959	1.0015	0.9955	95.2535	$< 10^{-4}$	$< 10^{-4}$	531.174
se	0.0267	0.0268	0.0345					
rmse	0.0296	0.0276	0.0367					
2S t-dist coef	0.9832	0.9935	1.0011	0.9941	108.1559	$< 10^{-4}$	0.4628	2,144.238
se	0.0452	0.0426	0.0362					
rmse	0.0482	0.0430	0.0362					
nse	0.0015	0.0015	0.0015					
3S t-dist coef	0.9891	0.9955	1.0011	0.9947	96.6751	$< 10^{-4}$	$< 10^{-4}$	2,794.758
se	0.0322	0.0300	0.0361					
rmse	0.0340	0.0304	0.0361					
nse	0.0015	0.0015	0.0016					
FB coef	1.0007	0.9996	1.0015	0.9970	96.1373			8,797.542
se	0.0251	0.0230	0.0362					
rmse	0.0251	0.0230	0.0362					
nse	0.0013	0.0013	0.0019					

2S bootstrap (3S bootstrap) : 2S (or 3S) with individual block resampling bootstrap.

2S (3S) t-dist: 2S (3S) with multivariate t-distributions. FB: Full Bayesian. The parameters π_t are omitted from the table.

Table 4: Hausman-Taylor World (IV, Robust 2S, 3S and Full Bayesian)

$N = 100$, $T = 5$, $\rho = 0.8$, $\varepsilon = 0.5$, replications=1,000

		β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ	Computation Time (secs.)
True		1	1	1	1	1	1	4			
HT	coef	0.9976	0.9981	1.0037	1.0173	0.9958	0.9924	4.2904			3.637
	se	0.0395	0.0394	0.0444	0.2143	0.1857					
	rmse	0.0405	0.0385	0.0456	0.2071	0.1903					
2S bootstrap	coef	0.9965	0.9983	1.0002	1.0031	1.0320	0.9935	3.8244	$< 10^{-4}$	0.5041	335.815
	se	0.0529	0.0506	0.0624	0.2056	0.1029					
	rmse	0.0530	0.0506	0.0624	0.2057	0.1078					
3S bootstrap	coef	0.9905	0.9921	1.0002	1.0136	1.0453	0.9947	3.7420	$< 10^{-4}$	$< 10^{-4}$	635.644
	se	0.0375	0.0356	0.0628	0.2051	0.0799					
	rmse	0.0387	0.0365	0.0628	0.2055	0.0919					
2S t-dist	coef	0.9938	0.9997	1.0003	1.0471	0.9582	0.9925	7.4181	$< 10^{-4}$	0.3320	2,507.205
	se	0.0459	0.0420	0.0450	1.1895	0.3458					
	rmse	0.0463	0.0419	0.0450	1.1899	0.3481					
	nse	0.0016	0.0016	0.0019	0.0209	0.0095					
3S t-dist	coef	0.9886	0.9920	1.0004	1.0172	1.0454	0.9937	3.7171	$< 10^{-4}$	$< 10^{-4}$	3,297.587
	se	0.0288	0.0268	0.0449	0.1966	0.0615					
	rmse	0.0310	0.0279	0.0449	0.1972	0.0764					
	nse	0.0014	0.0014	0.0020	0.0071	0.0045					
FB	coef	0.9864	0.9875	1.0004	1.0173	1.0599	0.9954	3.6260			8,564.863
	se	0.0205	0.0189	0.0450	0.1935	0.0459					
	rmse	0.0246	0.0226	0.0450	0.1941	0.0754					
	nse	0.0013	0.0013	0.0022	0.0054	0.0033					

2S bootstrap (3S bootstrap) : 2S (or 3S) with individual block resampling bootstrap.

2S (3S) t-dist: 2S (3S) with multivariate t-distributions. FB: Full Bayesian.

Table 5: Hausman-Taylor World, Robustness to ε-contamination

$N = 100$, $T = 5$, $\rho = 0.8$, replications=1,000

		β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ
	True	1	1	1	1	1	1	4		
	HT coef	0.9976	0.9981	1.0037	1.0173	0.9958	0.9924	4.2904	1	1
	se	0.0395	0.0394	0.0444	0.2143	0.1857				
	rmse	0.0405	0.0385	0.0456	0.2071	0.1903				
$\varepsilon = 0$	2S bootstrap coef	0.9835	0.9851	0.9983	0.7195	0.9423	0.9941	3.5298	1	1
	se	0.0524	0.0501	0.0623	0.3006	0.1043				
	rmse	0.0549	0.0522	0.0623	0.4111	0.1192				
	3S bootstrap coef	0.9854	0.9839	0.9977	0.9725	1.0085	0.9904	3.3834	1	1
	se	0.0355	0.0342	0.0626	0.1944	0.0761				
	rmse	0.0384	0.0378	0.0627	0.1963	0.0766				
$\varepsilon = 0.1$	2S bootstrap coef	0.9965	0.9983	1.0002	1.0045	1.0320	0.9935	3.8240	$< 10^{-4}$	0.9008
	se	0.0529	0.0506	0.0624	0.2058	0.1029				
	rmse	0.0530	0.0506	0.0624	0.2058	0.1078				
	3S bootstrap coef	0.9905	0.9919	1.0003	1.0140	1.0453	0.9956	3.7408	$< 10^{-4}$	$< 10^{-4}$
	se	0.0376	0.0358	0.0629	0.2042	0.0797				
	rmse	0.0388	0.0367	0.0629	0.2046	0.0917				
$\varepsilon = 0.5$	2S bootstrap coef	0.9965	0.9983	1.0002	1.0031	1.0320	0.9935	3.8244	$< 10^{-4}$	0.5041
	se	0.0529	0.0506	0.0624	0.2056	0.1029				
	rmse	0.0530	0.0506	0.0624	0.2057	0.1078				
	3S bootstrap coef	0.9905	0.9921	1.0002	1.0136	1.0453	0.9947	3.7420	$< 10^{-4}$	$< 10^{-4}$
	se	0.0375	0.0356	0.0628	0.2051	0.0799				
	rmse	0.0387	0.0365	0.0628	0.2055	0.0919				
$\varepsilon = 0.9$	2S bootstrap coef	0.9965	0.9983	1.0002	1.0015	1.0320	0.9935	3.8249	$< 10^{-4}$	0.1079
	se	0.0529	0.0506	0.0624	0.2058	0.1029				
	rmse	0.0530	0.0506	0.0624	0.2058	0.1078				
	3S bootstrap coef	0.9905	0.9919	1.0003	1.0140	1.0453	0.9956	3.7408	$< 10^{-4}$	$< 10^{-4}$
	se	0.0376	0.0358	0.0629	0.2042	0.0797				
	rmse	0.0388	0.0367	0.0629	0.2046	0.0917				

Table 6: Departure from Normality: The Skewed t-distribution

$N = 100$, $T = 5$, $\varepsilon = 0.5$, $\rho = 0.8$, replications=1,000

	β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ
	MUNDLAK-TYPE FIXED EFFECTS WORLD[†]								
True	1	1	1			5.0459	6.3257		
Mundlak Coef	1.0007	1.0021	1.0015			6.6390	6.3816		
se	0.0357	0.0357	0.0898						
rmse	0.0351	0.0357	0.0884						
3S bootstrap coef	0.9979	1.0036	1.0017			6.4584	6.4126	0.0363	$< 10^{-4}$
se	0.0521	0.0522	0.0867						
rmse	0.0531	0.0541	0.0901						
	CHAMBERLAIN-TYPE FIXED EFFECTS WORLD[‡]								
True	1	1	1			6.6292	95.5874		
MCS coef	1.0005	1.0000	1.0008			6.9099	91.8444		
se	0.0334	0.0333	0.0808						
rmse	0.0368	0.0374	0.0850						
3S boot strap coef	0.9982	0.9943	1.0004			7.4226	95.7630	0.0005	$< 10^{-4}$
se	0.0546	0.0588	0.0871						
rmse	0.0583	0.0672	0.0927						
	HAUSMAN-TAYLOR								
True	1	1	1	1	1	6.9095	4		
HT coef	0.9979	0.9977	1.0113	0.9901	0.9997	6.8540	5.1608		
se	0.0897	0.0902	0.1137	0.2680	0.3478				
rmse	0.0897	0.0891	0.1152	0.2529	0.3427				
3S bootstrap coef	0.9905	0.9865	1.0051	0.9479	1.0186	6.9214	4.3901	0.0041	$< 10^{-4}$
se	0.0986	0.0955	0.1607	0.3237	0.2039				
rmse	0.0990	0.0964	0.1608	0.3278	0.2047				

[†] The parameter π is omitted.

[‡] The parameters $\pi_1, \ldots, \pi_5$ are omitted.

Table 7: Sensitivity to Endogenous Covariates

$N = 100$, $T = 5$, $\varepsilon = 0.5$, replications=1,000

	β_{11}	β_{12}	β_2	η_1	η_2	σ^2_u	σ^2_μ	λ_β	λ_μ
			CHAMBERLAIN-TYPE FIXED EFFECTS WORLD[†]						
True	1	1	1	1	1	8.0000	112.0657		
MCS coef	1.0010	1.0010	0.9999			7.9944	115.3597		
se	0.0374	0.0374	0.0503						
rmse	0.0401	0.0413	0.0518						
3S bootstrap coef	1.0005	0.9996	0.9996			7.9584	111.4192	$< 10^{-4}$	$< 10^{-4}$
se	0.0403	0.0404	0.0499						
rmse	0.0410	0.0418	0.0511						
3S t-dist coef	0.9984	1.0001	0.9990			7.9520	112.5857	$< 10^{-4}$	$< 10^{-4}$
se	0.0393	0.0404	0.0505						
rmse	0.0393	0.0404	0.0505						
nse	0.0020	0.0020	0.0022						
FB coef	1.0016	1.0011	0.9995			7.9517	112.5285	$< 10^{-4}$	$< 10^{-4}$
se	0.0395	0.0406	0.0503						
rmse	0.0395	0.0406	0.0505						
nse	0.0020	0.0020	0.0022						
			HAUSMAN-TAYLOR WORLD						
True	1	1	1	1	1	8	32		
HT coef	0.9904	0.9880	1.0059	0.9879	1.0195	7.8967	43.0953		
se	0.1082	0.1092	0.0511	0.6813	0.6150				
rmse	0.1139	0.1093	0.0521	0.6424	0.5878				
3S bootstrap coef	0.9341	0.9312	1.0011	0.9894	1.2898	7.9191	17.1702	$< 10^{-4}$	$< 10^{-4}$
se	0.0668	0.0648	0.0720	0.4734	0.1386				
rmse	0.0938	0.0945	0.0720	0.4735	0.3212				
3S t-dist coef	0.9270	0.9261	1.0005	0.9902	1.3023	7.9441	16.6030	$< 10^{-4}$	$< 10^{-4}$
se	0.0433	0.0416	0.0513	0.4345	0.0947				
rmse	0.0848	0.0848	0.0513	0.4344	0.3168				
nse	0.0021	0.0021	0.0023	0.0090	0.0047				
FB coef	0.9325	0.9305	1.0008	0.9893	1.2948	7.9402	16.9453		
se	0.0438	0.0426	0.0513	0.4412	0.0951				
rmse	0.0805	0.0815	0.0513	0.4411	0.3097				
nse	0.0024	0.0024	0.0023	0.0107	0.0054				

† The parameters $\pi_1, \ldots, \pi_5$ are omitted from the table.