Robust linear static panel data models using epsilon-contamination

Baltagi, Badi H. and Bresson, Georges and Chaturvedi, Anoop and Lacroix, Guy

Syracuse University, USA, Université Paris II, France, University of Allahabad, India, Université Laval, Canada

14 November 2014

Robust linear static panel data models using ε-contamination

Badi H. Baltagi*
Syracuse University, USA

Georges Bresson
Université Paris II, France

Anoop Chaturvedi
University of Allahabad, India

Guy Lacroix
Université Laval, Canada

November 14, 2014

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 coefficients 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. We illustrate the performance of our estimator relative to classic panel estimators using data on earnings and crime.

Keywords: ε-contamination, hyper g-priors, type-II maximum likelihood posterior density, panel data, robust Bayesian estimator, three-stage hierarchy.
JEL classification: C11, C23, C26.

*Corresponding author. Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244–1020 USA. Tel.: +1 315 443 1630. *E-mail:* bbaltagi@maxwell.syr.edu (B.H. Baltagi), bresson-georges@orange.fr (G. Bresson), anoopchaturv@gmail.com (A. Chaturvedi), Guy.Lacroix@ecn.ulaval.ca (G. Lacroix).

We thank participants at the research seminars of Maastricht University (NL) and Université Laval (Canada) for helpful comments and suggestions. Usual disclaimers apply.

1 Introduction

The choice of which classic panel data estimator to employ for a linear static regression model depends upon the hypothesized correlation between the individual effects and the regressors. Random effects assume that the regressors are uncorrelated with the individual effects, while fixed effects assume that all of the regressors are correlated with the individual effects (see Mundlak (1978) and Chamberlain (1980)). When a subset of the regressors are correlated with the individual effects, one employs the instrumental variables estimator of Hausman and Taylor (1981). In contrast, a Bayesian needs to specify the distributions of the priors (and the hyperparameters in hierarchical models) to estimate the model. It is well-known that Bayesian models can be sensitive to misspecification of these distributions. In empirical analyses, the choice of specific distributions is often made out of convenience. 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.

Often the subjective information available to the experimenter may not be enough for correct elicitation of a single prior distribution for the parameters, which is an essential requirement for the implementation of classical Bayes procedures. The robust Bayesian approach relies upon a class of prior distributions and selects an appropriate prior in a data dependent fashion. An interesting class of prior distributions suggested by Berger (1983, 1985) is the ε-contamination class, which combines the elicited prior for the parameters, termed as base prior, with a possible contamination class of prior distributions and implements Type II maximum likelihood (ML-II) procedure for the selection of prior distribution for the parameters. The primary advantage of using such a contamination class of prior distributions is that the resulting estimator obtained by using ML-II procedure performs well even if the true prior distribution is away from the elicited base prior distribution.

The objective of our paper is to propose a robust Bayesian approach to linear static panel data models. This approach departs from the standard Bayesian model in two ways. First, we consider the ε-contamination class of prior distributions for the model parameters (and for the individual effects). The base elicited prior is assumed to be contaminated and the contamination is hypothesized to belong to some suitable class of prior distributions. Second, both the base elicited priors and the ε-contaminated priors use Zellner's (1986) g-priors rather than the standard Wishart distributions for the variance-covariance matrices. The paper contributes to the panel data literature by presenting a general robust Bayesian framework. It encompasses the above mentioned conventional frequentist specifications and their associated estimation methods and is presented in Section 2.

In Section 3 we derive the Type II maximum likelihood posterior mean and the variance-covariance matrix of the coefficients in a two-stage hierarchy model. We show that the ML-II posterior mean of the coefficients is a shrinkage estimator, *i.e.*, a weighted average of the Bayes estimator under a base prior and the data-dependent empirical Bayes estimator. Furthermore, we show in a panel data context that the ε-contamination model is capable of extracting more information from the data and is thus superior to the classical Bayes estimator based on a single base prior.

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 differences between the two-stage and the three-stage hierarchy models pertain 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 the robust Bayesian estimators through extensive Monte Carlo experiments. These include the standard random effects model as well as Mundlak-type, Chamberlain-type and Hausman-Taylor-type models. We find that the three-stage hierarchy model outperforms the standard frequentist estimation methods. Section 6 compares the relative performance of the robust Bayesian estimators and the standard classical panel data estimators with real applications using panel data on earnings and crime. We conclude the paper in Section 7.

2 The general setup

Let us specify a Gaussian linear mixed model:

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

where X'_{it} is a $(1 \times K_1)$ vector of explanatory variables excluding the intercept, and β is a $(K_1 \times 1)$ vector of slope 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, ε_{it} is a remainder term assumed to be normally distributed, *i.e.* $\varepsilon_{it} \sim N\left(0, \tau^{-1}\right)$. The distribution of ε_{it} is parametrized in terms of its precision τ rather that its variance $\sigma^2_\varepsilon \left(= 1/\tau\right)$. In the statistics literature, the elements of β do not differ across i and are referred to as *fixed effects* whereas the b_i are referred to as *random effects*.[1] The resulting model in (1) is a Gaussian mixed linear model. 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 b_i is assumed to be correlated (uncorrelated) with all the X'_{it}s, they will be termed fixed (random) effects.[2]

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), Greenberg (2008), Koop (2003), Chib (2008), Zheng et al. (2008), Rendon (2013)):

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

where y is $(NT \times 1)$, X is $(NT \times K_1)$, W is $(NT \times K_2)$, ε is $(NT \times 1)$, and $\Sigma = \tau^{-1} I_{NT}$ is $(NT \times NT)$. The parameters depend upon hyperparameters which follow random distributions. The second stage (also called *fixed effects model* in the Bayesian literature) updates the distribution of the parameters. The third stage (also called *random effects model* in the Bayesian literature) updates the distribution of the hyperparameters. As stated by Smith (1973, pp. 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.

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

[2] 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.

In that case, the hyperparameters only concern the variance-covariance matrix of the b coefficients[3] and the precision τ. As is well-known, Bayesian models may be sensitive to possible 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 Fernandez, Ley and Steel (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), Chaturvedi and Singh (2012) among others, we consider the ε-contamination class of prior distributions for (β, b, τ):

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

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

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

β_0, b_0, g_0 and h_0 are known scalar hyperparameters of the base prior $\pi_0(\beta, b, \tau \mid g_0, h_0)$. The probability density function (henceforth pdf) of the base prior $\pi_0(.)$ is given by:

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

The possible class of contamination Q is defined as:

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

with

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

where β_q, b_q, g_q and h_q are unknown. 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:

[3] Note that in (2), the prior distribution of β and b are assumed to be independent, so $\mathrm{Var}[\theta]$ 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: $\beta_0 \sim N(\beta_{00}, \Lambda_{\beta_0})$, $b_0 \sim N(b_{00}, \Lambda_{b_0})$ and $\Lambda_\beta^{-1} \sim Wish(\nu_\beta, R_\beta)$ (see for instance, Greenberg (2008), Hsiao and Pesaran (2008), Koop (2003), Bresson and Hsiao (2011)).

1. a one-step estimation of the ML-II posterior distribution[4] of β, b and τ;
2. or a two-step approach as follows:
   (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 slope 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 difficult. 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:

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

and

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

with

$$\begin{aligned}
\pi_0(\beta, \tau \mid g_0) &= \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \tau^{-1} |\Lambda_X|^{1/2} \exp\left(-\frac{\tau g_0}{2}\left(\beta - \beta_0 \iota_{K_1})'\Lambda_X(\beta - \beta_0 \iota_{K_1})\right)\right), \\
\pi_0(b, \tau \mid h_0) &= \left(\frac{\tau h_0}{2\pi}\right)^{\frac{K_2}{2}} \tau^{-1} |\Lambda_W|^{1/2} \exp\left(-\frac{\tau h_0}{2}(b - b_0 \iota_{K_2})'\Lambda_W(b - b_0 \iota_{K_2})\right).
\end{aligned}$$

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

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[5]:

$$\begin{aligned}
m(y^* \mid \pi_0, b, g_0) &= \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0(\beta, \tau \mid g_0) \times p(y^* \mid X, b, \tau) \ d\beta \ d\tau \qquad (8) \\
&= \widetilde{H}\left(\frac{g_0}{g_0 + 1}\right)^{K_1/2} \left(1 + \left(\frac{g_0}{g_0 + 1}\right)\left(\frac{R^2_{\beta_0}}{1 - R^2_{\beta_0}}\right)\right)^{-\frac{NT}{2}}
\end{aligned}$$

with

$$\widetilde{H} = \frac{\Gamma\left(\frac{NT}{2}\right)}{\pi^{\left(\frac{NT}{2}\right)} v(b)^{\left(\frac{NT}{2}\right)}}, \qquad (9)$$

[4] *"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 $m(y \mid \pi)$ over Γ. This process is called Type II maximum likelihood by Good (1965)"* (Berger and Berliner (1986), page 463.)

[5] Derivation of all the following expressions can be found in the Appendix.

$$R^2_{\beta_0} = \frac{(\widehat{\beta}(b) - \beta_0 \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_0 \iota_{K_1})}{(\widehat{\beta}(b) - \beta_0 \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_0 \iota_{K_1}) + v(b)}, \tag{10}$$

$\widehat{\beta}(b) = \Lambda_X^{-1} X' y^*$ and $v(b) = (y^* - X\widehat{\beta}(b))'(y^* - X\widehat{\beta}(b))$ and where $\Gamma(\cdot)$ is the Gamma function.

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

$$m(y^* \mid q, b, g_0) = \widetilde{H} \left(\frac{g_q}{g_q + 1} \right)^{\frac{K_1}{2}} \left(1 + \left(\frac{g_q}{g_q + 1} \right) \left(\frac{R^2_{\beta_q}}{1 - R^2_{\beta_q}} \right) \right)^{-\frac{NT}{2}}, \tag{11}$$

where

$$R^2_{\beta_q} = \frac{(\widehat{\beta}(b) - \beta_q \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_q \iota_{K_1})}{(\widehat{\beta}(b) - \beta_q \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_q \iota_{K_1}) + v(b)}. \tag{12}$$

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

$$m(y^* \mid \pi, b, g_0) = (1 - \varepsilon) m(y^* \mid \pi_0, b, g_0) + \varepsilon m(y^* \mid q, b, g_0) \tag{13}$$

and

$$\sup_{\pi \in \Gamma} m(y^* \mid \pi, b, g_0) = (1 - \varepsilon) m(y^* \mid \pi_0, b, g_0) + \varepsilon \sup_{q \in Q} m(y^* \mid q, b, g_0). \tag{14}$$

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

$$\widehat{\beta}_q = \left(\iota'_{K_1} \Lambda_X \iota_{K_1} \right)^{-1} \iota'_{K_1} \Lambda_X \widehat{\beta}(b) \tag{15}$$

and

$$\begin{aligned} \widehat{g}_q &= \min(g_0, g^*) \qquad (16) \\ \text{with } g^* &= \max\left[\left(\frac{(NT - K_1)}{K_1} \frac{(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1})}{v(b)} - 1 \right)^{-1}, 0 \right] \\ &= \max\left[\left(\frac{(NT - K_1)}{K_1} \left(\frac{R^2_{\widehat{\beta}_q}}{1 - R^2_{\widehat{\beta}_q}} \right) - 1 \right)^{-1}, 0 \right]. \end{aligned}$$

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

$$m(y^* \mid \widehat{q}, b, g_0) = \widetilde{H} \left(\frac{\widehat{g}_q}{\widehat{g}_q + 1} \right)^{\frac{K_1}{2}} \left(1 + \left(\frac{\widehat{g}_q}{\widehat{g}_q + 1} \right) \left(\frac{R^2_{\widehat{\beta}_q}}{1 - R^2_{\widehat{\beta}_q}} \right) \right)^{-\frac{NT}{2}}. \tag{17}$$

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

$$\begin{aligned} \widehat{\pi}^*(\beta \mid g_0) &= \int_0^\infty \widehat{\pi}^*(\beta, \tau \mid g_0)\, d\tau \\ &= \widehat{\lambda}_{\beta, g_0} \int_0^\infty \pi_0^*(\beta, \tau \mid g_0)\, d\tau + \left(1 - \widehat{\lambda}_{\beta, g_0} \right) \int_0^\infty q^*(\beta, \tau \mid g_0)\, d\tau \\ &= \widehat{\lambda}_{\beta, g_0} \pi_0^*(\beta \mid g_0) + \left(1 - \widehat{\lambda}_{\beta, g_0} \right) \widehat{q}^*(\beta \mid g_0) \qquad (18) \end{aligned}$$

with

$$\widehat{\lambda}_{\beta,g_0}=\left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}_q}{\widehat{g}_q+1}}{\frac{g_0}{g_0+1}}\right)^{K_1/2}\left(\frac{1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right)}{1+\left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\beta}_q}}{1-R^2_{\widehat{\beta}_q}}\right)}\right)^{\frac{NT}{2}}\right]^{-1}. \tag{19}$$

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

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

$$M_{0,\beta}=\frac{(g_0+1)}{v(b)}\Lambda_X \text{ and } \xi_{0,\beta}=1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right). \tag{20}$$

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

$$\beta_*(b\mid g_0)=\frac{\widehat{\beta}(b)+g_0\beta_0\iota_{K_1}}{g_0+1}. \tag{21}$$

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

$$\xi_{q,\beta}=1+\left(\frac{\widehat{g}_q}{\widehat{g}_q+1}\right)\left(\frac{R^2_{\widehat{\beta}_q}}{1-R^2_{\widehat{\beta}_q}}\right) \text{ and } M_{q,\beta}=\left(\frac{(\widehat{g}_q+1)}{v(b)}\right)\Lambda_X, \tag{22}$$

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:

$$\widehat{\beta}_{EB}(b\mid g_0)=\frac{\widehat{\beta}(b)+\widehat{g}_q\widehat{\beta}_q\iota_{K_1}}{\widehat{g}_q+1}. \tag{23}$$

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

$$\begin{aligned}\widehat{\beta}_{ML-II} &= E\left[\widehat{\pi}^*(\beta\mid g_0)\right] \qquad (24)\\ &= \widehat{\lambda}_{\beta,g_0}E\left[\pi_0^*(\beta\mid g_0)\right]+\left(1-\widehat{\lambda}_{\beta,g_0}\right)E\left[\widehat{q}^*(\beta\mid g_0)\right]\\ &= \widehat{\lambda}_{\beta,g_0}\beta_*(b\mid g_0)+\left(1-\widehat{\lambda}_{\beta,g_0}\right)\widehat{\beta}_{EB}(b\mid g_0).\end{aligned}$$

The ML-II posterior mean of β, given b and g_0 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 mean 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 mean 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.

The ML-II posterior variance-covariance matrix of β is given by (see Berger (1985) p. 207):

$$\begin{aligned} Var\left(\widehat{\beta}_{ML-II}\right) &= \widehat{\lambda}_{\beta,g_0} Var\left[\pi_0^*\left(\beta \mid g_0\right)\right] + \left(1-\widehat{\lambda}_{\beta,g_0}\right) Var\left[\widehat{q}^*\left(\beta \mid g_0\right)\right] \\ &+ \widehat{\lambda}_{\beta,g_0}\left(1-\widehat{\lambda}_{\beta,g_0}\right)\left(\beta_*(b \mid g_0) - \widehat{\beta}_{EB}\left(b \mid g_0\right)\right)\left(\beta_*(b \mid g_0) - \widehat{\beta}_{EB}\left(b \mid g_0\right)\right)' \\ &= \widehat{\lambda}_{\beta,g_0}\left(\frac{\xi_{0,\beta}}{NT-2}\frac{v\left(b\right)}{g_0+1}\right)\Lambda_X^{-1} \qquad (25) \\ &+ \left(1-\widehat{\lambda}_{\beta,g_0}\right)\left(\frac{\xi_{q,\beta}}{NT-2}\frac{v\left(b\right)}{\widehat{g}_q+1}\right)\Lambda_X^{-1} \\ &+ \widehat{\lambda}_{\beta,g_0}\left(1-\widehat{\lambda}_{\beta,g_0}\right)\left(\beta_*(b \mid g_0) - \widehat{\beta}_{EB}\left(b \mid g_0\right)\right)\left(\beta_*(b \mid g_0) - \widehat{\beta}_{EB}\left(b \mid g_0\right)\right)'. \end{aligned}$$

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:

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

with

$$\widehat{\lambda}_{b,h_0} = \left[1 + \frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{h}}{\widehat{h}+1}}{\frac{h_0}{h_0+1}}\right)^{K_2/2}\left(\frac{1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)}{1+\left(\frac{\widehat{h}}{\widehat{h}+1}\right)\left(\frac{R_{\widehat{b}_q}^2}{1-R_{\widehat{b}_q}^2}\right)}\right)^{\frac{NT}{2}}\right]^{-1}, \qquad (26)$$

where

$$R_{b_0}^2 = \frac{(\widehat{b}\left(\beta\right) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}\left(\beta\right) - b_0\iota_{K_2})}{(\widehat{b}\left(\beta\right) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}\left(\beta\right) - b_0\iota_{K_2}) + v\left(\beta\right)}, \qquad (27)$$

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

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

$$\widehat{b}_q = \left(\iota_{K_2}'\Lambda_W\iota_{K_2}\right)^{-1}\iota_{K_2}'\Lambda_W\widehat{b}\left(\beta\right) \qquad (28)$$

and

$$\begin{aligned} \widehat{h}_q &= \min\left(h_0, h^*\right) \qquad (29) \\ \text{with } h^* &= \max\left[\left(\frac{(NT-K_2)}{K_2}\frac{(\widehat{b}\left(\beta\right) - \widehat{b}_q\iota_{K_2})'\Lambda_W(\widehat{b}\left(\beta\right) - \widehat{b}_q\iota_{K_2})}{v\left(\beta\right)} - 1\right)^{-1}, 0\right] \\ &= \max\left[\left(\frac{(NT-K_2)}{K_2}\left(\frac{R_{\widehat{b}_q}^2}{1-R_{\widehat{b}_q}^2}\right) - 1\right)^{-1}, 0\right]. \end{aligned}$$

$\pi_0^*\left(b \mid h_0\right)$ 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

$$M_{0,b} = \frac{(h_0+1)}{v\left(\beta\right)}\Lambda_W \text{ and } \xi_{0,b} = 1 + \left(\frac{h_0}{h_0+1}\right)\frac{(\widehat{b}\left(\beta\right) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}\left(\beta\right) - b_0\iota_{K_2})}{v\left(\beta\right)}. \qquad (30)$$

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

$$b_*(\beta \mid h_0) = \frac{\widehat{b}(\beta) + h_0 b_0 \iota_{K_2}}{h_0 + 1}. \tag{31}$$

$q^* (b \mid h_0)$ 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

$$\xi_{1,b} = 1 + \left(\frac{\widehat{h}_q}{\widehat{h}_q + 1}\right) \frac{(\widehat{b}(\beta) - \widehat{b}_q \iota_{K_2})' \Lambda_W (\widehat{b}(\beta) - \widehat{b}_q \iota_{K_2})}{v(\beta)} \text{ and } M_{1,b} = \left(\frac{\widehat{h} + 1}{v(\beta)}\right) \Lambda_W \tag{32}$$

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

$$\widehat{b}_{EB} (\beta \mid h_0) = \frac{\widehat{\beta}(b) + \widehat{h}_q \widehat{b}_q \iota_{K_2}}{\widehat{h}_q + 1}. \tag{33}$$

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

$$\widehat{b}_{ML-II} = \widehat{\lambda}_b b_*(\beta \mid h_0) + \left(1 - \widehat{\lambda}_\beta\right) \widehat{b}_{EB} (\beta \mid h_0) \tag{34}$$

and the ML-II posterior variance-covariance matrix of b is given by:

$$\begin{aligned} Var\left(\widehat{b}_{ML-II}\right) &= \widehat{\lambda}_{b,h_0} \left(\frac{\xi_{0,b}}{NT-2} \frac{v(\beta)}{h_0 + 1}\right) \Lambda_W^{-1} \\ &+ \left(1 - \widehat{\lambda}_{b,h_0}\right) \left(\frac{\xi_{1,b}}{NT-2} \frac{v(\beta)}{\widehat{h}_q + 1}\right) \Lambda_W^{-1} \\ &+ \widehat{\lambda}_{b,h_0} \left(1 - \widehat{\lambda}_{b,h_0}\right) \left(b_*(\beta \mid h_0) - \widehat{b}_{EB} (\beta \mid h_0)\right) \left(b_*(\beta \mid h_0) - \widehat{b}_{EB} (\beta \mid h_0)\right)'. \end{aligned} \tag{35}$$

As our estimator is a shrinkage estimator, it is not necessary to draw thousands of multivariate t-distributions to compute the mean and variance after burning draws. We can use an iterative shrinkage approach as suggested by Maddala *et al.* (1997) (see also Baltagi *et al.* (2008)) to compute the ML-II posterior mean and variance-covariance matrix of β and b.

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 objective 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), Greenberg (2008), Chib (2008), Zheng *et al.* (2008) among others, hyperparameters only concern the variance-covariance matrix of the b coefficients. Because we use g-priors at 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.

Zellner and Siow (1980) proposed a Cauchy prior on g which is not as popular as the g-prior since closed form expressions for the marginal likelihoods are not available. More recently, and as an alternative to the Zellner-Siow's prior, Liang *et al.* (2008) (see also Cui and George (2008)) have proposed a Pareto type II hyper-g prior whose pdf is defined as:

$$p(g) = \frac{(k-2)}{2}(1+g)^{-\frac{k}{2}} \ , g > 0, \tag{36}$$

which is a proper prior for $k > 2$. One advantage of the hyper-g prior is that the posterior distribution of g, given a model, is available in closed form. Unfortunately, the normalizing constant is a Gaussian hypergeometric function and a Laplace approximation is usually required to compute the integral of its representation for large samples, NT, and large R^2. Liang *et al.* (2008) have shown that the best choices for k are given by[6] $2 < k \leq 4$. Maruyama and George (2011, 2014) proposed a generalized hyper-g prior:

$$p(g) = \frac{g^{c-1}(1+g)^{-(c+d)}}{B(c,d)} \ , c > 0, \ d > 0, \tag{37}$$

where $B(\cdot)$ is the Beta function. This Beta-prime (or Pearson Type VI) hyper prior for g is a generalization of the Pareto type II hyper-g prior since the expression in (36) is equivalent to that in (37) when $c = 1$. In that specific case, $d = \frac{(k-2)}{2}$. Using the generalized hyper-g prior specification, the three-stage hierarchy of the model can be defined as:

$$\begin{aligned} \text{First stage}: & \ y \sim N(X\beta + Wb, \Sigma), \ \Sigma = \tau^{-1} I_{NT} \\ \text{Second stage}: & \ \beta \sim N\left(\beta_0 \iota_{K_1}, (\tau g_0 \Lambda_X)^{-1}\right) \ , b \sim N\left(b_0 \iota_{K_2}, (\tau h_0 \Lambda_W)^{-1}\right) \\ \text{Third stage}: & \ h_0 \sim \beta'(c,d) \ \rightarrow p(h_0) = \frac{h_0^{c-1}(1+h_0)^{-(c+d)}}{B(c,d)} \ , c > 0, \ d > 0. \end{aligned} \tag{38}$$

As our objective is to account for the uncertainty about the prior means of the base prior $\pi_0(\beta, b, \tau)$, we do not need to introduce an ε-contamination class of prior distributions for the hyperparameters of the third stage of the hierarchy. Moreover, Berger (1985, p. 232) has stressed that the choice of a specific functional form for the third stage matters little. Sinha and Jayaraman (2010a, 2010b) 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's (1985) assertion that the form of the second stage prior (the third stage of the hierarchy) does not affect the Bayes decision. 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

$$\begin{aligned} m(\widetilde{y} \mid \pi_0, \beta) &= \int_0^\infty m(\widetilde{y} \mid \pi_0, \beta, h_0)\, p(h_0) dh_0 \\ &= \frac{\widetilde{H}}{B(c,d)} \int_0^1 (\varphi)^{\frac{K_2}{2}+c-1}(1-\varphi)^{d-1}\left(1+\varphi\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right)^{-\frac{NT}{2}} d\varphi \end{aligned} \tag{39}$$

[6] In their Monte Carlo simulations, Liang *et al.* (2008) use $k = 3$ and $k = 4$.

which can be written as:

$$m\left(\widetilde{y}\mid\pi_0,\beta\right)=\frac{B(d,\frac{K_2}{2}+c)}{B\left(c,d\right)}\widetilde{H}\times_2 F_1\left(\frac{NT}{2};\frac{K_2}{2}+c;\frac{K_2}{2}+c+d;-\left(\frac{R^2_{\widehat{b}_0}}{1-R^2_{\widehat{b}_0}}\right)\right), \quad (40)$$

where ${}_2F_1(.)$ is the Gaussian hypergeometric function (see Abramovitz and Stegun (1970) and the Appendix). As shown by Liang *et al.* (2008), numerical overflow is problematic for moderate to large NT and large $R^2_{\widehat{b}_0}$. As the Laplace approximation involves an integral with respect to a normal kernel, we follow the suggestion of Liang et al. (2008) and develop an expansion after a change of variable given by $\phi=\log\left(\frac{h_0}{h_0+1}\right)$ (see the Appendix).

Similar to the conditional predictive density corresponding to the contaminated prior on β (see eq(17)), the unconditional predictive density corresponding to the contaminated prior on b is given by:

$$m\left(\widetilde{y}\mid\widehat{q},\beta\right)=\int_0^\infty m\left(\widetilde{y}\mid\widehat{q},\beta,h_0\right)p(h_0)dh_0 \quad (41)$$

$$=\frac{\widetilde{H}}{B\left(c,d\right)}\times\int_0^{h^*}\left(\begin{array}{c}\left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left(1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right)\right)^{-\frac{NT}{2}}\\ \times h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d}\end{array}\right)dh_0$$

$$+\frac{\widetilde{H}}{B\left(c,d\right)}\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right)\right)^{-\frac{NT}{2}}\times$$

$$\int_{h^*}^\infty h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d}dh_0.$$

$$m\left(\widetilde{y}\mid\widehat{q},\beta\right)= \quad (42)$$

$$=\frac{\widetilde{H}}{B\left(c,d\right)}\left\{\begin{array}{c}\frac{2.\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c}}{K_2+2c}\\ \times F_1\left(\frac{K_2}{2}+c;1-d;\frac{NT}{2};\frac{K_2}{2}+c+1;\frac{h^*}{h^*+1};-\frac{h^*}{h^*+1}\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right)\right)\\ +\left\{\begin{array}{c}\left[\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{\widehat{b}_q}}{1-R^2_{\widehat{b}_q}}\right)\right)^{-\frac{NT}{2}}\right]\\ \times\left[\begin{array}{c}B\left(c,d\right)-\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c}\\ \times{}_2F_1\left(c;d-1;c+1;\frac{h^*}{h^*+1}\right)\end{array}\right]\end{array}\right\}\end{array}\right\},$$

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

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

Hence, we can write

$$\widehat{\lambda}_b = \int_0^\infty \widehat{\lambda}_{b,h_0} p(h_0) dh_0 = \left[1 + \left(\frac{\varepsilon}{1-\varepsilon}\right) . \frac{m\left(\widetilde{y} \mid \widehat{q}, \beta\right)}{m\left(\widetilde{y} \mid \pi_0, \beta\right)}\right]^{-1}. \tag{43}$$

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

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

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

$$\begin{aligned}
\widehat{b}_{EB}\left(\beta\right) &= \int_0^\infty \widehat{b}_{EB}\left(\beta \mid h_0\right) p(h_0) dh_0 \qquad (44)\\
&= \frac{1}{B\left(c,d\right)}\left[\begin{array}{c} \widehat{b}\left(\beta\right) \int_0^{h^*} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_q \iota_{K_2} \int_0^{h^*} h_0^{c}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ + \left\{\widehat{b}\left(\beta\right)\left(\frac{1}{g^*+1}\right) + \widehat{b}_q \iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\} \int_{h^*}^\infty h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \end{array}\right]\\
&= \frac{1}{B\left(c,d\right)}\left[\begin{array}{c} \widehat{b}\left(\beta\right) \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1\left(c; -d; c+1; \frac{h^*}{h^*+1}\right) \\ + \widehat{b}_q \iota_{K_2} \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1\left(c+1; 1-d; c+2; \frac{h^*}{h^*+1}\right) \\ + \left\{\widehat{b}\left(\beta\right)\left(\frac{1}{h^*+1}\right) + \widehat{b}_q \iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\} \\ \times \left[\begin{array}{c} B\left(c,d\right) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2 F_1\left(c; d-1; c+1; \frac{h^*}{h^*+1}\right) \end{array}\right] \end{array}\right]
\end{aligned}$$

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

$$\begin{aligned}
\widehat{\pi}^*\left(b\right) &= \int_0^\infty \widehat{\pi}^*\left(b,\tau\right) d\tau = \widehat{\lambda}_b \int_0^\infty \pi_0^*\left(b,\tau\right) d\tau + \left(1-\widehat{\lambda}_b\right) \int_0^\infty q^*\left(b,\tau\right) d\tau \qquad (45)\\
&= \widehat{\lambda}_b \pi_0^*\left(b\right) + \left(1-\widehat{\lambda}_b\right) \widehat{q}^*\left(b\right).
\end{aligned}$$

$\pi_0^*\left(b\right)$ 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

$$M_{0,b} = \frac{(h_0+1)}{v\left(\beta\right)} \Lambda_W \text{ and } \xi_{0,b} = 1 + \left(\frac{h_0}{h_0+1}\right)\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right). \tag{46}$$

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

$$\xi_{q,b} = 1 + \left(\frac{\widehat{h}_q}{\widehat{h}_q+1}\right)\left(\frac{R_{\widehat{b}_q}^2}{1-R_{\widehat{b}_q}^2}\right) \text{ and } M_{q,b} = \left(\frac{(\widehat{h}_q+1)}{v\left(\beta\right)}\right)\Lambda_W. \tag{47}$$

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

$$\begin{aligned}\widehat{b}_{ML-II} &= E\left[\widehat{\pi}^*(b)\right] = \widehat{\lambda}_b E\left[\pi_0^*(b)\right] + \left(1-\widehat{\lambda}_b\right) E\left[\widehat{q}^*(b)\right] \\ &= \widehat{\lambda}_b b_*(\beta) + \left(1-\widehat{\lambda}_b\right)\widehat{b}_{EB}(\beta)\end{aligned} \quad (48)$$

and the ML-II posterior variance-covariance matrix of b is given by:

$$\begin{aligned}Var\left(\widehat{b}_{ML-II}\right) = \widehat{\lambda}_b &\left(\frac{\xi_{0,b}}{NT-2}.\frac{v(\beta)}{h_0+1}\right)\Lambda_W^{-1} \\ &+ \left(1-\widehat{\lambda}_b\right)\left(\frac{\xi_{q,b}}{NT-2}\frac{v(\beta)}{\widehat{h}_q+1}\right)\Lambda_W^{-1} \\ &+ \widehat{\lambda}_b\left(1-\widehat{\lambda}_b\right)\left(b_*(\beta)-\widehat{b}_{EB}(\beta)\right)\left(b_*(\beta)-\widehat{b}_{EB}(\beta)\right)'.\end{aligned} \quad (49)$$

The main differences with the two-stage hierarchy model relate 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 $\widehat{\lambda}_{b,h_0}$). Once again, as our estimator is a shrinkage estimator, it is not necessary to draw thousands of multivariate t-distributions to compute the mean and the variance after burning draws. We can use an iterative shrinkage approach to calculate the ML-II posterior mean and variance-covariance matrix of β and b.

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:

$$\begin{aligned}y_{it} &= x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + Z_{1,i}\eta_1 + Z_{2,i}\eta_2 + \mu_i + \varepsilon_{it} \\ \text{for } i &= 1,...,N \ , \ t = 1,...,T\end{aligned} \quad (50)$$

with

$$x_{1,1,it} = 0.7x_{1,1,it-1} + \delta_i + \zeta_{it} \quad (51)$$

$$x_{1,2,it} = 0.7x_{1,2,it-1} + \theta_i + \varsigma_{it} \quad (52)$$

$$\varepsilon_{it} \sim N\left(0,\tau^{-1}\right), \ (\delta_i, \theta_i, \zeta_{it}, \varsigma_{it}) \sim U(-2,2) \quad (53)$$

$$\text{and } \beta_{1,1} = \beta_{1,2} = \beta_2 = 1. \quad (54)$$

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

$$\eta_1 = \eta_2 = 0 \quad (55)$$

$$x_{2,it} = 0.7x_{2,it-1} + \kappa_i + u_{it} \ , \ (\kappa, u_{it}) \sim U(-2,2) \quad (56)$$

$$\mu_i \sim N\left(0,\sigma_\mu^2\right) \ , \ \rho = \frac{\sigma_\mu^2}{\sigma_\mu^2+\tau^{-1}} = 0.3, \ 0.8. \quad (57)$$

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

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

$$\begin{aligned}
\eta_1 &= \eta_2 = 0; & (58)\\
x_{2,it} &= \delta_{2,i} + \omega_{2,it}\ ,\ \delta_{2,i} \sim N(m_{\delta_2}, \sigma^2_{\delta_2}),\ \omega_{2,it} \sim N(m_{\omega_2}, \sigma^2_{\omega_2}); & (59)\\
m_{\delta_2} &= m_{\omega_2} = 1,\ \sigma^2_{\delta_2} = 8,\ \sigma^2_{\omega_2} = 2; & (60)\\
\mu_i &= \overline{x}_{2,i}\pi + \nu_i,\ \nu_i \sim N(0, \sigma^2_\nu),\ \overline{x}_{2,i} = \frac{1}{T}\sum_{t=1}^{T} x_{2,it}; & (61)\\
\sigma^2_\nu &= 1,\ \pi = 0.8. & (62)
\end{aligned}$$

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

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

$$\begin{aligned}
\eta_1 &= \eta_2 = 0; & (63)\\
x_{2,it} &= \delta_{2,i} + \omega_{2,it}\ ,\ \delta_{2,i} \sim N(m_{\delta_2}, \sigma^2_{\delta_2}),\ \omega_{2,it} \sim N(m_{\omega_2}, \sigma^2_{\omega_2}); & (64)\\
m_{\delta_2} &= m_{\omega_2} = 1,\ \sigma^2_{\delta_2} = 8,\ \sigma^2_{\omega_2} = 2; & (65)\\
\mu_i &= x_{2,i1}\pi_1 + x_{2,i2}\pi_2 + ... + x_{2,iT}\pi_T + \nu_i,\ \nu_i \sim N(0, \sigma^2_\nu); & (66)\\
\sigma^2_\nu &= 1,\ \pi_t = (0.8)^{T-t} \text{ for } t = 1, ..., T. & (67)
\end{aligned}$$

$x_{1,1,it}$ 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:

$$\begin{aligned}
\eta_1 &= \eta_2 = 1; & (68)\\
x_{2,it} &= 0.7x_{2,it-1} + \mu_i + u_{it}\ ,\ u_{it} \sim U(-2, 2); & (69)\\
Z_{1,i} &= 1,\ \forall i; & (70)\\
Z_{2,i} &= \mu_i + \delta_i + \theta_i + \xi_i,\ \xi_i \sim U(-2, 2); & (71)\\
\mu_i &\sim N\left(0, \sigma^2_\mu\right)\ ,\ \text{and } \rho = \frac{\sigma^2_\mu}{\sigma^2_\mu + \tau^{-1}} = 0.3,\ 0.8. & (72)
\end{aligned}$$

$x_{1,1,it}$ 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 ε_{it}.

For each set-up, we vary the size of our panel. We choose several (N, T) pairs with $N = 100$, 500 and $T = 5$, 10. We also choose $N = 50$, $T = 20$ as is typical for U.S. state panel data or country macro-panels. We generate the data by choosing initial values of $x_{1,1,it}$ and $x_{1,2,it}$ to be zero. We generate $x_{1,1,it}$, $x_{1,2,it}$, ε_{it}, ζ_{it}, u_{it}, ς_{it}, $\omega_{2,it}$ over $T + T_0$ time periods and we drop the first $T_0 (= 50)$ observations to reduce the dependence on initial values. We also use the robust Bayesian estimators for the two-stage hierarchy (2S) and for the three-stage hierarchy (3S) with $\varepsilon = 0.5$.

We must define the initial hyperparameters $\beta_0, b_0, g_0, h_0, \tau$ 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 the UIP, the RIC and the BRIC for the g priors.[7] In the normal regression case, and following Kass and Wasserman (1995), the unit information prior (UIP) corresponds to $g_0 = h_0 = 1/NT$, leading to Bayes factors that behave like the Bayesian Information Criterion (BIC). Foster and George (1994) calibrated priors for model selection based on the Risk inflation

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

criterion (RIC) and recommended the use of $g_0 = 1/K_1^2$, $h_0 = 1/N^2$. Fernández *et al.* (2001) recommended the BRIC (mix of BIC and RIC) using $g_0 = 1/\max(NT, K_1^2)$, $h_0 = 1/\max(NT, N^2)$. We use the UIP since the RIC and the BRIC lead to very small h_0 priors.

For the three-stage hierarchy (3S), we need to choose the coefficients (c, d) of the generalized hyper-g priors. Liang *et al.* (2008) stressed that the best parameter for the Pareto type II distribution was $k = 4$ which corresponds to $c = d = 1$ 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 principle (*i.e.*, h_0 close to $1/NT$), we chose $c = 0.1$ and $d = 1$. As our 2S and 3S estimators are shrinkage estimators (see eq.(24), eq.(34) for 2S and eq.(48) for 3S), we can use an iterative shrinkage approach as suggested by Maddala *et al.* (1997) with only 50 iterations. For the three-stage hierarchy (3S), we could use Gaussian hypergeometric functions ${}_2F_1$ and Appel functions F_1 with Laplace approximations but we prefer to solve the integrals numerically with adaptive quadrature methods (see Davis and Rabinowitz (1984), Press *et al.* (2007)). For each experiment, we run 1000 replications and we compute the mean, standard error and root mean squared error (RMSE) of the coefficients.

5.2 The results of the Monte Carlo study

5.2.1 The random effects world

Let us rewrite our general model (2): $y = X\beta + Wb + \varepsilon$, $\varepsilon \sim N(0, \Sigma)$, $\Sigma = \tau^{-1} I_{NT}$ as $y = X\beta + Z_\mu \mu + \varepsilon$ where $Z_\mu = I_N \otimes \iota_T$ is $(NT \times N)$, 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 $\nu = Z_\mu \mu + \varepsilon$ with $\operatorname{Var}(\nu) = \sigma_\mu^2 (I_N \otimes J_T) + \sigma_\varepsilon^2 I_{NT}$, where $J_T = \iota_T \iota_T'$ and are estimated using Feasible Generalized Least Squares (FGLS), (see Hsiao (2003) or Baltagi (2013)).

For the random effects world, we compare the standard FGLS estimator and our 2S and 3S estimators. In this specification, $X = [x_{1,1}, x_{1,2}, x_2]$, $W = Z_\mu$ and $b = \mu$. The results in Table 1 are based on $N = 100$, $T = 5$ with $\varepsilon = 0.5$. The proportion of heterogeneity in the total variance, measured by the ratio of the variance of the individual effects to the total variance (ρ). This is allowed to be either 30% or 80%. Table 1 shows that the 2S and 3S robust estimators have good properties. The estimated coefficients are very close to the true values. More interestingly, their standard errors (se) are much smaller than those of FGLS. Indeed, the standard errors of the latter estimator are nearly twice as large as those of the 2S and 3S estimators. The bias and RMSE, however, are similar to those of FGLS. Estimates of the remainder variance ($\sigma_\varepsilon^2 \equiv \tau^{-1}$) are the same and very close to the true value ($\sigma_\varepsilon^2 = 1$). The robust 3S also correctly estimates the variance of the individual effects (σ_μ^2). The 2S estimator yields unbiased coefficients but leads to a biased σ_μ^2. The weights λ_β and λ_b show the trade-off between the Bayes estimators ($\beta_*(b)$ and $b_*(\beta)$) and the empirical Bayes estimators ($\widehat{\beta}_{EB}(b)$ and $\widehat{b}_{EB}(\beta)$). In the 2S model, $\lambda_\beta = 28\%$ ($\lambda_b = 49\%$) which indicates that the empirical Bayes estimator $\widehat{\beta}_{EB}(b)$ ($\widehat{b}_{EB}(\beta)$) accounts for 72% (51%) of the weight in estimating the slope coefficients. In the 3S model, these ratios decrease considerably, from 22% to 11% for λ_β when ρ increases from 0.3 to 0.8. Furthermore, λ_b dramatically drops to zero. This means that only the empirical Bayes estimator is used in the estimation of the individual effects $(b \equiv \mu)$. In addition, the standard errors of the 2S and 3S are now three times smaller and the estimate of σ_μ^2 of the 3S corresponds perfectly to the true value. These results are confirmed when we increase the size of the sample of the short panel (large N, small T) (see Appendix, Tables A2 and A3).[8] Note that when ρ increases from 30% to 80%, the bias in the variance

[8]For the sake of brevity, we only present results of the three-stage hierarchy (3S) in what follows.

of the individual effects (σ^2_μ) is reduced and is smaller than -1.2% for $N = 100$, $T = 10$. It is also smaller than -0.04% for $N = 500$, $T = 10$. Even for a macro-panel (N small, T large), these results still hold (see Appendix Table A3 for $N = 50$, $T = 20$, $\rho = 0.8$ with $\varepsilon = 0.5$) and the bias in the variance of the individual effects (σ^2_μ) is smaller than -1.3%.

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 $y = X\beta + PX\pi + Z_\mu\varpi + \varepsilon$, 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 and our robust 3S estimator of the same specification. As $\mu_i = \overline{x}_{2,i}\pi + \nu_i$, the transformed model is given by: $y = x_{1,1}\beta_{1,1} + x_{1,2}\beta_{1,2} + x_2\beta_2 + Px_2\pi + Z_\mu\nu + \varepsilon$. In this specification, $X = [x_{1,1}, x_{1,2}, x_2, Px_2]$, $W = Z_\mu$ and $b = \nu$. The results are presented in Table 2 for $\varepsilon = 0.5$. Once again, they show the very good performance of the 3S estimator. Irrespective of the size of N and T, the estimated coefficients are very close to their true values and their standard errors (se) are smaller with the robust approach than with FGLS. They are much smaller for $\beta_{1,1}$ and $\beta_{1,2}$ — whose respective variables are uncorrelated with μ_i — but the difference with FGLS is smaller for β_2 and π. The bias and RMSE are similar to those of FGLS. Estimates of the remainder variance ($\sigma^2_\varepsilon \equiv \tau^{-1}$) are very close to the true value ($\sigma^2_\varepsilon = 1$). The weights λ_β and λ_b confirm that there is no trade-off between the Bayes estimators and the empirical Bayes estimators. Both λ_β and λ_b tend to zero, which means that only the empirical Bayes estimators are used in the estimation of the slope coefficients and the individual effects.[9] The same results hold for (N small, T large) macro-type panel, (see Appendix Table A4 for $N = 50$, $T = 20$ with $\varepsilon = 0.5$).

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: $y = X\beta + Z_\mu\underline{X}\Pi + Z_\mu\varpi + \varepsilon$. 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)) with our robust 3S estimator.[10] These are based on the transformed model: $y_{it} = x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + \sum_{t=1}^{T} x_{2,it}\pi_t + \nu_i + \varepsilon_{it}$ or $y = x_{1,1}\beta_{1,1} + x_{1,2}\beta_{1,2} + x_2\beta_2 + \underline{x_2}\Pi + Z_\mu\nu + \varepsilon$. In that specification, $X = \left[x_{1,1}, x_{1,2}, x_2, \underline{x_2}\right]$, $W = Z_\mu$ and $b = \nu$. Table 3 reports results for ($N = 100, 500, T = 5$). The estimated slope coefficients for the MCS and 3S are very close to the true values, but the standard errors (se) of the latter are between 10% to 20% smaller than those of MCS. The bias and RMSE of our robust 3S estimator are similar to those of MCS. Focusing on the five π_t coefficients, both MCS and 3S yield good estimates

[9]The concatenation of $[X, PX]$ does not change $R^2_{\beta_0}$ (as compared to the RE world) but it increases $R^2_{\beta_q}$ while remaining below 0.5. It therefore drives λ_β to zero.

[10]See the Appendix for a short presentation of the MCS estimator.

but the standard errors of the latter (se) are in most cases roughly 10% smaller. The 3S and the MCS give very close results both for the remainder variance (σ_ε^2) and the variance of the individual effects (σ_μ^2). Just as with the Mundlak-type FE world, the weights λ_β and λ_b confirm that there is no trade-off between the Bayes estimators and the empirical Bayes estimators.[11] Only the empirical Bayes estimators are used in the estimation of the slope coefficients and the individual effects irrespective of the value of N. One can note that σ_μ^2 is biased for MCS but not for 3S.

When we increase T from 5 to 10, we estimate ten π_t coefficients. The convexity of these time-varying coefficients is strong (from $\pi_1 = 0.13$ to $\pi_{10} = 1$) (see Tables A5-A6 in the Appendix) and both estimators manage to estimate the π_t parameters precisely. Likewise, the β parameters are very close to their true values and the standard errors are very similar across estimators. As a consequence, the RMSE's are nearly identical. Results in Tables 3, A5 and A6 show that 3S yields more precise estimates for small N. Whenever N or T increase, both MCS and 3S generate somewhat similar parameter estimates (β's and π_t), standard errors and RMSE's. The main advantage of 3S is that it provides unbiased estimates of σ_ε^2 and σ_μ^2 irrespective of N and T. The advantages of 3S over MCS are also illustrated in Table A7 in the Appendix. There we consider a typical macro panel data set consisting of $N = 50$, and $T = 20$ observations. Table A7 shows that both estimators yield parameter estimates (β's and π_t) that are very close to their true values. Yet, the RMSE associated with 3S are systematically smaller that those of MCS. In addition, 3S and MCS yield estimates for both σ_ε^2 and σ_μ^2 that are close to their true values.

5.2.4 The Hausman-Taylor world

The Hausman-Taylor model (henceforth HT, see Hausman and Taylor (1981)) posits that $y = X\beta + Z\eta + Z_\mu\mu + \varepsilon$, 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.[12] For our general model (2): $y = X\beta + Wb + \varepsilon$, we assume that ($\overline{X'_{2,i}}$, Z'_{2i} and μ_i) are jointly normally distributed:

$$\left(\begin{array}{c} \mu_i \\ \left(\begin{array}{c} \overline{X'_{2,i}} \\ Z'_{2i} \end{array} \right) \end{array} \right) \sim N \left(\left(\begin{array}{c} 0 \\ \left(\begin{array}{c} E_{\overline{X'_2}} \\ E_{Z'_2} \end{array} \right) \end{array} \right), \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right) \right), \tag{73}$$

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:

$$\mu_i \mid \overline{X'_{2,i}}, Z'_{2i} \sim N \left(\Sigma_{12}\Sigma_{22}^{-1}. \left(\begin{array}{c} \overline{X'_{2,i}} - E_{\overline{X'_2}} \\ Z'_{2i} - E_{Z'_2} \end{array} \right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right). \tag{74}$$

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

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}} \right) \theta_X + \left(Z'_{2i} - E_{Z'_2} \right) \theta_Z + \varpi_i, \tag{75}$$

where $\varpi_i \sim N\left(0, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$ is uncorrelated with ε_{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

[11] The concatenation of $[X, Z_\mu \underline{X}]$ increases the set of information to estimate β and Π. It does not change $R^2_{\beta_0}$ (as compared to the RE world) but it increases strongly $R^2_{\beta_q}$ while remaining below 0.5, therefore driving λ_β to zero.

[12] See the Appendix for a short presentation of the Hausman-Taylor estimator.

to avoid possible collinearity problems, we assume that the individual effects are given by:

$$\mu_i = \left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right)\theta_X + f\left[\left(\overline{X'_{2,i}} - E_{\overline{X'_2}}\right) \odot \left(Z'_{2i} - E_{Z'_2}\right)\right]\theta_Z + \varpi_i, \qquad (76)$$

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 non-linear 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 (76) 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, $y_{it} = x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + Z_{1,i}\eta_1 + Z_{2,i}\eta_2 + \mu_i + \varepsilon_{it}$ or $y = X_1\beta_1 + x_2\beta_2 + Z_1\eta_1 + Z_2\eta_2 + Z_\mu\mu + \varepsilon$. Then, we assume that

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

We propose adopting the following strategy: If the correlation between μ_i and Z_{2i} is quite large (> 0.2), use $f\,[.] = \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right)^2 \odot \left(Z_{2i} - E_{Z_2}\right)^s$ 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: $\widehat{\mu} = \left(Z'_\mu Z_\mu\right)^{-1} Z'_\mu \widehat{y}$ 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 $b = \nu$.

For the Hausman-Taylor world, we compare the IV method proposed by Hausman and Taylor (1981) with our robust 3S estimator. Table 4 gives the results for $N = 100$, $T = (5, 10)$, $\varepsilon = 0.5$ and $\rho = 0.3, 0.8$. It shows very good estimates of the slope coefficients with 3S, except for η_2 which is slightly biased. The coefficient β_2 of the time-varying variable x_2, (correlated with μ_i), is also well estimated. Similarly, the coefficient η_1 of the time-invariant variable Z_1, (uncorrelated with μ_i), is also well estimated. In contrast, the coefficient η_2 of the time-invariant variable Z_2, (correlated with μ_i), is slightly biased (3% to 4.7% for $(N = 100, T = 5)$ and 2.1% to 2.9% for $(N = 100, T = 10)$). This bias does not change when N increases (see Table A8 in the Appendix). However, the standard errors are considerably lower, especially for the coefficients η_1 and η_2 of the two time-invariant variables. The 95% confidence intervals obtained with 3S are much narrower and are entirely nested within those obtained with the IV procedure of Hausman-Taylor. For instance, from Table 4, the average over 1,000 replications of the 95% confidence intervals for η_2 are:

95% confidence intervals for η_2		3S		IV HT	
		min	max	min	max
$N = 100, T = 5$	$\rho = 0.3$	0.965	1.096	0.740	1.249
	$\rho = 0.8$	0.984	1.107	0.643	1.356
$N = 100, T = 10$	$\rho = 0.3$	0.971	1.070	0.838	1.156
	$\rho = 0.8$	0.983	1.076	0.721	1.296

The HT procedure is known to generate large confidence intervals for all the coefficients of the time-invariant variables. Despite the fact that our 3S method leads to a slight bias for the coefficient η_2 of the time-invariant variable Z_2 (correlated with μ_i), the uncertainty about the permissible values is significantly reduced compared to HT.

While the biases are similar to those of HT, the RMSE are much smaller (for instance, the RMSE of η_2 is three times smaller when $N = 100$, $T = 10$ and $\rho = 0.8$). Whereas 3S and HT fit the remainder variance rather well (σ^2_ε), 3S tends to slightly over-estimate the individual effects variance (σ^2_μ) when ρ is small, and under-estimate it when ρ is large.

When $\rho = 0.3$, the bias of σ^2_μ for HT declines from 17.05% to 4.57% when T doubles. The comparable decline for 3S is from 36.40% to 15.04% when T doubles. This bias shrinks considerably when $\rho = 0.8$. In fact, for HT, there is a reduction in the bias from 8.16% to 2.75% when T doubles. The comparable reduction in the bias for 3S is from -0.57% to 2.49%. These results continue to hold when N increases (see Table A8 in the Appendix).

Just like the Mundlak and Chamberlain-type FE worlds, the weights λ_β and λ_b indicate that there is no trade-off between the Bayes estimators and the empirical Bayes estimators. Only the empirical Bayes estimators are used in the estimation of the slope coefficients and the individual effects. For a typical macro-panel (N small, T large), all these results carry through, and in some cases are even improved (see Table A9 in the Appendix for $N = 50$, $T = 20$, $\rho = 0.8$ with $\varepsilon = 0.5$). We see that the bias on the coefficient η_2 of the time-invariant variable Z_2 (correlated with μ_i) is reduced to 1.5%, and more importantly the standard errors are 8 times smaller than those obtained for the IV procedure of Hausman-Taylor. Once again the 95% confidence interval for η_2 obtained with 3S $[0.966; 1.065]$ is smaller and nested within that obtained for the HT estimator $[0.606; 1.397]$. Moreover, the bias of σ^2_μ for HT is larger (3.06%) than that for 3S (-0.52%).

To investigate the properties of our proposed strategy, we computed the biases ($\eta_2 - \widehat{\eta}_{2,3S}$) under $s = 1, 2, 3$. Figures 1 and 2 in the Appendix plot the ratios of the biases for $s = 2, 3$ relative to the bias for $s = 1$ for different sample sizes. The figures confirm that when the correlation between μ_i and Z_{2i} is more than 20%, it is best to use $s = 1$ to reduce the bias. Whereas when the correlation between μ_i and Z_{2i} is less than 20%, it is best to use $s = 2, 3$ to reduce the bias.

5.3 Sensitivity to ε and non-normality

As a final check on the properties of our proposed 3S estimator, we conducted two additional sets of experiments. First, we checked the sensitivity of our results to changing the values of ε, the contamination part of prior distributions. We allowed ε to vary between 10% and 90%. Only the results for the RE world and the Hausman-Taylor world ($N = 100, T = 5, \rho = 0.8$) are reported in Tables A10 and A11 in the Appendix. For the RE and HT worlds, this does not change the estimated slope coefficients, standard errors, biases or RMSE of the coefficients. It also does not change the estimated values of the remainder variances (σ^2_ε). Only for HT do we observe some differences in the variances of the individual effects (σ^2_μ). The closer we are to the intermediate values ($\varepsilon = 0.3, 0.7$), the more important is the bias (-2.5%). For extreme values ($\varepsilon = 0.1$ or 0.9), the bias is smaller (-1.75%). For $\varepsilon = 0.5$, we get the smallest bias (-0.75%). Moving from $\varepsilon = 0.1$ to $\varepsilon = 0.9$ leads to a W shape for the bias on σ^2_μ.

Last, but not least, we checked the sensitivity of various estimators to a non-normal framework. The remainder disturbances (ε_{it}) were assumed to follow a right-skewed t-distribution $ST(0, df = 3, shape = 2)$ (see Fernández and Steel (1998)) instead of the $N(0, 1)$ (see equation (53)). Results in Tables A12-A14 in the Appendix show that, irrespective of the estimator considered, our 3S significantly dominates in terms of bias and precision of the slope parameters for RE, Chamberlain-type fixed effects and Hausman-Taylor worlds. In addition, the estimated variances of the individual effects and remainder terms are much closer to the true theoretical values compared to the classical estimators. What is remarkable is that our 3S estimator remains unbiased and has very small standard errors relative to the classic estimators. It also yields variances of the individual effects σ^2_μ and remainder terms σ^2_ε that are very similar to the theoretical ones. For example, for the Hausman-Taylor world, Table A14 in the Appendix shows that the bias of our 3S estimator for η_2 is -0.38%, while that for HT is 1.02%. But most surprising, the 95% confidence interval of η_2 is very narrow [0.8447; 1.1628] as compared to the wide 95% confidence interval [0.3035; 1.6761] for HT. The estimates of σ_ε^2 of our 3S estimator (7.221) and the HT estimator (7.211) are close to the theoretical variance $\left(\sigma_\varepsilon^2 = 7.227\right)$. However, this is not the case for the estimated individual effects variance σ_μ^2. Our 3S estimator (4.571) is relatively closer to the theoretical value ($\sigma_\mu^2 = 4$) as compared to that of HT estimator (5.558). Last but not least, λ_β is small but slightly more important than that for the Gaussian cases.

6 Applications

6.1 The Cornwell-Rupert earnings equations

Cornwell and Rupert (1988) estimate a returns to schooling example based on a panel of 595 individuals observed over the period 1976 − 82 and drawn from the Panel Study of Income Dynamics (PSID). In particular, log wage is regressed on years of education (ED), weeks worked (WKS), years of full-time work experience (EXP), occupation (OCC=1, if the individual is in a blue-collar occupation), residence (SOUTH = 1, SMSA = 1, if the individual resides in the South, or in a standard metropolitan statistical area), industry (IND = 1, if the individual works in a manufacturing industry), marital status (MS = 1, if the individual is married), sex and race (FEM = 1, BLK = 1, if the individual is female or black), union coverage (UNION = 1, if the individual's wage is set by a union contract) (see also Baltagi and Khanti-Akom (1990)). We let $X1 = (OCC, SOUTH, SMSA, IND)$, $X2 = (EXP, EXP2, WKS, MS, UNION)$, $Z1 = (FEM, BLK)$ and $Z2 = ED$. For the Mundlak estimation, we drop $Z1$ and $Z2$ and we consider that only the variables in $X2$ are correlated with the individual effects.

The estimation results are reported in Table 5. There are very few differences between the Within, the FGLS estimates on the transformed model (*i.e.*, the Mundlak-type FE) and our 3S estimator. Since we assume that only the $X2$ variables are correlated with the individual effects, Within estimates do not exactly match the Mundlak-type FE. One can note that the FE estimates are slightly different from those of the two other methods (Mundlak-type FE and 3S), especially for OCC, $SOUTH$ and $SMSA$. But for the main variables of the earnings equation, we get similar results. A comparison between the Mundlak-type FE and 3S shows that the estimate of IND becomes significantly different from zero. With the three-stage robust Bayesian estimator, we get more precise estimates of all coefficients. Estimation of the π values from the 3S and the FGLS on the transformed model are quite similar. The estimated variances of the individual effects (σ_μ^2) and the residuals (σ_ε^2) are roughly the same for 3S and Mundlak-type fixed effects.

For the HT model, we need to reintroduce $Z1$ and $Z2$ into the model. The assumption that there is correlation between the individual effects and the explanatory variables $X2$ and $Z2$ justifies the use of the IV method with instruments given by $A_{HT} = [Q_X X1, Q_W X2, PX1, Z1]$ where $Q_W = I_{NT} - P$ is the within-transform. To choose the s parameter of our function $f\left[.\right] = \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right)^2 \odot \left(Z_{2i} - E_{Z_2}\right)^s$, we estimate the OLS proxy of the individual effects $\widehat{\mu} = \left(Z_\mu' Z_\mu\right)^{-1} Z_\mu' \widehat{y}$ where $\widehat{y}$ are the fitted values of the pooling regression $y = X_1\beta_1 + X_2\beta_2 + Z_1\eta_1 + Z_2\eta$ and then compute the correlation between $\widehat{\mu}$ and Z_2. The estimated correlation is large (0.612), so we set $s = 1$. From the estimates reported in Table 6, we see little differences between the HT and our 3S estimators. Of course, the RE estimates are biased but they are presented here for the sake of comparison with the HT and 3S estimates. The three-stage robust Bayesian estimator leads to more precise and significant coefficients compared to those of the IV estimator, except for $SMSA$. With the IV method, we get non significant effects for OCC, $SOUTH$ and IND and a surprising negative effect for $SMSA$. In contrast, with the 3S estimator, OCC and $SOUTH$ have the expected negative effects. IND has an expected positive effect but $SMSA$ has a non significant effect. With 3S, gender and race effects are now significant and the plausible negative gender impact dominates the negative race effect. If EXP and $EXP2$ have the same impact in both the 3S and HT, the effect of ED is slightly lower (11.43% against 13.79%).[13] More interestingly, the 95% confidence interval of ED is narrow [11.02%; 11.83%] as compared to the one obtained with the IV method [9.63%; 17.56%]. This sizeable difference between the standard errors of 3S and those of the IV method are expected from our Monte Carlo study. However, there is no statistical difference between these two estimates, since the confidence interval of ED for IV nests the one for 3S. From an economic policy point of view, though, the effect of education on earnings is better estimated with 3S than with IV. It is difficult to imagine an economic adviser telling a policy-maker that the returns to schooling effects can vary between 9% and 17%. Yet, one may wonder whether the average education effect estimated with 3S may be under-estimated, the difference being less than 2.4%.[14]

6.2 The Cornwell-Trumbull crime model

Cornwell and Trumbull (1994) estimated an economic model of crime using panel data on 90 counties in North Carolina over the period 1981 − 1987. The empirical model relates the crime rate to a set of explanatory variables which include deterrent variables as well as variables measuring returns to legal opportunities. All variables are in logs except for the regional dummies (*west*, *central*). The explanatory variables include the probability of arrest P_A , the probability of conviction given arrest P_C, the probability of a prison sentence given a conviction P_P, the number of policemen per capita as a measure of the county's ability to detect crime (*Police*), the population density, (*Density*), percent minority (*pctmin*), regional dummies for western and central counties. Opportunities in the legal sector are captured by the average weekly wage in the county by industry. These industries are: transportation, utilities and communication (*wtuc*); manufacturing ($wmfg$).

From Table 7, there is not much difference between the MCS and 3S estimates on the transformed model.[15] All the confidence intervals for MCS and 3S estimates overlap. Estimation of the π_t coefficients obtained from 3S lead to more statistically significant coefficients than those from MCS. We only report coefficients that are statistically significant at the 5% level. But, more interestingly, we note a strong coherency between the Within estimates (FE) and 3S. The MCS estimates are slightly different from the FE estimates, with, for example, P_C having an estimate of −0.23 for MCS as compared to −0.31 for FE and 3S. Note, however, that the 95% confidence intervals overlap with each other. The estimated variances of the individual effects (σ^2_μ) and the residuals (σ^2_ε) are sighltly different (0.05 for 3S and 0.07 for MCS).

We also estimated a Mundlak-type FE model. Table 8 reveals slight differences between the Within, the Mundlak-type and 3S estimation results. The most notable differences concern the dummies, the π values and the standard errors between 3S and Mundlak-type FE. Estimation of all the coefficients by 3S are more precise than those of the FGLS

[13] Baltagi and Bresson (2012), using a robust HT estimator, show that the returns to education are roughly the same but the gender effect becomes significant and the race effect becomes smaller as compared to those obtained in the classical HT case.

[14] Recall that the coefficient of the endogeneous time-invariant variable was slightly biased (1.7% to 5%) in the simulation study for 3S, even though the RMSE for 3S, was lower than that for HT.

[15] Results are obtained using one hundred iterations on the MCS estimator to match the results of Baltagi *et al.* (2009).

on the transformed model for Mundlak-type FE. But, more interestingly, we note once again a strong coherency between the Within estimates (FE), the 3S and the FGLS on the transformed model. Just as with the MCS estimation, the estimated variances of the individual effects (σ_μ^2) and of the residuals (σ_ε^2) are roughly the same (0.03 for Mundlak-type FE and 0.04 for 3S).

7 Conclusion

To our knowledge, our paper is the first to analyze the static linear panel data model using an ε-contamination approach with two-stage and three-stage hierarchies. 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 have shown that our approach encompasses a variety of specifications such as random effects, Hausman-Taylor, Mundlak, and Chamberlain-type models. The frequentist approach, on the other hand, requires separate estimators for each model.

Following Singh and Chaturvedi (2012), we estimate the Type II maximum likelihood (ML-II) posterior distribution of the coefficients, β, and the individual effects, b, using a two-step procedure. Indeed, we first subtract Wb from y and derive the ML-II posterior distribution of β given b and g_0 (the Zellner's g-prior of the base elicited prior of the variance-covariance matrix of β). It turns out that the ML-II posterior density of β is a weighted average of the conditional posterior density of β based upon the base prior and the conditional posterior density of β based on the ε-contaminated prior. We show that each conditional posterior density of β is the pdf of a multivariate t-distribution which also depends on both the Bayes estimator of β under the base prior g_0 and the data-dependent empirical Bayes estimator of β. If the base prior is consistent with the data, the ML-II posterior density of β 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 of β is then close to the conditional posterior density derived from the ε-contaminated prior. Moreover, we derive the ML-II posterior mean and variance-covariance matrix of β given b and g_0. In the second step, we subtract $X\beta$ from y, and again derive the ML-II posterior distribution of b given β and h_0 (the Zellner's g-prior of the base elicited prior for the variance-covariance matrix of b). Similar conclusions as in the first step obtain. These derivations are useful in that they show how the shrinkage estimators arise. They are also useful in that they avoid having to draw thousands of multivariate t-distributions in order to compute the means and the variances after burning draws. Our approach only requires a weighted average of the Bayes and the empirical Bayes estimators and is relatively easy to implement.

Our approach is derived both for a two-stage and a three-stage hierarchy model. As stressed in the literature, the Bayesian approach introduces a third stage in the hierarchical model in order to discriminate between *fixed effects* and *random effects.* In general, and more specifically in the context of panel data, hyperparameters are used only to model the variance-covariance matrix of the individual effects b. We go one-step further and use Zellner's g-priors in the second stage on β *and* b, assuming g_0 is fixed and known. We need only define mixtures of g-priors on the precision matrix of b, or equivalently on h_0. For this purpose, we use a generalized hyper-g prior which is a Beta-prime (or Pearson Type VI) hyper prior for h_0 and we restrict the ε-contamination class of prior distributions to the second stage of the hierarchy only (*i.e.*, for (β, b)). The expression of the ML-II posterior density of b is little affected by this specification. On the other hand, the predictive densities of the Bayes estimator, the empirical Bayes estimator and the weights are now Gaussian and Appell hypergeometric functions for the estimators based on the base elicited prior and on the ε-contaminated prior, respectively. These functions are known to generate overflows for moderate to large samples. This is likely to be problematic for microeconomic panel data. However, we could use Laplace approximations of the integrals to circumvent this difficulty.

The finite sample performance of the two-stage and three-stage hierarchy estimators are investigated using Monte Carlo experiments. The experimental design includes a random effects world, a Mundlak-type world, a Chamberlain-type world and a Hausman-Taylor-type world. Using unit information prior in the two-stage hierarchy for the Zellner's g-priors g_0 and h_0 and a Beta-prime distribution in the three-stage hierarchy for h_0, our simulation results underscore the relatively superior performance of the three-stage hierarchy estimator, irrespective of the data generating process considered. Indeed, estimated $\beta's$ and $b's$ are always very close to their true values. Moreover, their biases and RMSE are close and often smaller than those of the conventional estimators. In the two-stage hierarchy, estimated weights show that 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 checked the sensitivity of our results to the values of ε, the contamination part of the prior distributions. Our results are very robust, even when ε is allowed to vary between 10% and 90%. Lastly, we have shown that our 3S estimators are significantly better behaved than the classical estimators when the remainder disturbances are not normally distributed.

The major conclusions from the Monte Carlo experiments is that our Bayesian approach, which encompasses a variety of specifications, leads to similar and often better performance than that obtained by conventional methods. The simulation results also hold in the empirical examples using panel data from earnings and crime. Analyses of earnings data (Within, Mundlak-type world, Hausman-Taylor-type world) and crime data (Within, Chamberlain-type world) show that our approach yields very similar results to those of conventional estimators (Feasible GLS,Within, Minimum Chi Square, Instrumental Variables), but often times outperforms them in the sense of being statistically more precise and definitely more robust.

The main originality of this paper lies in the application of the ε-contamination class to the linear static panel data model. The framework we develop is very general and encompasses various specifications. 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. and I.A. Stegun, 1970, *Handbook of Mathematical Functions*, Dover Publications, Inc. New York.

Appell, P., 1882, Sur les fonctions hypergéométriques de deux variables, *Journal de Mathématiques Pures et Appliquées,* 3ème série, (in French) , 8, 173–216. http://portail.mathdoc.fr/JMPA/PDF/JMPA_1882_3_8_A8_0.pdf.

Baltagi, B.H., 2013, *Econometric Analysis of Panel Data*, fifth edition, Wiley, Chichester, UK.

Baltagi, B.H. and G. Bresson, 2012, A robust Hausman-Taylor estimator, in *Advances in Econometrics: Essays in Honor of Jerry Hausman*, vol. 29, (Baltagi B.H., Carter Hill, R. Newey, W.K. and H.L. White, eds.), Emerald Group Publishing Limited, 175-214.

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

Baltagi, B.H., Bresson, G. and A. Pirotte, 2008, To pool or not to pool?, in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, (Mátyás, L. and P. Sevestre, eds.), Chap. 16, Advanced Studies in Theoretical and Applied Econometrics, Springer, Amsterdam, 517-546.

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

Baltagi, B.H. and S. Khanti-Akom, 1990, On efficient estimation with panel data: an empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics*, 5, 401-06.

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

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

Berger, J. and M. Berliner, 1986, Robust Bayes and empirical Bayes analysis with ε-contaminated priors, *Annals of Statistics*, 14, 2, 461-486.

Bresson, G. and C. Hsiao, 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, 4, 501-529.

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.

Chib, S., 2008, Panel data modeling and inference: a Bayesian primer, in *The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice* (Matyas, L. and P. Sevestre, eds), Springer, 479-516.

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

Cornwell, C. and P. Rupert, 1988, Efficient estimation with panel data: an empirical comparison of instrumental variables estimators, *Journal of Applied Econometrics*, 3, 149-155.

Cornwell, C. and W.N. Trumbull, 1994, Estimating the economic model of crime with panel data, *Review of Economic and Statistics*, 76, 360-366.

Cui, W. and E.I. George, 2008, Empirical Bayes vs. fully Bayes variable selection, *Journal of Statistical Planning and Inference*, 138, 888-900.

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

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

Fernandez, C., Ley, E. and M.F.J. Steel, 2001, Benchmark priors for Bayesian model averaging, *Journal of Econometrics*, 100, 381-427.

Foster, D. P. and E.I. George, 1994, The risk inflation criterion for multiple regression, *The Annals of Statistics*, 22, 1947–1975.

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. and W.E. Taylor, 1981, Panel data and unobservable individual effects, *Econometrica*, 49, 1377-1398.

Hsiao, C., 2003, *Analysis of Panel Data*, second edition, Cambridge University Press, Cambridge, MA.

Hsiao, C. and M.H. Pesaran, 2008, Random coefficient models, in *The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice* (Matyas, L. and P. Sevestre, eds), Springer, 185-214.

Kass, R. E. and L. Wasserman, 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.

Liang, F., Paulo, R., Molina, G., Clyde, M.A. and J.O. Berger, 2008, Mixtures of g priors for Bayesian variable selection, *Journal of the American Statistical Association*, 103, 481, 410-423.

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

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

Maddala, G.S., Trost, R.P., Li, H. and F. Joutz, 1997, Estimation of short-run and long-run elasticities of energy demand from panel using shrinkage estimators, *Journal of Business and Economic Statistics*, 15, 90-100.

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

Maruyama, Y. and E.I. George, 2014, Posterior odds with a generalized hyper-g prior, *Econometric Reviews*, 33, 1-4, 251-269.

Moreno, E. and L.R. Pericchi, 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 and B.P Flannery, 2007, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 3rd ed., New York.

Rendon, S.R., 2013, Fixed and random effects in classical and Bayesian regression, *Oxford Bulletin of Economics and Statistics,* 75, 3, 460-476.

Serlenga, L. and Y. Shin, 2007, Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors, *Journal of Applied Econometrics*, 22, 361-381.

Singh, A. and A. Chaturvedi, 2012, Robust Bayesian analysis of autoregressive fixed effects panel data model, *Department of Statistics, University of Allahabad*, working paper.

Sinha, P. and P. Jayaraman, 2010a, Robustness of Bayes decisions for normal and lognormal distributions under hierarchical priors, *Faculty of Management Studies, University of Delhi*, working paper MPRA n° 22416.

Sinha, P. and P. Jayaraman, 2010b, Bayes reliability measures of lognormal and inverse Gaussian distributions under ML-II $\varepsilon-$contaminated class of prior distributions, *Defence Science Journal*, 60, 4, 442-450.

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.

Strawderman, W. E., 1971, Proper Bayes minimax estimators of the multivariate normal mean, *The Annals of Mathematical Statistics*, 42, 385–388.

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*, (Goel, P.K., Zellner, A. and B. de Finetti, eds), Studies in Bayesian Econometrics, vol. 6, North-Holland, Amsterdam, 389–399.

Zellner, A., and A. Siow, 1980, Posterior odds ratios for selected regression hypotheses, in *Bayesian Statistics: Proceedings of the First International Meeting*, (Bernardo, M.H., DeGroot, J.M., Lindley, D.V. and A.F.M. Smith, eds), University of Valencia Press, 585–603.

Zheng, Y., Zhu, J. and D. Li, 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 (RE, Robust two-stage and Robust three-stage) , $N = 100, T = 5, \varepsilon = 0.5$

			β_{11}	β_{12}	β_2	σ_ε^2	σ_μ^2	λ_β	λ_b
		true	1	1	1	1	0.428571		
$\rho = 0.3$	RE	coef	0.999413	1.000269	1.000614	1.003209	0.439283		
		se	0.018034	0.017995	0.017987				
		rmse	0.018490	0.018312	0.018250				
	2S	coef	0.998982	0.998718	0.996877	0.996431	0.524295	0.281039	0.498841
		se	0.009711	0.009685	0.009686				
		rmse	0.043036	0.043164	0.043901				
	3S	coef	0.995925	0.996070	0.996831	0.996400	0.441509	0.224000	$< 10^{-6}$
		se	0.009802	0.009802	0.009821				
		rmse	0.025551	0.025923	0.025853				
		true	1	1	1	1	4		
$\rho = 0.8$	RE	coef	0.999333	0.999454	0.999220	1.004497	3.943506		
		se	0.033463	0.033530	0.033488				
		rmse	0.033543	0.032842	0.034219				
	2S	coef	0.996988	0.998563	0.997820	0.997612	4.062508	0.279963	0.493310
		se	0.009667	0.009695	0.009685				
		rmse	0.041191	0.042294	0.043960				
	3S	coef	0.996486	0.996962	0.997768	0.995032	4.000670	0.116000	$< 10^{-6}$
		se	0.009746	0.009725	0.009743				
		rmse	0.036711	0.036959	0.037201				

Table 2: Mundlak-type FE (Mundlak and Robust three-stage), $\varepsilon = 0.5$

			β_{11}	β_{12}	β_2	π	σ^2_ε	σ^2_μ	λ_β	λ_b
		true	1	1	1	0.8	1	6.332799		
$N = 100, T = 5$	Mundlak	coef	1.000379	1.000026	1.000225	0.798594	0.997845	6.360992		
		se	0.023233	0.023295	0.035373	0.047462				
		rmse	0.022653	0.023244	0.036145	0.047717				
	3S	coef	0.999247	0.998581	1.000122	0.802392	0.991534	6.410207	$< 10^{-6}$	$< 10^{-6}$
		se	0.009659	0.009699	0.031597	0.033642				
		rmse	0.027365	0.028673	0.036099	0.047561				
		true	1	1	1	0.8	1	6.194526		
$N = 100, T = 10$	Mundlak	coef	0.999654	1.000224	0.999984	0.800496	0.997840	6.196425		
		se	0.018342	0.018387	0.023591	0.038385				
		rmse	0.019218	0.018609	0.023322	0.038276				
	3S	coef	0.999422	0.999832	0.999927	0.802940	0.994733	6.283714	$< 10^{-6}$	$< 10^{-6}$
		se	0.007244	0.007280	0.022357	0.023997				
		rmse	0.021718	0.021130	0.023307	0.038166				
		true	1	1	1	0.8	1	6.376910		
$N = 500, T = 5$	Mundlak	coef	1.000264	1.000447	0.999123	0.801186	1.001408	6.382758		
		se	0.010366	0.010369	0.015818	0.021108				
		rmse	0.010442	0.010196	0.015275	0.020788				
	3S	coef	1.000181	1.000301	0.999098	0.801989	0.998821	6.39379	$< 10^{-6}$	$< 10^{-6}$
		se	0.004307	0.004306	0.014157	0.015052				
		rmse	0.013118	0.012843	0.015281	0.020801				
		true	1	1	1	0.8	1	6.229901		
$N = 500, T = 10$	Mundlak	coef	0.999933	0.999596	0.999238	0.800899	0.999333	6.236728		
		se	0.008201	0.008199	0.010544	0.017104				
		rmse	0.008038	0.008018	0.010771	0.017407				
	3S	coef	0.999977	0.999577	0.999225	0.801419	0.999103	6.249716	$< 10^{-6}$	$< 10^{-6}$
		se	0.003221	0.003221	0.010004	0.010723				
		rmse	0.008970	0.009299	0.010770	0.017458				

Table 3: Chamberlain-type fixed effects world (MCS and Robust three-stage), $\varepsilon = 0.5$

$N = 100$	$T = 5$	β_{11}	β_{12}	β_2	π_1	π_2	π_3	π_4	π_5	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	0.4096	0.512	0.64	0.8	1	1	95.395060		
MCS	coef	0.999183	0.999338	0.998899	0.410068	0.512711	0.640541	0.798805	0.999782	0.930255	95.486405		
	se	0.011398	0.011444	0.015461	0.030681	0.030595	0.030668	0.030623	0.030637				
	rmse	0.026642	0.028338	0.038574	0.073181	0.072649	0.072438	0.074425	0.073688				
3S	coef	0.994889	0.994874	0.998748	0.421159	0.520266	0.642413	0.795123	0.988425	0.992827	95.763309	$< 10^{-6}$	$< 10^{-6}$
	se	0.009897	0.009936	0.031725	0.027224	0.027155	0.027204	0.027173	0.027180				
	rmse	0.029051	0.029593	0.034296	0.068681	0.069719	0.068357	0.070054	0.070790				

$N = 500$	$T = 5$	β_{11}	β_{12}	β_2	π_1	π_2	π_3	π_4	π_5	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	0.4096	0.512	0.64	0.8	1	1	96.527183		
MCS	coef	0.999953	1.000162	0.999901	0.410658	0.513867	0.639844	0.798942	0.998895	1.033564	90.106810		
	se	0.005217	0.015505	0.005226	0.007034	0.013921	0.013928	0.013942	0.013940				
	rmse	0.011064	0.011324	0.015999	0.031011	0.030940	0.030951	0.031271	0.032062				
3S	coef	0.999155	0.998555	0.999683	0.413109	0.513152	0.638680	0.799302	0.998823	0.999089	96.108612	$< 10^{-6}$	$< 10^{-6}$
	se	0.004319	0.004318	0.014169	0.011788	0.011802	0.011805	0.011819	0.011821				
	rmse	0.012777	0.013279	0.015534	0.031473	0.032603	0.031234	0.033800	0.030361				

Table 4: Hausman-Taylor world (HT and Robust three-stage) , $\varepsilon = 0.5$

			β_{11}	β_{12}	β_2	η_1	η_2	σ_ε^2	σ_μ^2	λ_β	λ_b
$N = 100, T = 5$		true	1	1	1	1	1	1	0.428571		
	HT	coef	1.001485	1.001458	0.999838	1.000812	0.994872	0.991842	0.5035783		
		se	0.034189	0.034186	0.044596	0.086291	0.129795				
$\rho = 0.3$		rmse	0.034353	0.033172	0.045580	0.084166	0.126294				
	3S	coef	0.993741	0.993366	0.999001	0.996682	1.030501	0.993270	0.584580	$< 10^{-6}$	$< 10^{-6}$
		se	0.012522	0.012518	0.040304	0.042410	0.033576				
		rmse	0.026884	0.026289	0.045551	0.081499	0.065582				
		true	1	1	1	1	1	1	4		
	HT	coef	0.998016	0.998633	1.003067	0.993534	0.999907	0.992364	4.326409		
		se	0.039375	0.039461	0.044472	0.213941	0.182014				
$\rho = 0.8$		rmse	0.040585	0.040655	0.045269	0.210925	0.179451				
	3S	coef	0.991538	0.990087	1.001559	1.002670	1.047291	0.992645	3.976909	$< 10^{-6}$	$< 10^{-6}$
		se	0.012486	0.012499	0.040205	0.048986	0.031392				
		rmse	0.028299	0.028394	0.043379	0.043379	0.073468				
$N = 100, T = 10$		true	1	1	1	1	1	1	0.428571		
	HT	coef	0.999568	1.000147	0.999426	0.998493	0.997749	0.997583	0.445893		
		se	0.021031	0.021081	0.025618	0.075291	0.081110				
$\rho = 0.3$		rmse	0.021259	0.020867	0.025860	0.074090	0.078393				
	3S	coef	0.995979	0.995975	0.998673	0.996635	1.021223	0.997971	0.493050	$< 10^{-6}$	$< 10^{-6}$
		se	0.009402	0.009398	0.024399	0.031352	0.025324				
		rmse	0.018592	0.018068	0.025882	0.072129	0.044547				
		true	1	1	1	1	1	1	4		
	HT	coef	0.998635	0.998664	1.001816	1.002543	1.008771	0.997156	4.110078		
		se	0.024247	0.024228	0.025585	0.203937	0.146772				
$\rho = 0.8$		rmse	0.024111	0.023718	0.025604	0.195657	0.148177				
	3S	coef	0.995587	0.995311	1.000044	1.002547	1.029726	0.997562	3.901885	$< 10^{-6}$	$< 10^{-6}$
		se	0.009383	0.009388	0.024393	0.034524	0.023696				
		rmse	0.018168	0.017625	0.025217	0.190068	0.047691				

Table 5: Earnings equation - Within, Mundlak-type FE and Robust three-stage , N=595, T=7

	Within			Mundlak-type FE			Robust three-stage		
	Coef.	se	t-stat	Coef.	se	t-stat	Coef.	se	t-stat
occ	-0.021476	0.013784	-1.558110	-0.057646	0.013340	-4.321237	-0.041358	0.005021	-8.236431
south	-0.001861	0.034299	-0.054263	-0.059623	0.027301	-2.183962	-0.077931	0.004965	-15.696192
smsa	-0.042469	0.019428	-2.185936	0.017762	0.017865	0.994205	-0.000804	0.004827	-0.166487
ind	0.019210	0.015446	1.243670	0.005401	0.014734	0.366532	0.011196	0.004714	2.374961
exp	0.113208	0.002471	45.814092	0.113139	0.002501	45.231610	0.113189	0.002289	49.440550
exp^2	-0.000418	0.000055	-7.662880	-0.000416	0.000055	-7.519089	-0.000417	0.000051	-8.231529
wks	0.000836	0.000600	1.394009	0.000887	0.000607	1.461538	0.000856	0.000556	1.538495
ms	-0.029726	0.018984	-1.565873	-0.034033	0.019205	-1.772086	-0.033505	0.017567	-1.907311
union	0.032785	0.014923	2.196954	0.037554	0.015095	2.487821	0.035553	0.013745	2.586638
π_{exp}				-0.050872	0.008681	-5.860344	-0.047229	0.002468	-19.140152
π_{exp^2}				-0.000780	0.000192	-4.063800	-0.000843	0.000055	-15.448131
π_{wks}				0.121233	0.001914	63.346793	0.120230	0.000598	201.053552
π_{ms}				0.353727	0.058888	6.006833	0.368338	0.018659	19.740215
π_{union}				0.201914	0.046582	4.334567	0.184044	0.014673	12.543212
${\sigma^2}_{\epsilon}$	0.023044			0.023102			0.023120		
${\sigma^2}_{\mu_i}$	1.068764			1.023089			1.024564		

Table 6: Earnings equation - RE, Hausman-Taylor and Robust three-stage , N=595, T=7

	RE			Hausman-Taylor			Robust three-stage		
	Coef.	se	t-stat	Coef.	se	t-stat	Coef.	se	t-stat
occ	-0.050066	0.016647	-3.007550	-0.020705	0.013781	-1.502414	-0.031173	0.005995	-5.199445
south	-0.016618	0.026527	-0.626452	0.007440	0.031955	0.232822	-0.043300	0.005100	-8.489817
smsa	-0.013823	0.019993	-0.691405	-0.041833	0.018958	-2.206619	-0.000612	0.004906	-0.124775
ind	0.003744	0.017262	0.216904	0.013604	0.015237	0.892800	0.020606	0.004827	4.268827
exp	0.082054	0.002848	28.813762	0.113133	0.002471	45.785055	0.113273	0.002291	49.441092
exp^2	-0.000808	0.000063	-12.868578	-0.000419	0.000055	-7.671785	-0.000418	0.000051	-8.246418
wks	0.001035	0.000773	1.337866	0.000837	0.000600	1.396292	0.000840	0.000556	1.510123
ms	-0.074628	0.023005	-3.243971	-0.029851	0.018980	-1.572752	-0.033093	0.017578	-1.882644
union	0.063223	0.017070	3.703763	0.032771	0.014908	2.198181	0.033645	0.013761	2.444885
intercept	4.263670	0.097716	43.633217	2.912726	0.283652	10.268653	3.188821	0.048464	65.797345
fem	-0.339210	0.051303	-6.611855	-0.130924	0.126659	-1.033670	-0.275260	0.010881	-25.297219
blk	-0.210280	0.057989	-3.626221	-0.285748	0.155702	-1.835225	-0.063830	0.009139	-6.984141
ed	0.099659	0.005747	17.339476	0.137944	0.021248	6.491942	0.114338	0.002067	55.326848
${\sigma^2}_\epsilon$	0.023102			0.023044			0.023102		
${\sigma^2}_{\mu_i}$	0.068989			0.886993			0.903661		

Table 7: Crime model - Within, Chamberlain and Robust three-stage , N=90, T=7

	Within			Chamberlain MCS			Robust three-stage		
	Coef.	se	t-stat	Coef.	se	t-stat	Coef.	se	t-stat
Intercept				-5.444113	0.915646	-5.945651	-5.127992	0.466941	-10.982092
pctmin				0.204222	0.035669	5.725551	0.220576	0.018189	12.126575
west				-0.197924	0.090929	-2.176684	-0.178019	0.046370	-3.839087
central				-0.051538	0.045829	-1.124572	-0.039906	0.023371	-1.707533
P_A	-0.394180	0.032783	-12.023892	-0.341458	0.027587	-12.377321	-0.393988	0.032303	-12.196512
P_C	-0.310792	0.021435	-14.499532	-0.229714	0.016540	-13.888003	-0.310662	0.021121	-14.708738
P_P	-0.204072	0.032714	-6.238011	-0.180867	0.023683	-7.636995	-0.204023	0.032236	-6.329127
Police	0.420279	0.027045	15.539840	0.305634	0.026383	11.584585	0.419859	0.026650	15.754847
Density	0.491696	0.274325	1.792386	0.310666	0.313637	0.990528	0.491222	0.270311	1.817250
wtuc	0.025904	0.017872	1.449361	0.012717	0.011270	1.128406	0.025780	0.017611	1.463855
wmfg	-0.336215	0.064678	-5.198283	-0.233752	0.065701	-3.557796	-0.336067	0.063732	-5.273153
${\sigma^2}_{\epsilon}$	0.020159			0.0211187			0.020159		
${\sigma^2}_{\mu_i}$	0.129723			0.0492775			0.070899		
total	0.149882			0.0703963			0.091782		

π_t MCS	P_A	P_C	P_P	Police	Density	wtuc	wmfg
1981		0.133646			-7.008780		-1.963856
1982			-0.233336		18.242164	-0.097704	
1983		-0.251424	0.249742	-0.499259	-8.282196		2.284093
1984		0.376922		-0.301084			
1985	-0.518998	-0.149456		0.530658			
1986		-0.349304				1.218455	0.633301
1987	0.206110				6.230701	-1.162895	-0.840497

π_t Rob.	P_A	P_C	P_P	Police	Density	wtuc	wmfg
1981	0.117983	0.070619		-0.149993	-10.247563	0.190783	-1.159108
1982			-0.141819		24.772614	-0.180327	
1983		-0.130959	0.278137	-0.521681	-10.463678		1.486814
1984		0.240189		-0.206941	-4.813363	0.529217	
1985	-0.542961	-0.140629	-0.144892	0.681189			
1986		-0.212280			-4.057384	0.933911	
1987	0.170253	-0.113766			6.256856	-1.078718	-0.724185

Table 8: Crime model - Within, Mundlak-type FE and Robust three-stage , N=90, T=7

	Within			Mundlak-type FE			Robust three-stage		
	Coef.	se	t-stat	Coef.	se	t-stat	Coef.	se	t-stat
pctmin				0.129638	0.041172	3.148691	0.163177	0.009472	17.227065
west				-0.243537	0.101786	-2.392648	-0.187871	0.023417	-8.022858
central				-0.126991	0.061604	-2.061417	-0.030875	0.014173	-2.178498
P_A	-0.394180	0.032783	-12.023892	-0.394180	0.032783	-12.023892	-0.394185	0.030628	-12.870217
P_C	-0.310792	0.021435	-14.499532	-0.310792	0.021435	-14.499532	-0.310819	0.020025	-15.521278
P_P	-0.204072	0.032714	-6.238011	-0.204072	0.032714	-6.238011	-0.204128	0.030563	-6.678830
Police	0.420279	0.027045	15.539840	0.420279	0.027045	15.539840	0.420056	0.025267	16.624600
Density	0.491696	0.274325	1.792386	0.491696	0.274325	1.792386	0.491453	0.256289	1.917577
wtuc	0.025904	0.017872	1.449361	0.025904	0.017872	1.449361	0.025786	0.016697	1.544321
wmfg	-0.336215	0.064678	-5.198283	-0.336215	0.064678	-5.198283	-0.336235	0.060426	-5.564444
π_{P_A}				-0.226198	0.087865	-2.574373	-0.259431	0.035914	-7.223735
π_{P_C}				-0.226241	0.065209	-3.469495	-0.245583	0.024531	-10.011252
π_{P_P}				0.710404	0.202301	3.511612	0.724378	0.055169	13.130177
π_{Police}				-0.043026	0.056108	-0.766848	-0.066626	0.027683	-2.406758
$\pi_{Density}$				-0.304603	0.279441	-1.090044	-0.373481	0.256581	-1.455606
π_{wtuc}				-0.302400	0.117226	-2.579640	-0.286539	0.031452	-9.110364
π_{wmfg}				0.245118	0.125948	1.946191	0.164364	0.065341	2.515483
${\sigma^2}_\epsilon$	0.020159			0.020424			0.020180		
${\sigma^2}_{\mu_i}$	0.129723			0.028189			0.044533		

Robust linear static panel data models using ε-contamination
Appendix

Badi H. Baltagi
Syracuse University, USA

Georges Bresson
Université Paris II, France

Anoop Chaturvedi
University of Allahabad, India

Guy Lacroix
Université Laval, Canada

November 14, 2014

List of Tables

List of Figures

A Tables and Figures

Table A1: Random effects world (RE and Robust three-stage) , $\rho = 0.3$, $\varepsilon = 0.5$

			β_{11}	β_{12}	β_2	σ^2_ε	σ^2_μ	λ_β	λ_b
		true	1	1	1	1	0.428571		
$N = 100, T = 10$	RE	coef	1.000691	1.000169	1.000279	0.999261	0.426830		
		se	0.014858	0.014881	0.014830				
		rmse	0.014826	0.015006	0.015437				
	3S	coef	0.999051	0.998538	0.998584	0.996517	0.435226	0.218000	$< 10^{-6}$
		se	0.007302	0.007308	0.007283				
		rmse	0.018665	0.019004	0.018925				
$N = 500, T = 5$	RE	coef	0.999882	1.000149	0.999313	1.001947	0.429767		
		se	0.008023	0.008019	0.008031				
		rmse	0.008099	0.008315	0.008083				
	3S	coef	0.998704	0.998993	0.998774	1.000902	0.432230	0.221000	$< 10^{-6}$
		se	0.004313	0.004311	0.004318				
		rmse	0.011321	0.011051	0.011187				
$N = 500, T = 10$	RE	coef	0.999661	1.000077	1.000106	1.000488	0.427950		
		se	0.006638	0.006640	0.006644				
		rmse	0.006964	0.006855	0.006668				
	3S	coef	0.999520	0.999852	0.999795	0.999920	0.429220	0.189000	$< 10^{-6}$
		se	0.003224	0.003226	0.003228				
		rmse	0.008490	0.008608	0.008463				

Table A2: Random effects world (RE and Robust three-stage) , $\rho = 0.8$, $\varepsilon = 0.5$

			β_{11}	β_{12}	β_2	σ^2_ε	σ^2_μ	λ_β	λ_b
		true	1	1	1	1	4		
$N = 100, T = 10$	RE	coef	1.000317	1.000381	1.000050	0.999261	3.977972		
		se	0.022856	0.022848	0.022866				
		rmse	0.022094	0.022848	0.023307				
	3S	coef	0.999118	0.999282	0.998613	0.996152	3.989020	0.117000	$< 10^{-6}$
		se	0.007282	0.007275	0.007283				
		rmse	0.022687	0.023644	0.024120				
$N = 500, T = 5$	RE	coef	1.000142	1.000352	1.000284	1.001947	4.007488		
		se	0.014973	0.014966	0.014970				
		rmse	0.014973	0.014656	0.014725				
	3S	coef	0.999474	0.999684	0.999770	1.000633	4.010065	0.110000	$< 10^{-6}$
		se	0.004305	0.004302	0.004307				
		rmse	0.016369	0.015631	0.016360				
$N = 500, T = 10$	RE	coef	0.999612	1.000320	0.999817	1.000488	3.998938		
		se	0.010219	0.010220	0.010223				
		rmse	0.010470	0.009993	0.00979				
	3S	coef	0.999403	0.999968	0.999445	0.999851	3.999162	0.123000	$< 10^{-6}$
		se	0.003223	0.003224	0.003226				
		rmse	0.010795	0.010404	0.010309				

Table A3: Random effects world (RE and Robust three-stage) , $N = 50$, $T = 20$, $\rho = 0.8$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	σ_ε^2	σ_μ^2	λ_β	λ_b
	true	1	1	1	1	4		
RE	coef	1.000024	1.000487	0.999531	1.001594	4.014982		
	se	0.021518	0.021487	0.021501				
	rmse	0.020823	0.020969	0.022112				
3S	coef	0.998745	0.999067	0.998213	0.998570	4.013042	$< 10^{-6}$	$< 10^{-6}$
	se	0.007568	0.007562	0.007572				
	rmse	0.021284	0.021341	0.022572				

Table A4: Mundlak-type FE (Mundlak and Robust three-stage), $N = 50$, $T = 20$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	π	σ_ε^2	σ_μ^2	λ_β	λ_b
	true	1	1	1	0.8	1	6.042385		
Mundlak	coef	0.999526	0.998582	1.000859	0.798767	1.000251	6.053826		
	se	0.019208	0.019199	0.022949	0.048149				
	rmse	0.018978	0.019487	0.023712	0.049777				
3S	coef	0.999184	0.998374	1.000790	0.801838	0.997190	6.073324	$< 10^{-6}$	$< 10^{-6}$
	se	0.007527	0.007530	0.022337	0.024157				
	rmse	0.020250	0.020733	0.023698	0.049073				

Table A5: Chamberlain-type fixed effects world (MCS and Robust three-stage), $\varepsilon = 0.5$

$N = 100$	$T = 10$	β_{11}	β_{12}	β_2	σ_{ε^2}	σ^2_μ	λ_β	λ_b
	true	1	1	1	1	164.877911		
MCS	coef	0.999734	1.000641	0.999877	0.991008	147.854745		
	se	0.007557	0.007534	0.007097				
	rmse	0.028058	0.026801	0.028788				
3S	coef	0.995584	0.996618	0.999710	0.995767	164.127937	$< 10^{-6}$	$< 10^{-6}$
	se	0.007617	0.007636	0.022473				
	rmse	0.021145	0.021187	0.023026				

$N = 100$	$T = 10$	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
	true	0.134218	0.167772	0.209715	0.262144	0.327680	0.4096	0.512	0.64	0.8	1
MCS	coef	0.131258	0.170069	0.212625	0.257390	0.325159	0.417267	0.513975	0.639243	0.800052	0.994872
	se	0.021328	0.021370	0.021425	0.021351	0.021434	0.021407	0.021452	0.021464	0.021442	0.021361
	rmse	0.078583	0.080546	0.076578	0.079965	0.081544	0.083082	0.079140	0.080920	0.077519	0.078152
3S	coef	0.142346	0.176048	0.215688	0.263888	0.330449	0.406021	0.514393	0.634483	0.796446	0.987879
	se	0.021699	0.021662	0.021644	0.021735	0.021727	0.021696	0.021732	0.021830	0.021645	0.021750
	rmse	0.073779	0.074874	0.073688	0.075324	0.074875	0.074941	0.073863	0.069525	0.074327	0.074024

Table A6: Chamberlain-type fixed effects world (MCS and Robust three-stage), $\varepsilon = 0.5$

$N = 500$	$T = 10$	β_{11}	β_{12}	β_2	σ_{ε^2}	σ^2_μ	λ_β	λ_b
	true	1	1	1	1	166.265383		
MCS	coef	0.999862	1.000052	0.999703	0.988115	154.282781		
	se	0.003509	0.003504	0.003302				
	rmse	0.010927	0.010546	0.010900				
3S	coef	0.998971	0.999393	0.999875	0.999579	165.371009	$< 10^{-6}$	$< 10^{-6}$
	se	0.003257	0.003250	0.010016				
	rmse	0.009925	0.009006	0.010399				

$N = 500$	$T = 10$	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
	true	0.134218	0.167772	0.209715	0.262144	0.327680	0.4096	0.512	0.64	0.8	1
MCS	coef	0.133672	0.168554	0.208779	0.263908	0.328663	0.408839	0.510733	0.640110	0.800060	0.999796
	se	0.009850	0.009859	0.009864	0.009863	0.009857	0.009829	0.009852	0.009854	0.009869	0.009843
	rmse	0.031283	0.032089	0.031067	0.032428	0.032057	0.031566	0.031065	0.032900	0.032030	0.030432
3S	coef	0.136374	0.168731	0.210563	0.261646	0.329955	0.409903	0.513582	0.638542	0.796614	0.997833
	se	0.009187	0.009178	0.009195	0.009177	0.009185	0.009198	0.009190	0.009189	0.009188	0.009198
	rmse	0.031592	0.031640	0.034228	0.032080	0.031925	0.032124	0.032957	0.031616	0.031692	0.032376

Table A7: Chamberlain-type fixed effects world (MCS and Robust three-stage), $N = 50$, $T = 20$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	1	196.385034		
MCS	coef	1.001209	1.001209	1.000829	1.164418	178.392318		
	se	0.006429	0.006438	0.004028				
	rmse	0.056046	0.058138	0.048501				
3S	coef	0.995564	0.996689	1.000382	0.998992	196.392882	$< 10^{-6}$	$< 10^{-6}$
	se	0.009513	0.009512	0.022597				
	rmse	0.021361	0.020617	0.022751				

		π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}
	true	0.014412	0.018014	0.022518	0.028147	0.035184	0.043980	0.054976	0.068719	0.085899	0.107374
MCS	coef	0.019099	0.014041	0.017620	0.034780	0.030831	0.036222	0.054342	0.072291	0.091705	0.112630
	se	0.019430	0.019590	0.019447	0.019505	0.019331	0.019319	0.019429	0.019468	0.019350	0.019408
	rmse	0.158869	0.165085	0.169190	0.170240	0.162608	0.169416	0.171741	0.166212	0.163969	0.169013
3S	coef	0.020431	0.027743	0.028257	0.040232	0.039268	0.044844	0.062615	0.078016	0.091976	0.110463
	se	0.028773	0.028999	0.028759	0.028846	0.028593	0.028584	0.028744	0.028792	0.028629	0.028723
	rmse	0.120876	0.130771	0.132898	0.133862	0.127165	0.129353	0.129107	0.129427	0.133017	0.130166

		π_{11}	π_{12}	π_{13}	π_{14}	π_{15}	π_{16}	π_{17}	π_{18}	π_{19}	π_{20}
	true	0.134218	0.167772	0.209715	0.262144	0.327680	0.4096	0.512	0.64	0.8	1
MCS	coef	0.140036	0.151234	0.204868	0.264955	0.327814	0.415780	0.520843	0.643831	0.789928	1.001918
	se	0.019490	0.019475	0.019409	0.019347	0.019547	0.019283	0.019561	0.019488	0.019513	0.019495
	rmse	0.167770	0.171701	0.166056	0.170118	0.166856	0.166020	0.172992	0.167415	0.170720	0.168452
3S	coef	0.138969	0.158330	0.210069	0.265858	0.330364	0.407856	0.506267	0.632105	0.777525	0.977140
	se	0.028808	0.028832	0.028734	0.028653	0.028948	0.028527	0.028920	0.028850	0.028874	0.028834
	rmse	0.127481	0.129830	0.125890	0.128229	0.132065	0.128107	0.132103	0.135215	0.131976	0.132502

Table A8: Hausman-Taylor world (HT and Robust three-stage) , $\varepsilon = 0.5$

			β_{11}	β_{12}	β_2	η_1	η_2	σ_ε^2	σ_μ^2	λ_β	λ_b
$N = 500, T = 5$		true	1	1	1	1	1	1	0.428571		
	HT	coef	1.000126	0.999942	0.999749	1.001554	1.000775	0.998995	0.442810		
		se	0.015238	0.015216	0.019945	0.036077	0.057130				
$\rho = 0.3$		rmse	0.015228	0.015361	0.019812	0.036052	0.056667				
	3S	coef	0.993609	0.993411	0.999564	1.000962	1.038739	1.002063	0.554927	$< 10^{-6}$	$< 10^{-6}$
		se	0.005488	0.005481	0.017900	0.018130	0.014502				
		rmse	0.012874	0.013213	0.019852	0.035099	0.045958				
		true	1	1	1	1	1	1	4		
	HT	coef	0.999859	1.000146	1.000641	0.998729	0.995684	0.999588	4.081871		
		se	0.017629	0.017651	0.019951	0.092787	0.079676				
$\rho = 0.8$		rmse	0.017674	0.018519	0.020089	0.091737	0.083407				
	3S	coef	0.992433	0.992113	0.999896	0.998508	1.048793	1.002774	3.797021	$< 10^{-6}$	$< 10^{-6}$
		se	0.005478	0.005482	0.017910	0.018816	0.013601				
		rmse	0.013942	0.014401	0.019995	0.086621	0.055194				
$N = 500, T = 10$		true	1	1	1	1	1	1	0.428571		
	HT	coef	1.000026	0.999581	1.000249	0.999999	1.001866	0.999121	0.430370		
		se	0.009360	0.009363	0.011446	0.032681	0.035614				
$\rho = 0.3$		rmse	0.009093	0.009270	0.011917	0.033689	0.034860				
	3S	coef	0.996381	0.996352	1.000096	0.999679	1.026482	1.000485	0.481835	$< 10^{-6}$	$< 10^{-6}$
		se	0.004117	0.004119	0.010873	0.013541	0.010913				
		rmse	0.008412	0.008532	0.011909	0.033095	0.031538				
		true	1	1	1	1	1	1	4		
	HT	coef	1.000065	0.999783	0.999599	1.000271	1.001514	1.000335	4.035335		
		se	0.010858	0.010866	0.011455	0.091137	0.064883				
$\rho = 0.8$		rmse	0.010445	0.010817	0.011578	0.092349	0.064688				
	3S	coef	0.996137	0.995958	0.999253	0.999770	1.031606	1.001729	3.866142	$< 10^{-6}$	$< 10^{-6}$
		se	0.004114	0.004121	0.010883	0.013919	0.010305				
		rmse	0.008596	0.008759	0.011583	0.089565	0.035819				

Table A9: Hausman-Taylor world (HT and Robust three-stage), $N = 50$, $T = 20$, $\rho = 0.3, 0.8$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	η_1	η_2	σ_ε^2	σ_μ^2	λ_β	λ_b
$\rho = 0.3$	true	1	1	1	1	1	1	0.428571		
HT	coef	0.998745	0.998572	1.001290	0.996105	1.006161	0.994668	0.426619		
	se	0.019822	0.019872	0.022501	0.099907	0.084415				
	rmse	0.019790	0.019778	0.022199	0.099232	0.085181				
3S	coef	0.997139	0.997086	1.000238	0.992574	1.015758	1.011238	0.460064	$< 10^{-6}$	$< 10^{-6}$
	se	0.009821	0.009868	0.021997	0.033154	0.026932				
	rmse	0.017018	0.017172	0.022041	0.097142	0.043981				
$\rho = 0.8$	true	1	1	1	1	1	1	4		
HT	coef	0.998931	0.999136	1.001683	0.996153	1.001815	0.998384	4.066213		
	se	0.022007	0.022013	0.022504	0.292926	0.201864				
	rmse	0.022316	0.022878	0.023375	0.286015	0.215722				
3S	coef	0.997645	0.998117	0.999945	0.992022	1.015758	0.994894	3.930940	$< 10^{-6}$	$< 10^{-6}$
	se	0.002355	0.001883	0.000055	0.007978	0.025283				
	rmse	0.017658	0.017969	0.023018	0.272587	0.043631				

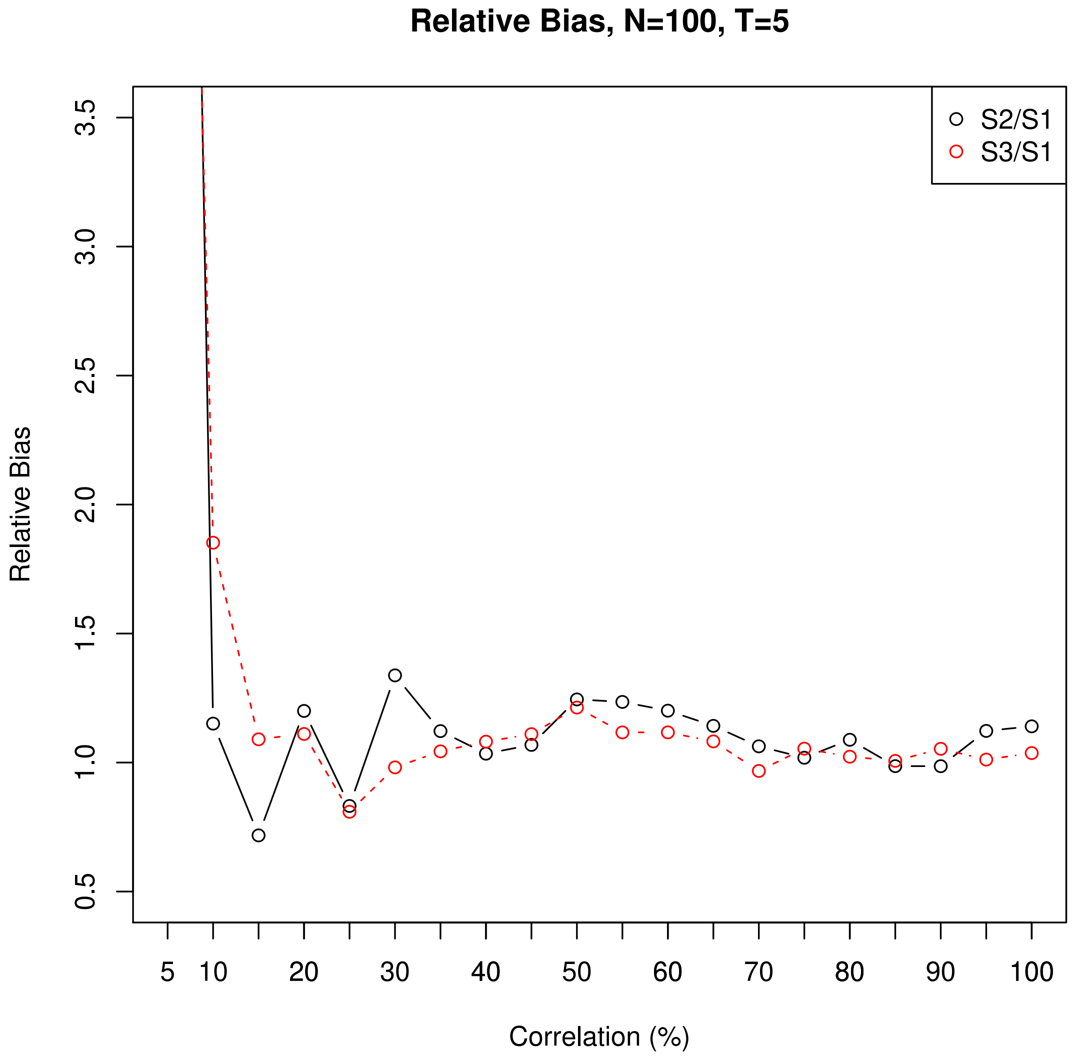

Figure 1: Hausman-Taylor world - Relative biases of $(\eta_2 - \hat{\eta}_2)$ for s=1, s=2 and s=3, N=100, T=5

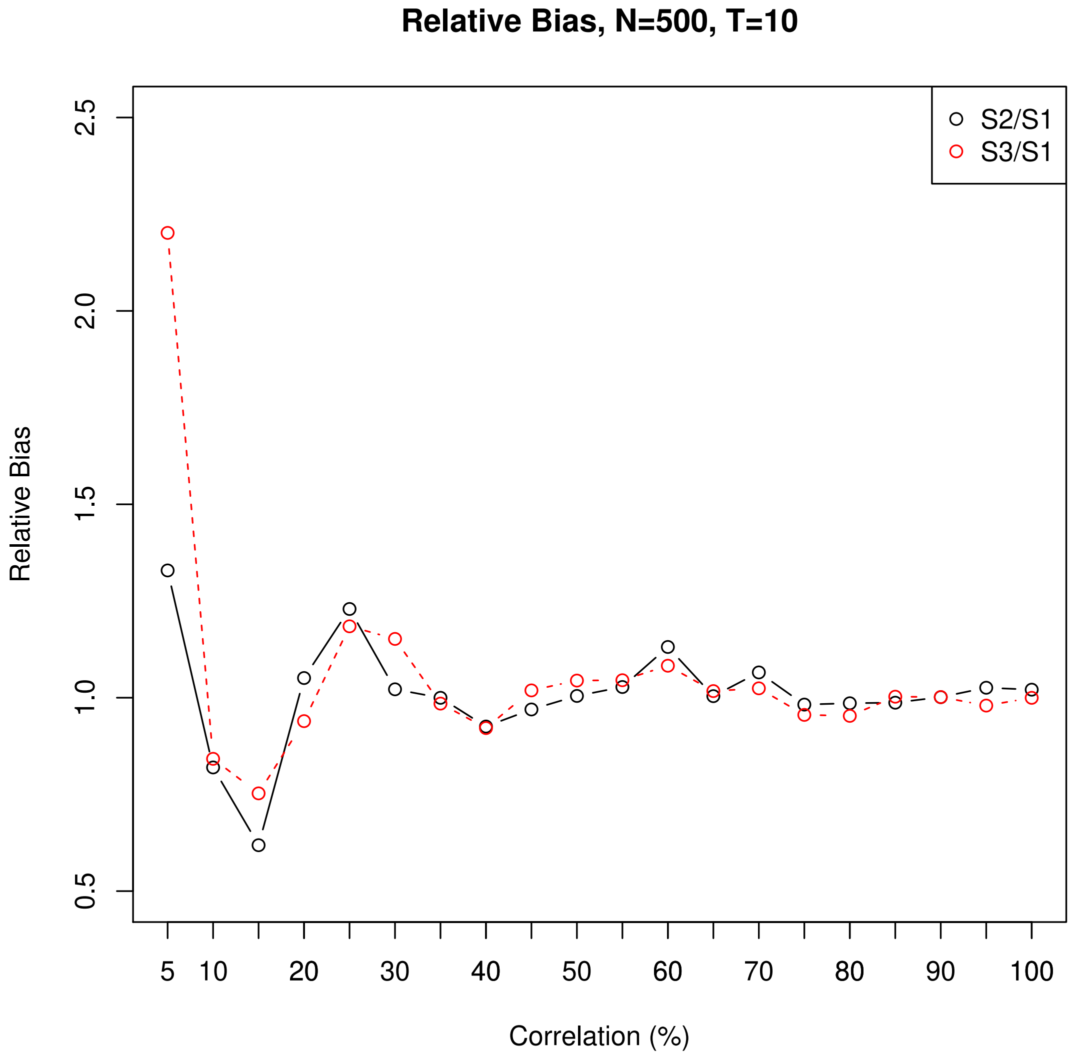

Figure 2: Hausman-Taylor world - Relative biases of $(\eta_2 - \hat{\eta}_2)$ for s=1, s=2 and s=3, N=500, T=10

Table A10: Random effects world (Robust three-stage), $N = 100$, $T = 5$, $\rho = 0.8$

		β_{11}	β_{12}	β_2	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	1	4		
$\varepsilon = 0.1$	coef	0.996484	0.996961	0.997766	0.995032	4.000669	0.116000	$< 10^{-6}$
	se	0.009746	0.009725	0.009743				
	rmse	0.036711	0.043164	0.037201				
$\varepsilon = 0.3$	coef	0.996486	0.996962	0.997768	0.995032	4.000670	0.116000	$< 10^{-6}$
	se	0.009746	0.009725	0.009743				
	rmse	0.036711	0.036959	0.037201				
$\varepsilon = 0.5$	coef	0.996486	0.996962	0.997768	0.995032	4.000670	0.116000	$< 10^{-6}$
	se	0.009746	0.009725	0.009743				
	rmse	0.036711	0.036959	0.037201				
$\varepsilon = 0.7$	coef	0.996486	0.996963	0.997768	0.995032	4.000670	0.116000	$< 10^{-6}$
	se	0.009746	0.009725	0.009743				
	rmse	0.036711	0.036959	0.037201				
$\varepsilon = 0.9$	coef	0.996486	0.996963	0.997768	0.995032	4.000670	0.116000	$< 10^{-6}$
	se	0.009746	0.009725	0.009743				
	rmse	0.036711	0.036959	0.037201				

Table A11: Hausman-Taylor world (Robust three-stage), $N = 100$, $T = 5$, $\rho = 0.8$

			β_{11}	β_{12}	β_2	η_1	η_2	σ^2_ε	σ^2_μ	λ_β	λ_b
$N = 100, T = 5$		true	1	1	1	1	1	1	4		
$\varepsilon = 0.1$	3S	coef	0.992970	0.992485	0.999145	0.999919	1.045505	0.993297	3.933178	$< 10^{-6}$	$< 10^{-6}$
		se	0.012493	0.012487	0.040310	0.048866	0.031320				
		rmse	0.027237	0.026860	0.045558	0.202091	0.072753				
$\varepsilon = 0.3$	3S	coef	0.991997	0.991802	0.999592	0.990634	1.046380	0.993812	3.894530	$< 10^{-6}$	$< 10^{-6}$
		se	0.012570	0.012554	0.040260	0.049227	0.031431				
		rmse	0.027794	0.028655	0.044864	0.199983	0.072734				
$\varepsilon = 0.5$	3S	coef	0.991538	0.990087	1.001559	1.002670	1.047291	0.992645	3.975909	$< 10^{-6}$	$< 10^{-6}$
		se	0.012486	0.012499	0.040205	0.048986	0.031392				
		rmse	0.028299	0.028394	0.043379	0.043379	0.073468				
$\varepsilon = 0.7$	3S	coef	0.992565	0.992406	1.002349	0.997065	1.045766	0.996919	3.909636	$< 10^{-6}$	$< 10^{-6}$
		se	0.012522	0.012567	0.040305	0.048530	0.031462				
		rmse	0.026568	0.026779	0.044342	0.190328	0.071724				
$\varepsilon = 0.9$	3S	coef	0.991470	0.991191	0.999245	1.002372	1.046966	0.992164	3.938627	$< 10^{-6}$	$< 10^{-6}$
		se	0.012517	0.012553	0.040298	0.048586	0.031387				
		rmse	0.027284	0.028120	0.045364	0.198722	0.073272				

Table A12: Skewed t-distribution - Random effects world (RE, Robust three-stage) , N=100, T=5, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	6.989414	4		
RE	coef	0.997920	0.999434	0.998785	6.930970	3.829401		
	se	0.195755	0.196371	0.195934				
	rmse	0.052993	0.051448	0.052255				
3S	coef	0.994568	0.994638	0.998761	6.945431	4.064348	0.204052	$< 10^{-6}$
	se	0.025057	0.025180	0.025087				
	rmse	0.070044	0.065293	0.068485				

Table A13: Skewed t-distribution - Chamberlain world (MCS, Robust three-stage) , $N = 100, T = 5$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	π_1	π_2	π_3	π_4	π_5	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	0.4096	0.512	0.64	0.8	1	6.819614	95.995993		
MCS	coef	0.999568	0.998549	1.000248	0.412606	0.510499	0.638155	0.799374	0.993463	6.572570	93.279720		
	se	0.072712	0.072807	0.227548	0.198072	0.198430	0.198743	0.198719	0.197763				
	rmse	0.036403	0.036606	0.085551	0.100238	0.100817	0.101374	0.100260	0.099848				
3S	coef	0.996461	0.995922	0.998294	0.421257	0.516924	0.641632	0.799025	0.989815	6.777786	96.314661	0.000121	$< 10^{-6}$
	se	0.025313	0.025357	0.081224	0.069478	0.069593	0.069735	0.069699	0.069502				
	rmse	0.056813	0.056295	0.091028	0.104254	0.103798	0.102057	0.104157	0.102902				

Table A14: Skewed t-distribution - Hausman-Taylor world (HT, Robust three-stage) , $N = 100, T = 5$, $\varepsilon = 0.5$

		β_{11}	β_{12}	β_2	η_1	η_2	σ^2_ε	σ^2_μ	λ_β	λ_b
	true	1	1	1	1	1	7.277104	4		
HT	coef	1.002909	1.003451	1.005721	1.005570	0.989824	7.210907	5.558088		
	se	0.090401	0.090200	0.113958	0.275183	0.350130				
	rmse	0.099392	0.103628	0.115303	0.243998	0.390042				
3S	coef	0.991418	0.992083	0.999541	0.989044	1.003796	7.220298	4.571552	0.003713	$< 10^{-6}$
	se	0.032428	0.032415	0.103877	0.126518	0.081135				
	rmse	0.073033	0.076466	0.115430	0.223265	0.162827				

B Derivations

B.1 The first step of the robust Bayesian estimator in the two-stage hierarchy[1]

B.1.1 Derivation of eq.(9)

As $y = X\beta + Wb + \varepsilon$, $\varepsilon \sim N(0, \Sigma)$ with $\Sigma = \tau^{-1} I_{NT}$, the joint probability density function (pdf) of y, given the observables and the parameters, is:

$$p(y \mid X, b, \tau, \beta) = \left(\frac{\tau}{2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{\tau}{2}(y - X\beta - Wb)'(y - X\beta - Wb)\right).$$

Let $y^* = y - Wb$. We can write (see Koop (2003), Bauwens *et al.* (2005) or Hsiao and Pesaran (2008) for instance):

$$(y^* - X\beta)'(y^* - X\beta) = y^{*\prime}y^* - y^{*\prime}X\beta - \beta' X' y^* + \beta' X' X\beta$$

and

$$p(y^* \mid X, b, \tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{\tau}{2}(y^* - X\beta)'(y^* - X\beta)\right).$$

Let $\widehat{\beta}(b) = (X'X)^{-1}X'y^* = \Lambda_X^{-1}X'y^*$ and $v(b) = (y^* - X\widehat{\beta}(b))'(y^* - X\widehat{\beta}(b))$. Then

$$\begin{aligned}
\Lambda_X \widehat{\beta}(b) &= X'y^* \text{ and } \widehat{\beta}'(b) \\
\Lambda_X &= y^{*\prime}X(y^* - X\beta)'(y^* - X\beta) \\
&= y^{*\prime}y^* - \widehat{\beta}'(b)\,\Lambda_X\beta - \beta'\Lambda_X\widehat{\beta}(b) + \beta'\Lambda_X\beta \\
&= y^{*\prime}y^* - \widehat{\beta}'(b)\,\Lambda_X\beta - \beta'\Lambda_X\widehat{\beta}(b) + \beta'\Lambda_X\beta \\
&\quad + \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) \\
&\quad + \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) \\
(y^* - X\beta)'(y^* - X\beta) &= y^{*\prime}y^* + (\beta - \widehat{\beta}(b))'\Lambda_X\left(\beta - \widehat{\beta}(b)\right) \\
&\quad - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) + \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) \\
&\quad - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b).
\end{aligned}$$

Since

$$\begin{aligned}
y^{*\prime}y^* - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) + \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) - \widehat{\beta}'(b)\,\Lambda_X\widehat{\beta}(b) &= y^{*\prime}y^* - y^{*\prime}X\widehat{\beta}(b) - \widehat{\beta}'(b)\,X'y^* \\
&\quad + \widehat{\beta}'(b)\,X'X\widehat{\beta}(b) \\
&= \left(y^* - X\widehat{\beta}(b)\right)'(y^* - X\widehat{\beta}(b)) \\
&= v(\beta)
\end{aligned}$$

then

$$(y^* - X\beta)'(y^* - X\beta) = (\beta - \widehat{\beta}(b))'\Lambda_X\left(\beta - \widehat{\beta}(b)\right) + v(\beta).$$

So the joint pdf can be written as:

$$p(y^* \mid X, b, \tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{NT}{2}} \exp\left(-\frac{\tau}{2}\left\{v(\beta) + (\beta - \widehat{\beta}(b))'\Lambda_X\left(\beta - \widehat{\beta}(b)\right)\right\}\right).$$

[1] Derivations of the second step of the robust Bayesian estimator in the two-stage hierarchy are not reported here since they follow strictly those of the first step.

The base prior of β is given by: $\beta \sim N\left(\beta_0 \iota_{K_1}, (\tau g_0 \Lambda_X)^{-1}\right)$ with $\Lambda_X = X'X$. Combining the pdf of y and the pdf of the base prior, we get the predictive density corresponding to the base prior:

$$\begin{aligned}
m\left(y^* \mid \pi_0, b, g_0\right) &= \int_0^\infty \int_{\mathbb{R}^{K_1}} \pi_0\left(\beta, \tau \mid g_0\right) \times p\left(y^* \mid X, b, \tau\right)\, d\beta\, d\tau \\
&= \int_0^\infty \int_{\mathbb{R}^{K_1}} p\left(\beta \mid \tau, \beta_0, g_0\right) \times p\left(\tau\right) \times p\left(y^* \mid X, b, \tau\right)\, d\beta\, d\tau \\
&= \int_0^\infty \int_{\mathbb{R}^{K_1}} \left\{ \begin{array}{c} \left(\frac{\tau}{2\pi}\right)^{\frac{NT}{2}} \left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}} \left(\frac{1}{\tau}\right) |\Lambda_X|^{1/2} \\ \times \exp\left(-\frac{\tau g_0}{2}(\beta - \beta_0 \iota_{K_1})' \Lambda_X (\beta - \beta_0 \iota_{K_1})\right) \\ \times \exp\left(-\frac{\tau}{2}\left\{v\left(b\right) + (\beta - \widehat{\beta}\left(b\right))' \Lambda_X (\beta - \widehat{\beta}\left(b\right)\right\}\right) \end{array} \right\} d\beta\, d\tau
\end{aligned}$$

$$m\left(y^* \mid \pi_0, b, g_0\right) = \int_0^\infty \int_{\mathbb{R}^{K_1}} \left\{ \begin{array}{c} \left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}} g_0^{\frac{K_1}{2}} \left(\tau\right)^{\frac{NT+K_1}{2}-1} |\Lambda_X|^{1/2} \\ \times \exp\left(-\frac{\tau}{2}\left\{v\left(\beta\right)\right\} - \frac{\tau}{2}\left\{ \begin{array}{c} (\beta - \widehat{\beta}\left(b\right))' \Lambda_X (\beta - \widehat{\beta}\left(b\right) \\ + g_0 (\beta - \beta_0 \iota_{K_1})' \Lambda_X (\beta - \beta_0 \iota_{K_{11}}) \end{array} \right\}\right) \end{array} \right\} d\beta\, d\tau.$$

First, we will simplify the expression inside the exponential:

$$F = (\beta - \widehat{\beta}\left(b\right))' \Lambda_X (\beta - \widehat{\beta}\left(b\right) + g_0 (\beta - \beta_0 \iota_{K_1})' \Lambda_X (\beta - \beta_0 \iota_{K_1}).$$

As the Bayes estimate of β is given by[2] (see Bauwens *et al.* (2005)):

$$\beta_*(b \mid g_0) = \left(\frac{\widehat{\beta}\left(b\right) + g_0 \beta_0 \iota_{K_1}}{n_*}\right) \text{ with } n_* = g_0 + 1,$$

then

$$\begin{aligned}
F &= n_*(\beta' \Lambda_X \beta - 2\beta_*'(b) \Lambda_X \beta) + g_0 \beta_0 \iota_{K_1}' \Lambda_X \iota_{K_{11}} + \widehat{\beta}'\left(b\right) \Lambda_X \widehat{\beta}\left(b\right) \\
&= n_*(\beta - \beta_*(b \mid g_0))' \Lambda_X (\beta - \beta_*(b \mid g_0)) - n_* \beta_*'(b) \Lambda_X \beta_*(b \mid g_0) \\
&\quad + g_0 \beta_0 \iota_{K_1}' \Lambda_X \iota_{K_1} + \widehat{\beta}'\left(b\right) \Lambda_X \widehat{\beta}\left(b\right) \\
&= n_*(\beta - \beta_*(b \mid g_0))' \Lambda_X (\beta - \beta_*(b \mid g_0)) + \frac{g_0}{n_*}\left(\beta_0 \iota_{K_1} - \widehat{\beta}\left(b\right)\right)' \Lambda_X \left(\beta_0 \iota_{K_1} - \widehat{\beta}\left(b\right)\right) \\
&= \left(g_0 + 1\right)\left(\beta - \beta_*(b \mid g_0)\right)' \Lambda_X (\beta - \beta_*(b \mid g_0)) \\
&\quad + \left(\frac{g_0}{g_0 + 1}\right)\left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right).
\end{aligned}$$

We can then write

$$\begin{aligned}
m\left(y^* \mid \pi_0, b, g_0\right) &= \int_0^\infty \int_{\mathbb{R}^{K_1}} \left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}} g_0^{\frac{K_1}{2}} \left(\tau\right)^{\frac{NT+K_1}{2}-1} |\Lambda_X|^{1/2} \\
&\quad \times \exp\left[\begin{array}{c} -\frac{\tau}{2}\left\{v\left(\beta\right)\right\} \\ -\frac{\tau}{2}\left\{ \begin{array}{c} \left(g_0 + 1\right)\left(\beta - \beta_*(b \mid g_0)\right)' \Lambda_X (\beta - \beta_*(b \mid g_0)) \\ + \left(\frac{g_0}{g_0+1}\right)\left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right) \end{array} \right\} \end{array} \right] d\beta\, d\tau \\
&= \int_0^\infty \left\{ \int_{\mathbb{R}^{K_1}} \exp\left[-\frac{\tau}{2}\left(g_0 + 1\right)\left(\beta - \beta_*(b \mid g_0)\right)' \Lambda_X (\beta - \beta_*(b \mid g_0))\right] d\beta \right\} \\
&\quad \times \left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}} g_0^{\frac{K_1}{2}} \left(\tau\right)^{\frac{NT+K_1}{2}-1} |\Lambda_X|^{1/2} \\
&\quad \times \left\{ \exp\left[-\frac{\tau}{2}\left\{v\left(b\right) + \left(\frac{g_0}{g_0 + 1}\right)\left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}\left(b\right) - \beta_0 \iota_{K_1}\right)\right\}\right] d\tau \right\}.
\end{aligned}$$

[2]Derivation of this estimator is presented below.

The multiple integral

$$I_{\mathbb{R}^{K_1}}=\int\limits_{\mathbb{R}^{K_1}}\exp\left(-\frac{\tau}{2}(g_0+1)(\beta-\beta_*(b\mid g_0))'\Lambda_X(\beta-\beta_*(b\mid g_0))\right)d\beta$$

can be written as

$$\begin{aligned}I_{\mathbb{R}^{K_1}} &= \int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}\exp\left(-\frac{\tau}{2}(g_0+1)(\beta-\beta_*(b\mid g_0))'\Lambda_X(\beta-\beta_*(b\mid g_0))\right)d\beta_1...d\beta_{K_1}\\ &= |D|^{-1}\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}\exp\left(-\frac{\tau}{2}\left(g_0+1\right)s's\right)ds_1...ds_{K_1},\end{aligned}$$

where $s=D(\beta-\beta_*(b\mid g_0))$. Then,

$$I_{\mathbb{R}^{K_1}}=|D|^{-1}\left[\int\limits_{-\infty}^{\infty}\exp\left(-\frac{\tau}{2}\left(g_0+1\right)s^2\right)\ ds\right]^{K_1},$$

and using the Gauss integral formula

$$\int\limits_{-\infty}^{\infty}\exp\left(-ax^2\right)\ dx=\sqrt{\frac{\pi}{a}}$$

we get

$$I_{\mathbb{R}^{K_1}}=|D|^{-1}\left(\frac{2\pi}{\tau(g_0+1)}\right)^{K_1/2}=|\Lambda_X|^{-1/2}\left(2\pi\right)^{K_1/2}.\left[\tau(g_0+1)\right]^{-K_1/2}.$$

Hence we can write

$$m\left(y\mid\pi_0,b,g_0\right)=\int\limits_{0}^{\infty}(2\pi)^{K_1/2}\left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}}g_0^{K_1/2}(g_0+1)^{-K_1/2}.\left|\Lambda_X\right|^{-1/2}\left|\Lambda_X\right|^{1/2}\tau^{-K_1/2}\tau^{\frac{NT+K_1}{2}-1}$$

$$\times\exp\left(-\frac{\tau}{2}\left\{v\left(b\right)+\left(\frac{g_0}{g_0+1}\right)(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})\right\}\right)d\tau$$

$$m\left(y\mid\pi_0,b,g_0\right)=(2\pi)^{-NT/2}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}\int\limits_{0}^{\infty}\tau^{\frac{NT}{2}-1}\exp\left(-\frac{\tau}{2}v\left(b\right)\left\{\begin{array}{c}1+\\ \left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right)\end{array}\right\}\right)d\tau,$$

where

$$R^2_{\beta_0}=\frac{(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})}{(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})+v\left(b\right)}.$$

We thus get

$$\begin{aligned}m\left(y\mid\pi_0,bg_0\right) &= \widetilde{H}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}\left(1+\left(\frac{g_0}{g_0+1}\right)\frac{(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})}{v\left(b\right)}\right)^{-\frac{NT}{2}}\\ &= \widetilde{H}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}\left(1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right)\right)^{-\frac{NT}{2}} \qquad (9)\end{aligned}$$

with

$$\widetilde{H}=\frac{\Gamma\left(\frac{NT}{2}\right)}{\pi^{\left(\frac{NT}{2}\right)}v\left(b\right)^{\left(\frac{NT}{2}\right)}}.$$

Q.E.D

B.1.2 Derivation of eq.(15) and eq.(16)

The maximization of $m\left(y^{*}\mid q,b,g_{0}\right)$ is equivalent to maximizing $\log m\left(y^{*}\mid q,b,g_{0}\right)$. Write:

$$\begin{aligned}\log m\left(y^{*}\mid q,b,g_{0}\right) &= \log\widetilde{H}+\frac{K_{1}}{2}\log\left(\frac{g_{q}}{g_{q}+1}\right)\\ &-\frac{NT}{2}\log\left(1+\left(\frac{g_{q}}{g_{q}+1}\right)\frac{(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})}{v\left(b\right)}\right).\end{aligned}$$

Next we derive the above expression with respect to β_q and g_q to obtain the first order conditions:

$$\frac{\partial\log m\left(y^{*}\mid q,b,g_{0}\right)}{\partial\beta_{q}}=0\text{ and }\frac{\partial\log m\left(y^{*}\mid q,b,g_{0}\right)}{\partial g_{q}}=0$$

The first term, $\left(\partial\log m\left(y^{*}\mid q,b\right)/\partial\beta_{q}\right)$, leads to

$$\begin{aligned}\frac{\partial\log m\left(y^{*}\mid q,b,g_{0}\right)}{\partial\beta_{q}} &= -\left(\frac{NT}{2}\right)\frac{\partial}{\partial\beta_{q}}\left\{\log\left(\begin{array}{c}1+\\ \left(\frac{g_{q}}{g_{q}+1}\right)\frac{(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})}{v(b)}\end{array}\right)\right\}\\ &= -\left(\frac{NT}{2}\right)\cdot\left[\frac{1}{1+\left(\frac{g_{q}}{g_{q}+1}\right)\cdot\frac{(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})}{v(\beta)}}\right]\\ &\quad\times\left(\frac{g_{q}}{g_{q}+1}\right)(-2)\,(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}\iota_{K_{1}}=0.\end{aligned}$$

Since

$$\left[1+\left(\frac{g_{q}}{g_{q}+1}\right)\frac{(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})}{v\left(\beta\right)}\right]^{-1}\neq0\text{ and finite}$$

it follows that

$$(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}\iota_{K_{1}}=0.$$

Thus

$$\widehat{\beta}_{q}=\left(\iota_{K_{1}}'\Lambda_{X}\iota_{K_{1}}\right)^{-1}\iota_{K_{1}}'\Lambda_{X}\widehat{\beta}\left(b\right). \tag{15}$$

The second term of the first order conditions is

$$\frac{\partial\log m\left(y^{*}\mid q,b\right)}{\partial g_{q}}=0.$$

This implies

$$\begin{aligned}\frac{\partial\log m\left(y^{*}\mid q,b,g_{0}\right)}{\partial g_{q}} &= \frac{\partial}{\partial g_{q}}\left\{\frac{K_{1}}{2}\log\left(\frac{g_{q}}{g_{q}+1}\right)\right\}\\ &-\left(\frac{NT}{2}\right)\frac{\partial}{\partial g_{q}}\left\{\log\left(\begin{array}{c}1+\\ \left(\frac{g_{q}}{g_{q}+1}\right)\frac{(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}(b)-\beta_{q}\iota_{K_{1}})}{v(b)}\end{array}\right)\right\}\\ &= \frac{K_{1}}{2}\left[\frac{1}{g_{q}\left(g_{q}+1\right)}\right]-\left(\frac{NT}{2}\right)\left[\frac{\left(\frac{R_{\beta_{q}}^{2}}{1-R_{\beta_{q}}^{2}}\right)}{(g_{q}+1)+g_{q}\left(\frac{R_{\beta_{q}}^{2}}{1-R_{\beta_{q}}^{2}}\right)}\frac{1}{g_{q}+1}\right]=0,\end{aligned}$$

with

$$R_{\beta_{q}}^{2}=\frac{(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})}{(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})'\Lambda_{X}(\widehat{\beta}\left(b\right)-\beta_{q}\iota_{K_{1}})+v\left(b\right)}.$$

Therefore

$$\frac{K_{1}}{2}\left[\frac{1}{g_{q}\left(g_{q}+1\right)}\right]=\left(\frac{NT}{2}\right)\left[\frac{\left(\frac{R_{\beta_{q}}^{2}}{1-R_{\beta_{q}}^{2}}\right)}{(g_{q}+1)\left[(g_{q}+1)+g_{q}\left(\frac{R_{\beta_{q}}^{2}}{1-R_{\beta_{q}}^{2}}\right)\right]}\right]$$

or equivalently

$$\frac{K_1}{2}\left[\frac{1}{g_q}\right]=\left(\frac{NT}{2}\right)\left[\frac{\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}{(g_q+1)+g_q\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}\right]$$

$$g_q=\left(\frac{K_1}{NT}\right)\left[\frac{g_q\left(\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)+1\right)+1}{\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}\right].$$

Hence

$$g_q=\frac{K_1}{NTB-K_1\left(\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)+1\right)}=\left(\frac{NT-K_1}{K_1}.\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)-1\right)^{-1}.$$

It follows that

$$\begin{aligned}\widehat{g}_q &= \min\left(g_0,g_q^*\right) \qquad (16)\\ \text{with } g_q^* &= \max\left[0,\left(\frac{NT-K_1}{K_1}\frac{(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})}{v(b)}-1\right)^{-1}\right]\\ &= \max\left[0,\left(\frac{NT-K_1}{K_1}\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)-1\right)^{-1}\right].\end{aligned}$$

Q.E.D

B.1.3 Derivation of eq.(18), eq.(19), eq.(20) and eq.(22)

If $\pi_0^*(\beta,\tau\mid g_0)$ denotes the posterior density of (β,τ) for the prior $\pi_0(\beta,\tau)$ and if $q^*(\beta,\tau\mid g_0)$ denotes the posterior density of (β,τ) for the prior $q(\beta,\tau)$, then the ML-II posterior density of (β,τ) is given by

$$\begin{aligned}\widehat{\pi}^*(\beta,\tau\mid g_0) &= \frac{p(y^*\mid X,b,\tau)\,\widehat{\pi}(\beta,\tau\mid g_0)}{\int_0^\infty\int_{\mathbb{R}^{K_1}}p(y^*\mid X,b,\tau)\,\widehat{\pi}(\beta,\tau\mid g_0)\;d\beta\;d\tau}\\ &= \frac{p(y^*\mid X,b,\tau)\{(1-\varepsilon)\pi_0(\beta,\tau\mid g_0)+\varepsilon\widehat{q}(\beta,\tau\mid g_0)\}}{\int_0^\infty\int_{\mathbb{R}^{K_1}}p(y^*\mid X,b,\tau)\{(1-\varepsilon)\pi_0(\beta,\tau\mid g_0)+\varepsilon\widehat{q}(\beta,\tau\mid g_0)\}\;d\beta\;d\tau}\\ &= \frac{(1-\varepsilon)p(y^*\mid X,b,\tau)\pi_0(\beta,\tau\mid g_0)+\varepsilon p(y^*\mid X,b,\tau)\widehat{q}(\beta,\tau\mid g_0)}{\begin{pmatrix}(1-\varepsilon)\int_0^\infty\int_{\mathbb{R}^{K_1}}p(y^*\mid X,b,\tau)\pi_0(\beta,\tau\mid g_0)\,d\beta d\tau\\ +\varepsilon\int_0^\infty\int_{\mathbb{R}^{K_1}}p(y^*\mid X,b,\tau)\widehat{q}(\beta,\tau\mid g_0)\,d\beta d\tau\end{pmatrix}}.\end{aligned}$$

Since

$$\begin{aligned}\widehat{\pi}^*(\beta,\tau\mid g_0) &= \frac{(1-\varepsilon)p(y^*\mid X,b,\tau)\pi_0(\beta,\tau\mid g_0)+\varepsilon p(y^*\mid X,b,\tau)\widehat{q}(\beta,\tau\mid g_0)}{(1-\varepsilon)m(y^*\mid\pi_0,b,g_0)+\varepsilon m(y^*\mid\widehat{q},b,g_0)}\\ &= \widehat{\lambda}_\beta\left(\frac{p(y^*\mid X,b,\tau)\pi_0(\beta,\tau\mid g_0)}{m(y^*\mid\pi_0,b,g_0)}\right)+\left(1-\widehat{\lambda}_\beta\right)\left(\frac{p(y^*\mid X,b,\tau)\widehat{q}(\beta,\tau\mid g_0)}{m(y^*\mid\widehat{q},b,g_0)}\right),\end{aligned}$$

then

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

with

$$\widehat{\lambda}_{\beta,g_0}=\frac{(1-\varepsilon)m(y^*\mid\pi_0,b,g_0)}{(1-\varepsilon)m(y^*\mid\pi_0,b,g_0)+\varepsilon m(y^*\mid\widehat{q},b,g_0)}.$$

$$\widehat{\lambda}_{\beta,g_0} = \left[1+\frac{\varepsilon m\left(y^* \mid \widehat{q},b,g_0\right)}{\left(1-\varepsilon\right)m\left(y^*\mid \pi_0,b,g_0\right)}\right]$$
$$= \left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}}{\widehat{g}+1}}{\frac{g_0}{g_0+1}}\right)^{K_1/2}\left(\frac{1+\left(\frac{g_0}{g_0+1}\right)\frac{(\widehat{\beta}(b)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_0\iota_{K_1})}{v(b)}}{1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})}{v(b)}}\right)^{\frac{NT}{2}}\right]^{-1}$$
$$= \left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}}{\widehat{g}+1}}{\frac{g_0}{g_0+1}}\right)^{K_1/2}\left(\frac{1+\left(\frac{g_0}{g_0+1}\right)\left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right)}{1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}\right)^{\frac{NT}{2}}\right]^{-1} \tag{19}$$

Integration of $\widehat{\pi}^*\left(\beta,\tau \mid g_0\right)$ with respect to τ leads to the marginal ML-II posterior density of β :

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

We must first define $\pi_0^*\left(\beta,\tau \mid g_0\right)$ and $q^*\left(\beta,\tau \mid g_0\right)$. As

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

where

$$m\left(y^*\mid \pi_0,b\right) = \frac{\Gamma\left(\frac{NT}{2}\right)}{\pi^{\left(\frac{NT}{2}\right)}v\left(b\right)^{\left(\frac{NT}{2}\right)}}\left(\frac{g_0}{g_0+1}\right)^{K_1/2}$$
$$\times\left(1+\left(\frac{g_0}{g_0+1}\right)\frac{(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1})}{v\left(b\right)}\right)^{-\frac{NT}{2}},$$

and where

$$p\left(y^*\mid X,b,\tau\right)\pi_0\left(\beta,\tau\mid g_0\right) = \left(\begin{array}{c}\left(\frac{\tau}{2\pi}\right)^{\frac{NT}{2}}\left(\frac{\tau g_0}{2\pi}\right)^{\frac{K_1}{2}}\tau^{-1}\left|\Lambda_X\right|^{1/2}\\ \times\exp\left(-\frac{\tau g_0}{2}(\beta-\beta_0\iota_{K_1})'\Lambda_X(\beta-\beta_0\iota_{K_1})\right)\\ \times\exp\left(-\frac{\tau}{2}\left\{v\left(b\right)+(\beta-\widehat{\beta}\left(b\right))'\Lambda_X(\beta-\widehat{\beta}\left(b\right)\right\}\right)\end{array}\right)$$
$$= \tau^{\left(\frac{NT+K_1}{2}-1\right)}\left|\Lambda_X\right|^{1/2}\left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}}g_0^{\frac{K_1}{2}}\times\exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\beta}\right),$$

with

$$\varphi_{\pi_0,\beta} = v\left(\beta\right)+\left(g_0+1\right)\left(\beta-\beta_*(b)\right)'\Lambda_X\left(\beta-\beta_*(b)\right)$$
$$+\left(\frac{g_0}{g_0+1}\right)\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right),$$

then

$$\pi_0^*\left(\beta,\tau\mid g_0\right) = L_0\left(b\right)\times\tau^{\left(\frac{NT+K_1}{2}-1\right)}\times\exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\beta}\right),$$

where

$$L_0\left(b\right) = \frac{2^{-\left(\frac{NT+K_1}{2}\right)}}{\Gamma\left(\frac{NT}{2}\right).\pi^{K_1/2}}.\left(g_0+1\right)^{\frac{K_1}{2}}.v\left(b\right)^{\frac{NT}{2}}.\left|\Lambda_X\right|^{1/2}$$
$$\times\left[\left(1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)}{v\left(b\right)}\right)^{\left(\frac{NT}{2}\right)}\right].$$

Similarly, the expression of $q^*(\beta,\tau \mid g_0)$ is defined as:

$$\begin{aligned} q^*(\beta,\tau \mid g_0) &= \frac{p(y^* \mid X,b,\tau)\,\widehat{q}(\beta,\tau \mid g_0)}{m(y^* \mid \widehat{q},b,g_0)} = \frac{p(y^* \mid X,b,\tau)\,\widehat{q}(\beta,\tau \mid g_0)}{\int_0^\infty \int_{\mathbb{R}^{K_1}} p(y^* \mid X,b,\tau)\,\widehat{q}(\beta,\tau \mid g_0)\ d\beta\ d\tau} \\ &= L_{\widehat{q}}(b) \times \tau^{\left(\frac{NT+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\widehat{q},\beta}\right), \end{aligned}$$

with

$$\begin{aligned} \varphi_{\widehat{q},\beta} &= v(\beta) + (\widehat{g}+1)\left(\beta - \widehat{\beta}_{EB}(b \mid g_0)\right)' \Lambda_X \left(\beta - \widehat{\beta}_{EB}(b \mid g_0)\right) \\ &\quad + \left(\frac{\widehat{g}}{\widehat{g}+1}\right)\left(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1}\right) \end{aligned}$$

and

$$\begin{aligned} L_{\widehat{q}}(b) &= \frac{2^{-(K_1)}}{\Gamma\left(\frac{NT}{2}\right)\pi^{K_1/2}} (\widehat{g}+1)^{\frac{K_1}{2}}\, v(b)^{\left(\frac{NT}{2}\right)} |\Lambda_X|^{1/2} \\ &\quad \times \left[\left(1 + \left(\frac{\widehat{g}}{\widehat{g}+1}\right) \frac{\left(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1}\right)}{v(\beta)}\right)^{\left(\frac{NT}{2}\right)}\right], \end{aligned}$$

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

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

Integration of $\widehat{\pi}^*(\beta,\tau \mid g_0)$ with respect to τ leads to the marginal ML-II posterior density of β :

$$\begin{aligned} \widehat{\pi}^*(\beta \mid g_0) = & \int_0^\infty \widehat{\pi}^*(\beta,\tau \mid g_0)\,d\tau = \widehat{\lambda}_{\beta,g_0} \int_0^\infty \pi_0^*(\beta,\tau \mid g_0)\,d\tau + \left(1 - \widehat{\lambda}_{\beta,g_0}\right) \int_0^\infty q^*(\beta,\tau \mid g_0)\,d\tau \\ = & \quad \widehat{\lambda}_{\beta,g_0}\pi_0^*(\beta \mid g_0) + \left(1 - \widehat{\lambda}_{\beta,g_0}\right)\widehat{q}^*(\beta \mid g_0) \quad (18) \end{aligned}$$

So,

$$\begin{aligned} \pi_0^*(\beta \mid g_0) &= \int_0^\infty \pi_0^*(\beta,\tau \mid g_0)\,d\tau \\ &= L_0(b) \int_0^\infty \tau^{\left(\frac{NT+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\pi_0,\beta}\right) d\tau \\ &= L_0(b) \times 2^{\left(\frac{NT+K_1}{2}\right)} \varphi_{\pi_0,\beta}^{\left(-\frac{NT+K_1}{2}\right)} \Gamma\left(\frac{NT+K_1}{2}\right). \end{aligned}$$

Then $\pi_0^*(\beta \mid g_0)$ is given by

$$\begin{aligned} \pi_0^*(\beta \mid g_0) &= \frac{\Gamma\left(\frac{NT+K_1}{2}\right)}{\Gamma\left(\frac{NT}{2}\right)\pi^{\frac{K_1}{2}}} |\Lambda_X|^{1/2} (g_0+1)^{\frac{K}{2}}\, v(b)^{\left(\frac{NT}{2}\right)} \times \varphi_{\pi_0,\beta}^{\left(-\frac{NT+K_1}{2}\right)} \\ &\quad \times \left(1 + \left(\frac{g_0}{g_0+1}\right) \frac{\left(\widehat{\beta}(b) - \beta_0 \iota_{K_1}\right)' \Lambda_X \left(\widehat{\beta}(\mu) - \beta_0 \iota_{K_1}\right)}{v(b)}\right)^{\left(\frac{NT}{2}\right)}. \end{aligned}$$

We therefore get

$$\pi_0^*(\beta \mid g_0) = \widetilde{H}_{\pi_0} \frac{(g_0+1)^{K_1/2}}{\left((g_0+1)\frac{(\beta-\beta_*(b))'\Lambda_X(\beta-\beta_*(b))}{v(b)} + \left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)}{v(b)} + 1\right)^{\frac{NT+K_1}{2}}},$$

with

$$\widetilde{H}_{\pi_0} = \frac{\Gamma\left(\frac{NT+K_1}{2}\right)|\Lambda_X|^{1/2}}{\pi^{K/2}\Gamma\left(\frac{NT}{2}\right)v(\beta)^{K_1/2}} \times \left(1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)}{v(b)}\right)^{\frac{NT}{2}}.$$

If we suppose that $M_{0,\beta} = \frac{(g_0+1)}{v(b)}\Lambda_X$, then $|M_{0,\beta}|^{1/2} = \left(\frac{g_0+1}{v(b)}\right)^{K_1/2}|\Lambda_X|^{1/2}$ and

$$\pi_0^*(\beta \mid g_0) = \frac{\Gamma\left(\frac{NT+K_1}{2}\right)|M_{0,\beta}|^{1/2}}{\pi^{K_1/2}\Gamma\left(\frac{NT}{2}\right)}(\xi_{0,\beta})^{NT/2}\left[(\beta-\beta_*(b))'M_{0,\beta}(\beta-\beta_*(b))+\xi_{0,\beta}\right]^{-\frac{NT+K_1}{2}},$$

$$\text{with } \xi_{0,\beta} = 1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)}{v(b)}. \tag{20}$$

So $\pi_0^*(\beta \mid g_0)$ is the pdf of a multivariate t-distribution with mean vector $\beta_*(b)$, variance-covariance matrix $\left(\frac{\xi_{0,\beta}M_{0,\beta}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) (see Bauwens *et al.* (2005)). $q^*(\beta \mid g_0)$ is defined equivalently by:

$$\widehat{q}^*(\beta \mid g_0) = \int_0^\infty \widehat{q}^*(\beta,\tau \mid g_0)\,d\tau = L_{\widehat{q}}(b)\int_0^\infty \tau^{\left(\frac{NT+K_1}{2}-1\right)} \times \exp\left(-\frac{\tau}{2}\varphi_{\widehat{q},\beta}\right)d\tau.$$

Then $q^*(\beta)$ is given by

$$q^*(\beta \mid g_0) = \widetilde{H}_q\frac{(\widehat{g}+1)^{K_1/2}}{\left\{(\widehat{g}+1)\frac{\left(\beta-\widehat{\beta}_{EB}(b)\right)'\Lambda_X\left(\beta-\widehat{\beta}_{EB}(b)\right)}{v(b)}+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(\mu)-\widehat{\beta}_q\iota_{K_1}\right)}{v(b)}+1\right\}^{\frac{NT+K_1}{2}}},$$

with

$$\widetilde{H}_q = \frac{\Gamma\left(\frac{NT+K_1}{2}\right)|\Lambda_X|^{1/2}}{\pi^{K_1/2}\Gamma\left(\frac{NT}{2}\right)v(b)^{K_1/2}} \times \left(1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)}{v(b)}\right)^{\frac{NT}{2}}.$$

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

$$\xi_{1,\beta 1} = 1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)}{v(b)} \text{ and } M_{1,\beta} = \left(\frac{(\widehat{g}+1)}{v(\beta)}\right)\Lambda_X. \tag{22}$$

Q.E.D

B.1.4 Derivation of eq.(21) and eq.(23).

To prove equation (23), start from Bayes's theorem:

$$p(\beta|y^*) \propto p(y^*|\beta)\,p(\beta).$$

As $y^* \sim N\left(X\beta, \tau^{-1}I_{NT}\right)$ and $\beta \sim N\left(\widehat{\beta}_q\iota_{K_1}, (\tau\widehat{g}\Lambda_X)^{-1}\right)$, then the product $p(y^*|\beta)\,p(\beta)$ is proportional to $\exp\left\{-\frac{1}{2}Q^*\right\}$ where Q^* is given by (see Koop (2003), Bauwens *et al.* (2005) or Hsiao and Pesaran (2008) for instance):

$$
\begin{aligned}
Q^* &= \tau\left(y^*-X\beta\right)'\left(y^*-X\beta\right)+\tau\widehat{g}\left(\beta-\widehat{\beta}_q\iota_{K_1}\right)'\Lambda_X\left(\beta-\widehat{\beta}_q\iota_{K_1}\right)\\
&= \tau y^{*'}y^*-\tau y^{*'}X\beta-\tau\beta'X'y^*+\tau\beta'X'X\beta\\
&\quad +\tau\widehat{g}\beta'\Lambda_X\beta-\tau\widehat{g}\beta'\Lambda_X\widehat{\beta}_q\iota_{K_1}-\tau\widehat{g}\widehat{\beta}_q\iota_{K_1}\Lambda_X\beta+\tau\widehat{g}\left(\widehat{\beta}_q\right)^2\iota_{K_1}'\Lambda_X\iota_{K_1}.
\end{aligned}
$$

We can write

$$
\begin{aligned}
Q^* &= \left\{\tau\widehat{g}\beta'\Lambda_X\beta+\tau\beta'X'X\beta-\tau\widehat{g}\widehat{\beta}_q\beta'\Lambda_X\iota_{K_1}-\tau\beta'Xy^*-\tau\widehat{g}\widehat{\beta}_q\iota_{K_1}\Lambda_X\beta-\tau y^{*'}X\beta\right\}\\
&\quad+\left\{\tau y^{*'}y^*+\tau\widehat{g}\left(\widehat{\beta}_q\right)^2\iota_{K_1}'\Lambda_X\iota_{K_1}\right\}\\
&=\beta'\left(\tau\widehat{g}\Lambda_X+\tau X'X\right)\beta-\beta'\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)-\tau\widehat{g}\widehat{\beta}_q\iota_{K_1}\Lambda_X\beta-\tau y^{*'}X\beta\\
&\quad+\left\{\tau y^{*'}y^*+\tau\widehat{g}\widehat{\beta}_q^2\iota_{K_1}'\Lambda_X\iota_{K_1}\right\}
\end{aligned}
$$

Let $D=\left(\tau\widehat{g}\Lambda_X+\tau X'X\right)^{-1}$. If we add and subtract $R'DR$ in Q^*, with $R=\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)$, then

$$
\begin{aligned}
Q^* &= \left\{\begin{array}{c}\beta'\left(\tau\widehat{g}\Lambda_X+\tau X'X\right)\beta-\beta'\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)-\tau\widehat{g}\widehat{\beta}_q\iota_{K_1}\Lambda_X\beta-\tau y^{*'}X\beta\\
+\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)\end{array}\right\}\\
&\quad+\left\{\begin{array}{c}\tau y^{*\prime}y^*+\tau\widehat{g}\overline{\widehat{\beta}\left(\mu\right)}^2\iota_{K_1}'\Lambda_X\iota_{K_1}\\
-\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)\end{array}\right\}\\
&= Q_1^*+Q_2^*.
\end{aligned}
$$

So

$$
\begin{aligned}
Q_1^* &= \left\{\begin{array}{c}\beta'\left(\tau\widehat{g}\Lambda_X+\tau X'X\right)\beta-\beta'\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)-\tau\widehat{g}\widehat{\beta}_q\iota_{K_1}\Lambda_X\beta-\tau y^{*'}X\beta\\
+\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)\end{array}\right\}\\
&= \beta'D^{-1}\beta-\beta'\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)-\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'\beta\\
&\quad+\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)\\
&= \beta'D^{-1}\beta-\beta'D^{-1}D\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)-\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D'D^{-1}\beta\\
&\quad+\left(\tau\widehat{g}\Lambda_X\widehat{\beta}_q\iota_{K_1}+\tau X'y^*\right)'D'D^{-1}D\left(\tau\widehat{g}\Lambda_X\beta+\tau X'y^*\right)
\end{aligned}
$$

Let $\widehat{\beta}_{EB}(b\mid g_0)=D\left(\tau X'y^*+\tau\widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right)$. Then

$$
\begin{aligned}
Q_1^* &= \beta'D^{-1}\beta-\beta'D^{-1}\widehat{\beta}_{EB}(b\mid g_0)-\widehat{\beta}_{EB}'(b\mid g_0)D^{-1}\beta+\widehat{\beta}_{EB}'(b\mid g_0)D^{-1}\widehat{\beta}_{EB}(b\mid g_0)\\
&= \left(\beta-\widehat{\beta}_{EB}(b\mid g_0)\right)'D^{-1}\left(\beta-\widehat{\beta}_{EB}(b\mid g_0)\right).
\end{aligned}
$$

As

$$
\begin{aligned}
Q_2^* &= \tau y^{*'}y^*+\tau\widehat{g}\widehat{\beta}_q^2\iota_{K_1}'\Lambda_X\iota_{K_1}\\
&\quad-\left(\tau X'y^*+\tau\widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right)'D\left(\tau X'y^*+\tau\widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right),
\end{aligned}
$$

and as far as the distribution of $p\left(\beta|y^*\right)$ is concerned, Q_2^* is a constant. So $\exp\left\{-\frac{1}{2}Q_2^*\right\}$ integrates to 1. Therefore, the marginal distribution of β given y^* is proportional to $\exp\left\{-\frac{1}{2}Q_1^*\right\}$. Consequently, the empirical Bayes estimator $\widehat{\beta}_{EB}(b\mid g_0)$ of β is given by

$$
\widehat{\beta}_{EB}(b\mid g_0)=D\left(\tau X'y^*+\tau\widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right),\ \text{with } D=\left(\tau\widehat{g}\Lambda_X+\tau X'X\right)^{-1}.
$$

Hence

$$\widehat{\beta}_{EB}(b \mid g_0) = D\left(\tau X'y^* + \tau\widehat{g}\widehat{\beta}_q \Lambda_X \iota_{K_1}\right) = \left((\widehat{g}+1)\Lambda_X\right)^{-1}\left(X'y^* + \widehat{g}\widehat{\beta}_q \Lambda_X \iota_{K_1}\right)$$
$$= \left((\widehat{g}+1)\right)^{-1}\left(\Lambda_X^{-1}X'y^* + \widehat{g}\widehat{\beta}_q \iota_{K_1}\right)$$
$$= \frac{\widehat{\beta}(b) + \widehat{g}\widehat{\beta}_q \iota_{K_1}}{\widehat{g}+1} = \widehat{\beta}(b) - \frac{\widehat{g}}{\widehat{g}+1}\left(\widehat{\beta}(b) - \widehat{\beta}_q \iota_{K_1}\right). \tag{23}$$

Using Bayes's theorem once again:

$$p(\beta|y^*) \propto p(y^*|\beta)\, p(\beta).$$

As $y^* \sim N\left(X\beta, \tau^{-1}I_{NT}\right)$ and $\beta \sim N\left(\beta_0 \iota_{K_1}, (\tau g_0 \Lambda_X)^{-1}\right)$, then, following the previous derivations, we can show that $\beta_*(b \mid g_0)$ is the Bayes estimate of β for the prior distribution $\pi_0(\beta, \tau \mid g_0)$:

$$\beta_*(b \mid g_0) = \frac{\widehat{\beta}(b) + g_0 \beta_0 \iota_{K_1}}{g_0 + 1}. \tag{21}$$

Q.E.D

B.2 The second step of the robust Bayesian estimator in the three-stage hierarchy

B.2.1 Derivation of eq.(40) and eq.(42)

In the second step of the two-stage hierarchy, we have derived the predictive density corresponding to the base prior conditional on h_0:

$$m(\widetilde{y} \mid \pi_0, \beta, h_0) = \widetilde{H}\left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left(1 + \left(\frac{h_0}{h_0+1}\right)\frac{(\widehat{b}(\beta) - b_0\iota_{K_2})'\Lambda_W(\widehat{b}(\beta) - b_0\iota_{K_2})}{v(\beta)}\right)^{-\frac{NT}{2}}$$
$$= \widetilde{H}\left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left(1 + \left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_0}}{1-R^2_{b_0}}\right)\right)^{-\frac{NT}{2}}.$$

Then, the unconditional predictive density corresponding to the base prior is given by

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

since

$$p(h_0) = \frac{h_0^{c-1}(1+h_0)^{-(c+d)}}{B(c,d)}, \; c > 0,\, d > 0$$

Let $\varphi = \frac{h_0}{h_0+1}$. Then $1 - \varphi = \frac{1}{h_0+1}$, $h_0 = \frac{\varphi}{1-\varphi}$ and $dh_0 = (1-\varphi)^{-2} d\varphi$, so

$$m(\widetilde{y} \mid \pi_0, \beta) = \frac{\widetilde{H}}{B(c,d)} \int_0^1 (\varphi)^{\frac{K_2}{2}+c+1}(1-\varphi)^{d-1}\left(1 + \varphi\left(\frac{R^2_{b_0}}{1-R^2_{b_0}}\right)\right)^{-\frac{NT}{2}} d\varphi \tag{40}$$
$$= \frac{B(d, \frac{K_2}{2}+c)}{B(c,d)}\widetilde{H} \times {}_2F_1\left(\frac{NT}{2}; \frac{K_2}{2}+c; \frac{K_2}{2}+c+d; -\left(\frac{R^2_{b_0}}{1-R^2_{b_0}}\right)\right),$$

where ${}_2F_1$ is the Gaussian hypergeometric function.[3] Following the lines of the second step of the robust estimator in the two-stage hierarchy, we have

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

so

$$\widehat{h}_q = \begin{cases} h_0 & \text{if} \quad h_0 \leq h^* \\ h^* & \text{if} \quad h_0 > h^* \end{cases}$$

and the predictive density corresponding to the contaminated prior conditional on h_0 is:

$$m\left(y \mid \widehat{q}, b, h_0\right) = \begin{cases} \widetilde{H}\left(\frac{h_0}{h_0+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} & \text{if} \quad h_0 \leq h^* \\ \widetilde{H}\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} & \text{if} \quad h_0 > h^*. \end{cases}$$

Then the unconditional predictive density corresponding to the contaminated prior is given by:

$$\begin{aligned} m\left(\widetilde{y} \mid \widehat{q}, \beta\right) &= \int_0^\infty m\left(y \mid \widehat{q}, \beta, h_0\right).p(h_0)dh_0 \\ &= \frac{\widetilde{H}}{B\left(c,d\right)} \times \int_0^{h^*} \left(\begin{array}{c} \left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left(1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} \\ \times h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} \end{array} \right) dh_0 \\ &+ \frac{\widetilde{H}}{B\left(c,d\right)}\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} \times \\ &\int_{h^*}^\infty h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0. \end{aligned}$$

---

[3] The Euler integral formula is given by (see Abramovitz and Stegun (1970)):

$$\begin{aligned} \int_0^1 (t)^{a_2-1}\left(1-t\right)^{a_3-a_2-1}\left(1-zt\right)^{-a_1} dt &= B(a_2, a_3-a_2) \times_2 F_1\left(a_1; a_2; a_3; z\right) \\ &= \frac{\Gamma\left(a_2\right)\Gamma\left(a_3-a_2\right)}{\Gamma\left(a_3\right)} \times_2 F_1\left(a_1; a_2; a_3; z\right) \end{aligned}$$

where ${}_2F_1\left(a_1; a_2; a_3; z\right)$ is the Gaussian hypergeometric function with ${}_2F_1\left(a_1; a_2; a_3; z\right) \equiv_2 F_1\left(a_2; a_1; a_3; z\right)$. This is a special function represented by the hypergeometric series. For $\mid z \mid < 1$, the hypergeometric function is defined by the power series:

$${}_2F_1\left(a_1; a_2; a_3; z\right) =_{j=0}^{\infty} \frac{(a_1)_j (a_2)_j}{(a_3)_j} \frac{z^j}{j!}$$

where $(a_1)_j$ is the Pochhammer symbol defined by

$$(a_1)_j = \begin{cases} 1 & \text{if} \quad j=0 \\ a_1\left(a_1+1\right)...\left(a_1+j-1\right) = \frac{\Gamma(a_1+j)}{\Gamma(a_1)} & \text{if} \quad j>0 \end{cases}$$

Let $\varphi = \frac{h_0}{h_0+1}$. Then

$$\begin{aligned}
m\left(\widetilde{y} \mid \widehat{q}, \beta\right) &= \frac{\widetilde{H}}{B\left(c,d\right)} \times \int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^{\frac{K_2}{2}+c-1} \left(1-\varphi\right)^{d-1} \left(1+\varphi\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} d\varphi \\
&+ \frac{\widetilde{H}}{B\left(c,d\right)} \times \left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}} \left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} \\
&\times \int_{\frac{h^*}{h^*+1}}^{1} \left(\varphi\right)^{c-1} \left(1-\varphi\right)^{d-1} d\varphi,
\end{aligned}$$

so we get two incomplete Gaussian hypergeometric functions. Let $\varphi = \left(\frac{h^*}{h^*+1}\right) t$. The solution of the first one is given by:

$$\int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^{\frac{K_2}{2}+c-1} \left(1-\varphi\right)^{d-1} \left(1+\varphi\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}} d\varphi =$$

$$\begin{aligned}
&= \left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c} \int_0^1 t^{\frac{K_2}{2}+c-1} \left(1-\left(\frac{h^*}{h^*+1}\right)t\right)^{d-1} \left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)t\right)^{-\frac{NT}{2}} dt \\
&= \left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c} \times \frac{\Gamma\left(\frac{K_2}{2}+c\right)}{\Gamma\left(\frac{K_2}{2}+c+1\right)} \\
&\times F_1\left(\frac{K_2}{2}+c; \frac{NT}{2}; 1-d; \frac{K_2}{2}+c+1; \left(\frac{h^*}{h^*+1}\right); -\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right) \\
&= \frac{2\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c}}{K_2+2c} \times F_1\left(\frac{K_2}{2}+c; \frac{NT}{2}; 1-d; \frac{K_2}{2}+c+1; \left(\frac{h^*}{h^*+1}\right); -\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right),
\end{aligned}$$

where $F_1(.)$ is the Appell hypergeometric function.[4] The second incomplete Gaussian hypergeometric function can be written as:

$$\begin{aligned}
\int_{\frac{h^*}{h^*+1}}^{1} \left(\varphi\right)^{c-1} \left(1-\varphi\right)^{d-1} d\varphi &= \int_0^1 \left(\varphi\right)^{c-1} \left(1-\varphi\right)^{d-1} d\varphi - \int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^{c-1} \left(1-\varphi\right)^{d-1} d\varphi \\
&= B\left(c,d\right) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times {}_2F_1\left(c; d-1; c+1; \left(\frac{h^*}{h^*+1}\right)\right).
\end{aligned}$$

[4]The Appell hypergeometric function (see Appell (1882), Abramovitz and Stegun (1970), Slater (1966)) is a formal extension of the hypergeometric function to two variables:

$$\begin{aligned}
F_1\left(a; b_1; b_2; c; x; y\right) &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{(a)_{j+k}(b_1)_j(b_2)_k}{(c)_{j+k}} \frac{x^j}{j!} \frac{y^k}{k!} \\
&= \frac{\Gamma\left(c\right)}{\Gamma\left(a\right)\Gamma\left(c-a\right)} \int_0^1 \left(t\right)^{a-1} \left(1-t\right)^{c-a-1} \left(1-xt\right)^{-b_1} \left(1-yt\right)^{-b_2} dt
\end{aligned}$$

where $(a_1)_j$ is the Pochhammer symbol.

Then the unconditional predictive density corresponding to the contaminated prior is given by:

$$m\left(\widetilde{y} \mid \widehat{q}, \beta\right) = \tag{42}$$

$$= \frac{\widetilde{H}}{B\left(c,d\right)} \left\{ \begin{array}{c} \frac{2.\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c}}{K_2+2c} \\ \times F_1\left(\frac{K_2}{2}+c; 1-d; \frac{NT}{2}; \frac{K_2}{2}+c+1; \frac{h^*}{h^*+1}; -\frac{h^*}{h^*+1}\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right) \\ + \left\{ \begin{array}{c} \left[\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right)^{-\frac{NT}{2}}\right] \\ \times \left[\begin{array}{c} B\left(c,d\right) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times {}_2F_1\left(c; d-1; c+1; \frac{h^*}{h^*+1}\right) \end{array} \right] \end{array} \right\} \end{array} \right\}$$

B.2.2 Derivation of eq.(44)

Under the contamination class of prior, the empirical Bayes estimator of b in the two-stage hierarchy (conditional on h_0) can be written as:

$$\begin{aligned} \widehat{b}_{EB}\left(\beta \mid h_0\right) &= \left(\frac{\widehat{b}\left(\beta\right) + \widehat{h}_q \widehat{b}_q \iota_{K_2}}{\widehat{h}_q + 1}\right) \\ &= \left\{ \begin{array}{ll} \left(\frac{1}{h_0+1}\right)\widehat{b}\left(\beta\right) + \left(\frac{h_0}{h_0+1}\right)\widehat{b}_q \iota_{K_2} & \text{if} \quad h_0 \leq h^* \\ \left(\frac{1}{h^*+1}\right)\widehat{b}\left(\beta\right) + \left(\frac{h^*}{h^*+1}\right)\widehat{b}_q \iota_{K_2} & \text{if} \quad h_0 > h^*. \end{array} \right. \end{aligned}$$

The (unconditional) empirical Bayes estimator of b for the three-stage hierarchy model is thus given by

$$\begin{aligned} \widehat{b}_{EB}\left(\beta\right) &= \int_0^\infty \widehat{b}_{EB}\left(\beta \mid h_0\right) p(h_0) dh_0 \\ &= \frac{1}{B\left(c,d\right)} \left[\begin{array}{c} \widehat{b}\left(\beta\right) \int_0^{h^*} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_q \iota_{K_2} \int_0^{h^*} h_0^c \left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ + \left\{\widehat{b}\left(\beta\right)\left(\frac{1}{g^*+1}\right) + \widehat{b}_q \iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\} \int_{h^*}^\infty h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \end{array} \right]. \end{aligned}$$

Let $\varphi = \frac{h_0}{h_0+1}$. Then

$$\widehat{b}_{EB}\left(\beta\right) = \frac{1}{B\left(c,d\right)} \left[\begin{array}{c} \widehat{b}\left(\beta\right) \int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^{c-1}\left(1-\varphi\right)^d d\eta + \widehat{b}_q \iota_{K_2} \int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^c \left(1-\varphi\right)^{d-1} d\varphi \\ + \left\{\widehat{b}\left(\beta\right)\left(\frac{1}{h^*+1}\right) + \widehat{b}_q \iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\} \int_{\frac{h^*}{h^*+1}}^1 \left(\varphi\right)^{c-1}\left(1-\varphi\right)^{d-1} d\varphi \end{array} \right].$$

We get three incomplete Gaussian hypergeometric functions. Let $\varphi = \left(\frac{h^*}{h^*+1}\right)t$. The solution of the first one is given by

$$\begin{aligned} \int_0^{\frac{h^*}{h^*+1}} \left(\varphi\right)^{c-1}\left(1-\varphi\right)^d d\eta &= \int_0^1 \left(\left(\frac{h^*}{h^*+1}\right)t\right)^{c-1}\left(1-\left(\frac{h^*}{h^*+1}\right)t\right)^d \left(\frac{h^*}{h^*+1}\right)dt \\ &= \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times {}_2F_1\left(c; -d; c+1; \frac{h^*}{h^*+1}\right). \end{aligned}$$

The solution of the second one is:

$$\int_0^{\frac{h^*}{h^*+1}} (\varphi)^c (1-\varphi)^{d-1} d\varphi = \int_0^1 \left((\frac{h^*}{h^*+1})t \right)^c \left(1 - (\frac{h^*}{h^*+1})t \right)^{d-1} (\frac{h^*}{h^*+1})dt$$
$$= \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1 \left(c+1; 1-d; c+2; \frac{h^*}{h^*+1} \right),$$

and the solution of the third one is:

$$\int_{\frac{h^*}{h^*+1}}^1 (\varphi)^{c-1} (1-\varphi)^{d-1} d\varphi = \int_0^1 (\varphi)^{c-1} (1-\varphi)^{d-1} d\varphi - \int_0^{\frac{h^*}{h^*+1}} (\varphi)^{c-1} (1-\varphi)^{d-1} d\varphi$$
$$= B(c,d) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1 \left(c; d-1; c+1; \left(\frac{h^*}{h^*+1} \right) \right).$$

It follows the empirical Bayes estimator of b for the three-stage hierarchy model is given by:

$$\widehat{b}_{EB}(\beta) = \frac{1}{B(c,d)} \left[\begin{array}{c} \widehat{b}(\beta) \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1 \left(c; -d; c+1; \frac{h^*}{h^*+1} \right) \\ +\widehat{b}_{q} \iota_{K_2} \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1 \left(c+1; 1-d; c+2; \frac{h^*}{h^*+1} \right) \\ + \left\{ \widehat{b}(\beta) \left(\frac{1}{h^*+1} \right) + \widehat{b}_{q} \iota_{K_2} \left(\frac{h^*}{h^*+1} \right) \right\} \\ \times \left[\begin{array}{c} B(c,d) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2 F_1 \left(c; d-1; c+1; \frac{h^*}{h^*+1} \right) \end{array} \right] \end{array} \right]. \quad (44)$$

C Laplace approximations

C.1 Laplace approximation of the predictive density based on the base prior

The unconditional predictive density corresponding to the base prior is given by

$$m(\widetilde{y} \mid \pi_0, \beta) = \int_0^\infty m(\widetilde{y} \mid \pi_0, \beta, h_0) .p(h_0) dh_0$$
$$= \frac{\widetilde{H}}{B(c,d)} \times \int_0^\infty \left\{ \begin{array}{c} \left(\frac{h_0}{h_0+1}\right)^{K_2/2} \left(1 + \left(\frac{h_0}{h_0+1}\right) \left(\frac{R^2_{b_0}}{1-R^2_{b_0}} \right) \right)^{-\frac{NT}{2}} \\ \times h_0^{c-1} \left(\frac{1}{1+h_0} \right)^{c+d} \end{array} \right\} dh_0$$
$$= \frac{B(d, \frac{K_2}{2}+c)}{B(c,d)} \widetilde{H} \times_2 F_1 \left(\frac{NT}{2}; \frac{K_2}{2}+c; \frac{K_2}{2}+c+d; -\left(\frac{R^2_{b_0}}{1-R^2_{b_0}} \right) \right).$$

As shown by Liang *et al.* (2008), numerical overflow is problematic for moderate to large NT and large $R^2_{b_0}$ in Gaussian hypergeometric functions. As the Laplace approximation involves an integral with respect to a normal kernel, we follow the suggestion of Liang *et al.* (2008) to develop the expansion after a change of variables to $\phi = \log \left(\frac{h_0}{h_0+1} \right)$. Thus $\frac{1}{h_0+1} = (1 - \exp[\phi])$, $h_0 = \frac{\exp[\phi]}{1-\exp[\phi]}$ and $dh_0 = \frac{\exp[\phi]}{(1-\exp[\phi])^2} d\phi$. Then:

$$m(\widetilde{y} \mid \pi_0, \beta) = \frac{\widetilde{H}}{B(c,d)} \int_{-\infty}^0 \exp\left[\phi \left(\frac{K_2}{2} + c \right) \right] (1-\exp[\phi])^{d-1} \left(1 + \exp[\phi] . \left(\frac{R^2_{b_0}}{1-R^2_{b_0}} \right) \right)^{-\frac{NT}{2}} d\phi \quad (1)$$

Let $l(\phi)$ be the logarithm of the integrand function of (1):

$$l(\phi) = \phi\left(\frac{K_2}{2}+c\right) + (d-1)\log\left(1-\exp[\phi]\right) - \frac{NT}{2}\log\left[1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]. \qquad (2)$$

The Laplace approximation is given by:

$$\int_{-\infty}^{0} \exp[l(\phi)]\,d\phi \simeq \sqrt{2\pi}.\widehat{\sigma}_l.\exp\left[l\left(\widehat{\phi}\right)\right], \text{ with } \widehat{\sigma}_l^2 = \left(-\left.\frac{d^2 l(\phi)}{d\phi^2}\right|_{\phi=\widehat{\phi}}\right)^{-1}. \qquad (3)$$

Setting $l'(\phi)=0$ gives a quadratic equation in $\exp[\phi]$:

$$\begin{aligned} l'(\phi) &= \left(\frac{K_2}{2}+c\right) - (d-1)\frac{\exp[\phi]}{1-\exp[\phi]} - \frac{NT\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)}{2\left[1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]} = 0 \\ &= \frac{1}{Den}\left\{\begin{array}{c} 2\left(\frac{K_2}{2}+c\right)\left(1-\exp[\phi]\right)\left[1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right] \\ -2(d-1)\exp[\phi]\left[1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right] \\ -NT\exp[\phi]\left(1-\exp[\phi]\right)\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right) \end{array}\right\} = 0 \end{aligned} \qquad (4)$$

with $Den = 2\left(1-\exp[\phi]\right)\left[1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]$. As $Den \neq 0$, the quadratic equation in $\exp[\phi]$ is given by:

$$\begin{aligned} &\exp[2\phi]\left[\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\{NT-K_2-2(c+d)+2\}\right] \\ &\qquad -\exp[\phi]\left[\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\{NT-K_2-2c\}+K_2+2(c+d)\right]+K_2+2c=0. \end{aligned} \qquad (5)$$

The roots are given by:

$$\left\{\exp\left[\widehat{\phi}\right]\right\}_{1,2} = \frac{C_1 \pm \sqrt{\Delta}}{C_2}, \qquad (6)$$

with

$$\begin{aligned} C_1 &= \left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\{NT-K_2-2c\}+K_2+2(c+d) \\ C_2 &= 2\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\{NT-K_2-2(c+d)\} \\ \Delta &= [C_1]^2 + 2C_2\left[-2c-K_1\right]. \end{aligned} \qquad (7)$$

As $h_0 \in]0,+\infty[$, then $\exp\left[\widehat{\phi}\right] \in]0,1[$ and only one root is positive, so:

$$\exp\left[\widehat{\phi}\right] = \frac{C_1+\sqrt{\Delta}}{C_2}. \qquad (8)$$

The corresponding variance is

$$
\begin{aligned}
\widehat{\sigma}_l^2 &= \left(- \left. \frac{d^2 l(\phi)}{d\phi^2} \right|_{\phi=\widehat{\phi}} \right)^{-1} \\
&= \left[\begin{array}{c} \frac{(d-1)\exp\left[2\widehat{\phi}\right]}{(1-\exp\left[\widehat{\phi}\right])^2} + \frac{(d-1)\exp\left[\widehat{\phi}\right]}{(1-\exp\left[\widehat{\phi}\right])} \\ -\frac{NT\exp\left[2\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)}{2\left[1+\exp\left[\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]^2} + \frac{NT\exp\left[\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)}{2\left[1+\exp\left[\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]} \end{array} \right]^{-1} \\
&= \left[\frac{(d-1)\exp\left[\widehat{\phi}\right]}{(1-\exp\left[\widehat{\phi}\right])^2} + \frac{NT\exp\left[\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)}{2\left[1+\exp\left[\widehat{\phi}\right]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right]^2} \right]^{-1}. \qquad (9)
\end{aligned}
$$

Then, the Laplace approximation of the predictive density based on the base prior is:

$$
\begin{aligned}
m(\widetilde{y} \mid \pi_0, \beta) &= \frac{\widetilde{H}}{B(c,d)} \int_{-\infty}^{0} \left\{ \begin{array}{c} \exp\left[\phi\left(\frac{K_{21}}{2}+c\right)\right](1-\exp[\phi])^{d-1} \\ \times\left(1+\exp[\phi]\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right)^{-\frac{NT}{2}} \end{array} \right\} d\phi \\
&\simeq \frac{\widetilde{H}\sqrt{2\pi}}{B(c,d)} \widehat{\sigma}_l \exp\left[l\left(\widehat{\phi}\right)\right], \qquad (10)
\end{aligned}
$$

with $l\left(\widehat{\phi}\right)$ given by (2) and (8) and $\widehat{\sigma}_l$ given by (9).

C.1.1 Laplace approximation of the predictive density based on the contaminated prior

As

$$
\widehat{h} = \left\{ \begin{array}{lll} h_0 & \text{if} & h_0 \leq h^* \\ h^* & \text{if} & h_0 > h^* \end{array} \right., \qquad (11)
$$

with

$$
h^* = \max\left[0, \left[\left(\frac{NT-K_2}{K_2}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right) - 1\right]^{-1}\right] \qquad (12)
$$

then,

$$
\begin{aligned}
m(\widetilde{y} \mid \widehat{q}, \beta) &= \int_0^{\infty} m(\widetilde{y} \mid \widehat{q}, \beta, h_0).p(h_0) dh_0 \\
&= \frac{\widetilde{H}}{B(c,d)} \int_0^{h^*} \left\{ \begin{array}{c} \left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left[1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}} \\ \times h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} \end{array} \right\} dh_0 \\
&\quad + \frac{\widetilde{H}}{B(c,d)} \left(\frac{h^*}{h^*+1}\right)^{K_2/2} \left[1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}} \qquad (13) \\
&\quad \times \int_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \\
&= \frac{\widetilde{H}}{B(c,d)} \left[I_1 + \left(\frac{h^*}{h^*+1}\right)^{K_2/2} \left[1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}} I_2 \right].
\end{aligned}
$$

Let $l_1(\phi)$ be the logarithm of the integrand function of I_1, with $\phi = \log\left(\frac{h_0}{h_0+1}\right)$:

$$l_1(\phi) = \phi\left(\frac{K_2}{2} + c\right) + (d-1)\log(1-\exp[\phi]) - \frac{NT}{2}\log\left[1 + \exp[\phi]\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right] \tag{14}$$

As $l_1(\phi)$ is similar to $l(\phi)$ in (2) (except the ratio of $R^2_{b_q}$), we get the same quadratic equation in $\exp[\phi]$ and the same roots $\left\{\exp\left[\widehat{\phi}\right]\right\}_{1,2}$. As $h_0 \in]0, h^*]$, then $\exp\left[\widehat{\phi}\right] \in \left]0, \frac{g^*}{g^*+1}\right]$, so the only root should be positive and bounded by $\left]0, \frac{h^*}{h^*+1}\right]$, *i.e*, $\exp\left[\widehat{\phi}\right] = \frac{C_1+\sqrt{\Delta}}{C_2}$ in (8) should lie within $\left]0, \frac{h^*}{h^*+1}\right]$. The corresponding variance is similar to (9) and the Laplace approximation of I_1 is

$$I_1 \simeq \sqrt{2\pi}\widehat{\sigma}_{l_1}\exp\left[l_1\left(\widehat{\phi}\right)\right], \tag{15}$$

with $l_1\left(\widehat{\phi}\right)$ given by (14) and (8) and $\widehat{\sigma}_{l_1}$ given by (9). As

$$I_2 = \int_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 = \int_{\log\left(\frac{h^*}{h^*+1}\right)}^{0} \exp[c\phi](1-\exp[\phi])^{d-1} d\phi. \tag{16}$$

Let $l_2(\phi)$ be the logarithm of the integrand function of I_2:

$$l_2(\phi) = c\phi + (d-1)\log(1-\exp[\phi]). \tag{17}$$

Setting $l_2'(\phi) = 0$ gives a first order equation in $\exp[\phi]$:

$$l_2'(\phi) = c - (d-1)\frac{\exp[\phi]}{1-\exp[\phi]} = 0, \tag{18}$$

and the root is given by:

$$\exp\left[\widehat{\phi}\right] = \frac{c}{c+d-1}. \tag{19}$$

As $h_0 \in [h^*, \infty[$, then $\exp\left[\widehat{\phi}\right] \in \left[\frac{h^*}{h^*+1}, 1\right[$, so $d \in \left[1, \frac{c-h^*}{h^*}\right[$. The corresponding variance is

$$\widehat{\sigma}^2_{l_2} = \left(-\left.\frac{d^2 l_2(\phi)}{d\phi^2}\right|_{\phi=\widehat{\phi}}\right)^{-1} = \left[\frac{(d-1)\exp\left[\widehat{\phi}\right]}{(1-\exp\left[\widehat{\phi}\right])^2}\right]^{-1} \tag{20}$$

and the Laplace approximation of I_2 is

$$I_2 \simeq \sqrt{2\pi}\widehat{\sigma}_{l_2}\exp\left[l_2\left(\widehat{\phi}\right)\right], \tag{21}$$

with $l_2\left(\widehat{\phi}\right)$ given by (17) and (19) and $\widehat{\sigma}_{l_2}$ given by (20). Then, the Laplace approximation of the predictive density based on the contaminated prior is:

$$\begin{aligned} m(\widetilde{y} \mid \widehat{q}, \beta) = & \frac{\widetilde{H}}{B(c,d)}\int_0^{h^*}\left\{\begin{array}{c}\left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left[1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right]^{-\frac{NT}{2}} \\ \times\, h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d}\end{array}\right\} dh_0 \\ & + \frac{\widetilde{H}}{B(c,d)}\left(\frac{h^*}{h^*+1}\right)^{K_2/2}\left[1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right]^{-\frac{NT}{2}} \\ & \times \int_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \\ \simeq & \frac{\widetilde{H}}{B(c,d)}\left[\begin{array}{c}\sqrt{2\pi}\widehat{\sigma}_{l_1}\exp\left[l_1\left(\widehat{\phi}\right)\right] \\ +\left(\frac{g^*}{g^*+1}\right)^{K_1/2}\left[1+\left(\frac{g^*}{g^*+1}\right)\left(\frac{R^2_u}{1-R^2_u}\right)\right]^{-\frac{NT}{2}}\sqrt{2\pi}\widehat{\sigma}_{l_2}\exp\left[l_2\left(\widehat{\phi}\right)\right]\end{array}\right]. \end{aligned} \tag{22}$$

C.1.2 Laplace approximation of the empirical Bayes estimator

Under the contamination class of prior, the empirical Bayes estimator of β for the three-stage hierarchy model is given by:

$$\begin{aligned}
\widehat{b}_{EB}(\beta) &= \int_0^\infty \widehat{b}_{EB}(\beta \mid h_0)\, p(h_0) dh_0 \\
&= \frac{1}{B(c,d)} \left[\begin{array}{c} \widehat{b}(\beta) \int_0^{h^*} h_0^{c-1} \left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_{q} \iota_{K_2} \int_0^{h^*} h_0^{c} \left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ + \left\{ \widehat{b}(\beta) \left(\frac{1}{g^*+1}\right) + \widehat{b}_q \iota_{K_2} \left(\frac{h^*}{h^*+1}\right) \right\} \int_{h^*}^{\infty} h_0^{c-1} \left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \end{array} \right] \\
&= \frac{1}{B(c,d)} \left[\begin{array}{c} \widehat{b}(\beta) \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1\left(c; -d; c+1; \frac{h^*}{h^*+1}\right) \\ + \widehat{b}_q \iota_{K_2} \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1\left(c+1; 1-d; c+2; \frac{h^*}{h^*+1}\right) \\ + \left\{ \widehat{b}(\beta) \left(\frac{1}{h^*+1}\right) + \widehat{b}_q \iota_{K_2} \left(\frac{h^*}{h^*+1}\right) \right\} \\ \times \left[\begin{array}{c} B(c,d) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2 F_1\left(c; d-1; c+1; \frac{h^*}{h^*+1}\right) \end{array} \right] \end{array} \right].
\end{aligned}$$

Let us write

$$\widehat{b}_{EB}(\beta) = \frac{\widehat{\beta}(b)}{B(c,d)} D_1 + \frac{\widehat{b}_q \iota_{K_2}}{B(c,d)} D_2 + \frac{\left\{ \widehat{b}(\beta) \left(\frac{1}{g^*+1}\right) + \widehat{b}_q \iota_{K_2} \left(\frac{h^*}{h^*+1}\right) \right\}}{B(c,d)} D_3,$$

with

$$\begin{aligned}
D_1 &= \int_0^{h^*} h_0^{c-1} \left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 = \int_{-\infty}^{\log\left(\frac{h^*}{h^*+1}\right)} \exp[c.\phi] \left(1 - \exp[\phi]\right)^d d\phi \\
D_2 &= \int_0^{h^*} h_0^{c} \left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 = \int_{-\infty}^{\log\left(\frac{h^*}{h^*+1}\right)} \exp[(c+1).\phi] \left(1 - \exp[\phi]\right)^{d-1} d\phi \\
D_3 &= \int_{h^*}^{\infty} h_0^{c-1} \left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \equiv I_2 = \int_{\log\left(\frac{h^*}{h^*+1}\right)}^{0} \exp[c.\phi] \left(1 - \exp[\phi]\right)^{d-1} d\phi.
\end{aligned}$$

Let $l_{D_1}(\phi)$ be the logarithm of the integrand function of D_1:

$$l_{D_1}(\phi) = c\phi + d \log\left(1 - \exp[\phi]\right).$$

Setting $l'_{D_1}(\phi) = 0$ gives a first order equation in $\exp[\phi]$:

$$l'_{D_1}(\phi) = c - d \frac{\exp[\phi]}{1 - \exp[\phi]} = 0$$

and the root is given by:

$$\exp\left[\widehat{\phi}\right] = \frac{c}{c+d}.$$

As $h_0 \in]0, \infty[$, then $\exp\left[\widehat{\phi}\right] \in]0, \infty[$.The corresponding variance is

$$\widehat{\sigma}^2_{l_{D_1}} = \left(- \left. \frac{d^2 l_{D_1}(\phi)}{d\phi^2} \right|_{\phi = \widehat{\phi}} \right)^{-1} = \left[\frac{d \exp\left[\widehat{\phi}\right]}{(1 - \exp\left[\widehat{\phi}\right])^2} \right]^{-1}$$

and the Laplace approximation of D_1 is

$$D_1 \simeq \sqrt{2\pi}\widehat{\sigma}^2_{l_{D_1}} \exp\left[l_{D_1}\left(\widehat{\phi}\right)\right]$$

Let $l_{D_2}(\phi)$ be the logarithm of the integrand function of D_2:

$$l_{D_2}(\phi) = (c+1)\,\phi + (d-1)\log\left(1-\exp\left[\phi\right]\right).$$

Setting $l'_{D_2}(\phi) = 0$ gives a first order equation in $\exp[\phi]$:

$$l'_{D_2}(\phi) = (c+1) - (d-1)\,\frac{\exp[\phi]}{1-\exp[\phi]} = 0$$

and the root is given by:

$$\exp\left[\widehat{\phi}\right] = \frac{c+1}{c+d+2}.$$

The corresponding variance is

$$\widehat{\sigma}^2_{l_{D_2}} = \left(-\left.\frac{d^2 l_{D_2}(\phi)}{d\phi^2}\right|_{\phi=\widehat{\phi}}\right)^{-1} = \left[\frac{(d-1)\exp\left[\widehat{\phi}\right]}{(1-\exp\left[\widehat{\phi}\right])^2}\right]^{-1}$$

and the Laplace approximation of D_2 is

$$D_2 \simeq \sqrt{2\pi}\widehat{\sigma}^2_{l_{D_2}} \exp\left[l_{D_2}\left(\widehat{\phi}\right)\right].$$

For I_2, the Laplace approximation of I_2 is

$$I_2 \simeq \sqrt{2\pi}\widehat{\sigma}_{l_2} \exp\left[l_2\left(\widehat{\phi}\right)\right]$$

Then the Laplace approximation of the empirical Bayes estimator of β on the contaminated prior is:

$$\begin{aligned}
\widehat{b}_{EB}(\beta) &= \frac{1}{B(c,d)}\left[\begin{array}{c} \widehat{b}(\beta)\int_0^{h^*} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_q \iota_{K_2}\int_0^{h^*} h_0^{c}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ +\left\{\widehat{b}(\beta)\left(\frac{1}{g^*+1}\right) + \widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\}\int_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \end{array}\right] \\
&\simeq \frac{\widehat{\beta}(b)}{B(c,d)}.\sqrt{2\pi}\widehat{\sigma}^2_{l_{D_1}}\exp\left[l_{D_1}\left(\widehat{\phi}\right)\right] + \frac{\widehat{b}_q\iota_{K_2}}{B(c,d)}\sqrt{2\pi}.\widehat{\sigma}^2_{l_{D_2}}\exp\left[l_{D_2}\left(\widehat{\phi}\right)\right] \\
&+ \frac{\left\{\widehat{b}(\beta)\left(\frac{1}{g^*+1}\right) + \widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\}}{B(c,d)}\sqrt{2\pi}\widehat{\sigma}_{l_2}\exp\left[l_2\left(\widehat{\phi}\right)\right].
\end{aligned} \tag{23}$$

D The minimum chi-square estimator

For the Chamberlain world, we have used the Minimum Chi-Square (MCS) estimator. Let $y_i = (y_{i1}, ..., y_{iT})'$ a $(T \times 1)$ vector and $x'_i = (x'_{i1}, ..., x'_{iT})$ a $(1 \times TK)$ matrix. Let us consider for generalization that $x'_{it} = \left[x^{(0)'}_{it}, x^{(1)'}_{it}, x^{(2)'}_{it}\right]$ where $x^{(0)'}_{it}$ is a $(1 \times K_0)$ vector of intercept and dummies, $x^{(1)'}_{it}$ is a $(1 \times K_1)$ vector of variables uncorrelated with μ_i and $x^{(2)'}_{it}$ is a $(1 \times K_2)$ vector of variables correlated with μ_i where $K = K_0 + K_1 + K_2$. In our simulation study: $K_0 = 0$, $x^{(1)'}_{it} = [x_{1,1,it}, x_{1,2,it}]$ is a $(1 \times K_1 (\equiv 2))$ vector of variables uncorrelated with μ_i and $x^{(2)'}_{it} = [x_{2,it}]$ is a $(1 \times K_2 (\equiv 1))$ vector of variables correlated with μ_i. The model is given by

$$y_{it} = x'_{it}\beta + \mu_i + \varepsilon_{it} \tag{24}$$

with

$$\mu_i = x^{(2)'}_{i1}\pi_1 + x^{(2)'}_{i2}\pi_2 + ... + x^{(2)'}_{iT}\pi_T + \nu_i \tag{25}$$

where π_t is a $(K_2 \times 1)$ vector of parameters. Substituting (24) into (25) gives for each t:

$$\begin{aligned} y_{it} &= x_{it}'\beta + x_{i1}^{(2)'}\pi_1 + x_{i2}^{(2)'}\pi_2 + ... + x_{iT}^{(2)'}\pi_T + r_{it} \\ &= x_{it}^{(0)'}\beta_0 + x_{it}^{(1)'}\beta_1 + x_{i1}^{(2)'}\pi_1 + x_{i2}^{(2)'}\pi_2 + ... + x_{it}^{(2)'}(\beta_2 + \pi_t) + ... + x_{iT}^{(2)'}\pi_T + r_{it} \\ &= x_{it}^{(0)'}\beta_0 + x_{it}^{(1)'}\beta_1 + x_{i}^{(2)'}\Pi_t + r_{it}, \end{aligned} \tag{26}$$

where $r_{it} = \nu_i + \varepsilon_{it}$, $x_i^{(2)'} = \left[x_{i1}^{(2)'}, x_{i2}^{(2)'}, ..., x_{iT}^{(2)'}\right]$ and $\Pi_t = \left(\pi_1', \pi_2', ..., (\beta_2 + \pi_t)', ..., \pi_T'\right)'$. The "reduced form" can be expressed as (see Chamberlain (1982), Hsiao (2003)):

$$y_i = X_i \Pi + r_i, \tag{27}$$

with y_i a $(1 \times T)$ vector, X_i a $(1 \times T[K_0 + K_1 + TK_2])$ vector and Π, a $(T[K_0 + K_1 + TK_2] \times T)$ matrix:

$$X_i = \left(\begin{array}{cccc} \left[x_{i1}^{(0)'}, x_{i1}^{(1)'}, x_i^{(2)'}\right] & \left[x_{i2}^{(0)'}, x_{i2}^{(1)'}, x_i^{(2)'}\right] & \cdots & \left[x_{iT}^{(0)'}, x_{iT}^{(1)'}, x_i^{(2)'}\right] \end{array} \right), \tag{28}$$

$$\Pi = \left(\begin{array}{cccc} \beta_0 & \beta_0 & \cdots & \beta_0 \\ \beta_1 & \beta_1 & \cdots & \beta_1 \\ \beta_2 + \pi_1 & \pi_1 & \cdots & \pi_1 \\ \pi_2 & \beta_2 + \pi_2 & \cdots & \pi_2 \\ \vdots & \vdots & \ddots & \vdots \\ \pi_T & \pi_T & \cdots & \beta_2 + \pi_T \end{array} \right). \tag{29}$$

The $([K_0 + K_1 + (T+1)K_2] \times 1)$ parameter vector of interest $\theta = (\beta_0', \beta_1', \beta_2', \pi_1', \pi_2', ..., \pi_T')'$, from the structural model is known to be related to the $(T[K_0 + K_1 + TK_2] \times T)$ matrix of reduced form parameters Π. In particular: $\text{vec}(\Pi) = h(\theta)$ for a known continuously differentiable function $h(.)$. CMS estimation of θ entails first estimating Π by $\widehat{\Pi}$ and then choosing an estimator $\widehat{\theta}$ of θ by making the distance between $\text{vec}\left(\widehat{\Pi}\right)$ and $h(\widehat{\theta})$ as small as possible. As with GMM, the CMS estimator uses an efficient weighted Euclidian measure of distance. Assuming that for an $(S \times S)$ definite positive matrix Ω

$$\sqrt{N}\text{vec}\left(\widehat{\Pi} - \Pi\right) \overset{asympt.}{\sim} N(0, \Omega), \tag{30}$$

with $S = T[K_0 + K_1 + TK_2]$, it turns out that the CMS solves

$$\min_{\theta} \left(\text{vec}\left(\widehat{\Pi}\right) - h(\theta)\right)' \widehat{\Omega}^{-1} \left(\text{vec}\left(\widehat{\Pi}\right) - h(\theta)\right). \tag{31}$$

The restrictions between the reduced form and structural parameters are given by

$$\text{vec}(\Pi) = h(\theta) = (I_T \otimes \beta) + \lambda \iota_T', \tag{32}$$

where $\beta = (\beta_0', \beta_1', \beta_2')'$, $\lambda = (0_0', 0_1', \pi_1', \pi_2', ..., \pi_T')'$ with 0_j a $(K_j \times 1)$ vector of zeros $(j = 0, 1)$. The appropriate estimator of the asymptotic variance-covariance matrix of $\widehat{\theta}$ is

$$\text{Avar}\left(\widehat{\theta}\right) = \frac{1}{N}\left(H'\widehat{\Omega}^{-1}H\right)^{-1} = \frac{1}{N}\left(H'\left[\text{Avar}\left(\text{vec}\left(\widehat{\Pi}\right)\right)\right]^{-1}H\right)^{-1}, \tag{33}$$

where $H = \frac{\partial h(\theta)}{\partial \theta'}$ is the $(S \times S)$ Jacobian of $h(\theta)$, *i.e.*, all 1s and 0s.

The Jacobian H is then defined as:

$$H = \begin{pmatrix} I_{K_0} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & I_{K_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & I_{K_2} & I_{K_2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & I_{K_2} & \cdots & 0 \\ & & & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & \cdots & I_{K_2} \\ I_{K_0} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & I_{K_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & I_{K_2} & 0 & \cdots & 0 \\ 0 & 0 & I_{K_2} & 0 & I_{K_2} & \cdots & 0 \\ & & & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & 0 & I_{K_2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ I_{K_0} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & I_{K_1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & I_{K_2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & I_{K_2} & \cdots & 0 \\ & & & & & \ddots & \\ 0 & 0 & I_{K_2} & 0 & 0 & \cdots & I_{K_2} \end{pmatrix}. \tag{34}$$

The estimator $\widehat{\Omega}$ of Avar$\sqrt{N}$vec$\left(\widehat{\Pi} - \Pi\right)$ is the robust asymptotic variance for system OLS:

$$\widehat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \left[\widehat{r}_i \widehat{r}_i' \otimes S_{xx}^{-1} \underline{X}_i \underline{X}_i' S_{xx}^{-1}\right], \tag{35}$$

where $\widehat{r}_i$ is the vector of the OLS residuals, $\underline{X}_i$ is the $([K_0 + K_1 + TK_2] \times T)$ matrix of X_i and $S_{xx} = \sum_{i=1}^{N} \left(\underline{X}_i \underline{X}_i'\right)/N$ (see Chamberlain (1982), Wooldridge (2002), Hsiao (2003)). If the conditional variance-covariance matrix of r_i is uncorrelated with $\underline{X}_i$ and is homoskedastic, then the estimator $\widehat{\Omega}$ will converge to

$$\widehat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \widehat{r}_i \widehat{r}_i' \otimes \left(\frac{1}{N} \sum_{i=1}^{N} \underline{X}_i \underline{X}_i'\right)^{-1}. \tag{36}$$

After the estimation of $\widehat{\theta}$, $\widehat{\Omega}$ is recalculated to estimate Avar$\left(\widehat{\theta}\right)$ and MCS can be iterated.

E The Hausman-Taylor estimator

For the Hausman-Taylor world, we used the IV method proposed by Hausman and Taylor (1981). For our model, $y_{it} = x_{1,1,it}\beta_{1,1} + x_{1,2,it}\beta_{1,2} + x_{2,it}\beta_2 + Z_{1,i}\eta_1 + Z_{2,i}\eta_2 + \mu_i + \varepsilon_{it}$ or $y = X_1\beta_1 + x_2\beta_2 + Z_1\eta_1 + Z_2\eta_2 + Z_\mu\mu + \varepsilon$. The HT procedure is defined by the following two-step consistent estimator of β and η (see Baltagi (2013)):

1. Perform the fixed effects (FE) or Within estimator obtained by regressing $\widetilde{y}_{it} = (y_{it} - \overline{y}_{i.})$, where $\overline{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$, on a similar within transformation of the regressors. Note that the Within transformation wipes out the Z_i variables since they are time invariant, and we only obtain an estimate of β which we denote by $\widetilde{\beta}_W$.

   - HT next averages the within residuals over time

     $$\widehat{d}_i = \overline{y}_{i.} - \bar{X}_{i.}' \widetilde{\beta}_W,$$

     where $\bar{X}_{i.}'$ is the vector of individual means $\bar{X}_{i.}' = [\overline{x}_{11i.}, \overline{x}_{12i.}, \overline{x}_{2i.},]$.
   - To get an estimate of η, HT suggest running a 2SLS of $\widehat{d}_i$ on $Z_i = [Z_{1,i}, Z_{2,i}]$ with the set of instruments $A = [X_1, Z_1]$ where $X_1 = [x_{1,1}, x_{1,2,}]$. This yields

     $$\widehat{\eta}_{2SLS} = (Z'P_AZ)^{-1}Z'P_A\widehat{d},$$

     where $P_A = A(A'A)^{-1}A'$.

2. HT suggest estimating the variance-components as follows:

$$\widehat{\sigma}_{\varepsilon}^2 = (y_{it} - X_{it}'\widetilde{\beta}_W)' Q_W (y_{it} - X_{it}'\widetilde{\beta}_W)/N(T-1)$$

and

$$\widehat{\sigma}_1^2 = (y_{it} - X_{it}'\widetilde{\beta}_W - Z_i'\eta_{2SLS})' P (y_{it} - X_{it}'\widetilde{\beta}_W - Z_i'\eta_{2SLS})/N,$$

where $\sigma_1^2 = T\sigma_{\mu}^2 + \sigma_{\varepsilon}^2$. Once the variance-components estimates are obtained, the model is transformed using $\widehat{\Omega}^{-1/2}$ where

$$\Omega^{-1/2} = \frac{1}{\sigma_1} P + \frac{1}{\sigma_{\varepsilon}} Q_W.$$

Note that $y^* = \widehat{\sigma}_{\varepsilon}\widehat{\Omega}^{-1/2} y$ has a typical element $y_{it}^* = y_{it} - \widehat{\theta}\bar{y}_{i.}$, where $\widehat{\theta} = 1 - (\widehat{\sigma}_{\varepsilon}/\widehat{\sigma}_1)$ and X_{it}^* and Z_i^* are defined similarly. In fact, the transformed regression becomes:

$$\widehat{\sigma}_{\varepsilon}\widehat{\Omega}^{-1/2} y_{it} = \widehat{\sigma}_{\varepsilon}\widehat{\Omega}^{-1/2} X_{it}\beta + \widehat{\sigma}_{\varepsilon}\widehat{\Omega}^{-1/2} Z_i\eta + \widehat{\sigma}_{\varepsilon}\widehat{\Omega}^{-1/2} u_{it},$$

where $u_{it} = \mu_i + \varepsilon_{it}$. The asymptotically efficient HT estimator is obtained by running a 2SLS on this transformed model using $A_{HT} = [\widetilde{X}, \bar{X}_1, Z_1]$ as the set of instruments. In this case, $\widetilde{X}$ denotes the within transformed X and $\bar{X}_1$ denotes the time average of X_1. More formally, the HT estimator under over-identification is given by:

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\eta} \end{pmatrix}_{HT} = \left[\begin{pmatrix} X^{*\prime} \\ Z^{*\prime} \end{pmatrix} P_{A_{HT}} (X^*, Z^*) \right]^{-1} \begin{pmatrix} X^{*\prime} \\ Z^{*\prime} \end{pmatrix} P_{A_{HT}} y^*,$$

where $P_{A_{HT}}$ is the projection matrix on $A_{HT} = [\widetilde{X}, \bar{X}_1, Z_1]$.

F References

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

Appell, P., 1882, Sur les fonctions hypergéométriques de deux variables, *Journal de Mathématiques Pures et Appliquées,* 3ème série, (in French) , 8, 173–216. http://portail.mathdoc.fr/JMPA/PDF/JMPA_1882_3_8_A8_0.pdf.

Baltagi, B.H., 2013, *Econometric Analysis of Panel Data*, fifth edition, Wiley, Chichester, UK.

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

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

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

Hsiao, C., 2003, *Analysis of Panel Data*, second edition, Cambridge University Press, Cambridge, MA.

Hsiao, C. and M.H. Pesaran, 2008, Random coefficient models, in *The Handbook of Panel Data, Fundamentals and Recent Developments in Theory and Practice* (Matyas, L. and P. Sevestre, eds), Springer, 185-214.

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

Liang, F., Paulo, R., Molina, G., Clyde, M.A. and J.O. Berger, 2008, Mixtures of g priors for Bayesian variable selection, *Journal of the American Statistical Association*, 103, 481, 410-423.

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

Wooldridge, J.M., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.