Robust linear static panel data models using ε-contamination

Supplementary Material

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

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

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 means 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.

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

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

Contents

A. Tables

List of Tables

Table A1: Random Effects World

FGLS, Robust 3S with individual block resampling bootstrap

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

(N,T,ρ)		β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	$\widehat{\lambda}_\beta$	$\widehat{\lambda}_\mu$
	True	1	1	1	1	0.4286		
(100,10,0.3)	FGLS coef	0.9989	1.0000	0.9998	0.9994	0.4311		
	se	0.0151	0.0148	0.0153				
	rmse	0.0143	0.0143	0.0155				
	3S bootstrap coef	0.9983	0.9994	0.9990	0.9967	0.4356	0.1282	$< 10^{-4}$
	se	0.0174	0.0174	0.0179				
	rmse	0.0178	0.0172	0.0191				
(500,5,0.3)	FGLS coef	0.9999	1.0000	1.0003	1.0004	0.4261		
	se	0.0081	0.0080	0.0079				
	rmse	0.0080	0.0086	0.0079				
	3S bootstrap coef	0.9995	0.9993	0.9994	0.9994	0.4275	0.1321	$< 10^{-4}$
	se	0.0103	0.0101	0.0101				
	rmse	0.0107	0.0109	0.0104				
	True	1	1	1	1	4		
(100,10,0.8)	FGLS coef	0.9986	1.0005	0.9998	0.9994	4.0143		
	se	0.0229	0.0228	0.0230				
	rmse	0.0224	0.0220	0.0236				
	3S bootstrap coef	0.9983	0.9999	0.9994	0.9969	3.9878	0.0781	$< 10^{-4}$
	se	0.0222	0.0225	0.0225				
	rmse	0.0226	0.0224	0.0242				
(500,5,0.8)	FGLS coef	0.9998	1.0000	1.0002	1.0004	3.9844		
	se	0.0151	0.0149	0.0149				
	rmse	0.0148	0.0159	0.0148				
	3S bootstrap coef	0.9995	0.9996	0.9998	0.9994	3.9761	0.0637	$< 10^{-4}$
	se	0.0149	0.0147	0.0147				
	rmse	0.0152	0.0160	0.0151				

Table A2: Fixed Effects Worlds

$N = 100, 500$, $T = 5, 10, \varepsilon = 0.5$, replications=1,000

		Mundlak-type Fixed Effects World†						
(N,T)		β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_β	λ_μ
	True	1	1	1	1	6.3334		
(100,10)	Mundlak coef	0.9987	0.9992	1.0001	1.0001	6.1968		
	se	0.0184	0.0184	0.0236				
	rmse	0.0184	0.0182	0.0240				
	3S bootstrap coef	0.9981	0.9989	0.9999	0.9938	6.2109	0.0002	$< 10^{-4}$
	se	0.0192	0.0193	0.0230				
	rmse	0.0201	0.0199	0.0244				
	True	1	1	1	1	6.3542		
(500,5)	Mundlak coef	1.0002	0.9998	0.9995	0.9978	6.3647		
	se	0.0104	0.0103	0.0158				
	rmse	0.0099	0.0102	0.0158				
	3S bootstrap coef	0.9997	0.9997	0.9997	0.9979	6.3666	$< 10^{-4}$	$< 10^{-4}$
	se	0.0116	0.0116	0.0153				
	rmse	0.0119	0.0119	0.0161				

		Chamberlain-type Fixed Effects World‡						
(N,T)		β_{11}	β_{12}	β_2	σ_u^2	σ_μ^2	λ_β	λ_μ
	True	1	1	1	1	165.2058		
(100,10)	MCS coef	1.0000	1.0019	0.9985	1.0092	170.1977		
	se	0.0236	0.0236	0.0224				
	rmse	0.0280	0.0278	0.0261				
	3S bootstrap coef	0.9968	0.9957	0.9981	0.9968	164.8366	$< 10^{-4}$	$< 10^{-4}$
	se	0.0198	0.0200	0.0229				
	rmse	0.0219	0.0213	0.0252				
	True	1	1	1	1	96.4161		
(500,5)	MCS coef	0.9996	0.9996	0.9994	1.0024	97.7024		
	se	0.0117	0.0117	0.0157				
	rmse	0.0118	0.0113	0.0159				
	3S bootstrap coef	0.9984	0.9987	0.9992	0.9972	95.9780	$< 10^{-4}$	$< 10^{-4}$
	se	0.0119	0.0118	0.0153				
	rmse	0.0121	0.0121	0.0160				

† The parameter π is omitted from the table. ‡ The parameters $\pi_1, \ldots \pi_5$ are omitted from the table.

Table A3: Hausman-Taylor World

IV, Robust 3S with individual block resampling bootstrap

$N = 100, 500$, $T = 5, 10$, $\varepsilon = 0.5$, replications=1,000

(N,T,ρ)		β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ
	True	1	1	1	1	1	1	0.4286		
(100,10,0.3)	HT coef	1.0001	1.0011	1.0005	1.0094	0.9967	0.9955	0.4494		
	se	0.0210	0.0210	0.0256	0.0754	0.0809				
	rmse	0.0209	0.0197	0.0258	0.0739	0.0781				
	3S bootstrap coef	0.9960	0.9975	0.9996	1.0070	1.0216	0.9945	0.4103	$< 10^{-4}$	$< 10^{-4}$
	se	0.0172	0.0172	0.0248	0.0445	0.0401				
	rmse	0.0188	0.0177	0.0266	0.0716	0.0461				
(500,5,0.3)	HT coef	0.9999	1.0003	0.9995	0.9990	0.9989	0.9990	0.4459		
	se	0.0152	0.0152	0.0199	0.0362	0.0572				
	rmse	0.0153	0.0152	0.0206	0.0349	0.0562				
	3S bootstrap coef	0.9935	0.9936	0.9995	0.9986	1.0388	0.9988	0.4042	$< 10^{-4}$	$< 10^{-4}$
	se	0.0110	0.0110	0.0195	0.0247	0.0246				
	rmse	0.0132	0.0131	0.0209	0.0344	0.0466				
	True	1	1	1	1	1	1	4		
(100,10,0.8)	HT coef	0.9990	1.0010	1.0016	1.0223	0.9948	0.9955	4.1497		
	se	0.0243	0.0243	0.0256	0.2078	0.1508				
	rmse	0.0244	0.0230	0.0262	0.2025	0.1554				
	3S bootstrap coef	0.9957	0.9973	0.9997	1.0204	1.0284	0.9945	3.7318	$< 10^{-4}$	$< 10^{-4}$
	se	0.0174	0.0174	0.0248	0.0490	0.0388				
	rmse	0.0190	0.0180	0.0266	0.1931	0.0484				
(500,5,0.8)	HT coef	0.9995	1.0001	1.0000	0.9957	0.9979	0.9990	4.0931		
	se	0.0176	0.0177	0.0199	0.0929	0.0799				
	rmse	0.0177	0.0177	0.0206	0.0911	0.0775				
	3S bootstrap coef	0.9920	0.9921	0.9995	0.9955	1.0501	0.9989	3.6510	$< 10^{-4}$	$< 10^{-4}$
	se	0.0111	0.0111	0.0195	0.0265	0.0240				
	rmse	0.0141	0.0140	0.0209	0.0869	0.0561				

Table A4: Random Effects World, Chamberlain-type Fixed Effects World and Hausman-Taylor World, replications=1,000

($N=500$, $T=5$, $\varepsilon=0.5$, $\rho=0.8$)		β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	$\widehat{\lambda}_\beta$	$\widehat{\lambda}_\mu$	Times (sec.)
Random Effects World	True	1	1	1			1	4			
	FGLS coef	0.9998	1.0000	1.0002			1.0004	3.9844			127.724
	se	0.0151	0.0149	0.0149							
	rmse	0.0148	0.0159	0.0148							
	3S bootstrap coef	0.9995	0.9996	0.9998			0.9994	3.9761	0.0637	$< 10^{-4}$	11431.7
	se	0.0149	0.0147	0.0147							
	rmse	0.0152	0.0160	0.0151							
	FB coef	0.9998	1.0000	1.0002			0.9995	3.9724			36578.9
	se	0.0149	0.0159	0.0148							
	rmse	0.0149	0.0159	0.0148							
	nse	0.0007	0.0007	0.0007							
Chamberlain-type Fixed Effects World†	True	1	1	1			1	96.4383			
	MCS coef	0.9993	0.9995	1.0004			1.0003	95.9580			73.811
	se	0.0117	0.0117	0.0157							
	rmse	0.0118	0.0116	0.0161							
	3S bootstrap coef	0.9987	0.9991	1.0001			1.0004	96.3790	$< 10^{-4}$	$< 10^{-4}$	8666.704
	se	0.0119	0.0118	0.0153							
	rmse	0.0121	0.0121	0.0160							
	FB coef	0.9993	0.9997	1.0004			0.9994	96.4231			47197.96
	se	0.0107	0.0103	0.0160							
	rmse	0.0108	0.0103	0.0160							
	nse	0.0005	0.0005	0.0007							
Hausman-Taylor World	True	1	1	1	1	1	1	4			
	HT coef	0.9995	1.0001	1.0000	0.9957	0.9979	0.9990	4.0931			104.151
	se	0.0176	0.0177	0.0199	0.0929	0.0799					
	rmse	0.0177	0.0177	0.0206	0.0911	0.0775					
	3S bootstrap coef	0.9920	0.9921	0.9995	0.9955	1.0501	0.9989	3.6510	$< 10^{-4}$	$< 10^{-4}$	8122.653
	se	0.0111	0.0111	0.0195	0.0265	0.0240					
	rmse	0.0141	0.0140	0.0209	0.0869	0.0561					
	FB coef	0.9845	0.9848	1.0004	0.9993	1.0664	0.9999	3.4951			50299.04
	se	0.0078	0.0076	0.0197	0.0868	0.0183					
	rmse	0.0173	0.0170	0.0197	0.0867	0.0689					
	nse	0.0004	0.0004	0.0009	0.0016	0.0010					

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

Table A5: Hausman-Taylor World

IV, Robust 3S with individual block resampling bootstrap

$N = 100$, $T = 5$, $\varepsilon = 0.5$ or $\varepsilon = 0$, $\rho = 0.8$, replications=1,000

		β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	$\widehat{\lambda}_\beta$	$\widehat{\lambda}_\mu$
	True	1	1	1	1	1	1	4		
HT	coef	0.9976	0.9981	1.0037	1.0173	0.9958	0.9924	4.2904		
	se	0.0395	0.0394	0.0444	0.2143	0.1857				
	rmse	0.0405	0.0385	0.0456	0.2071	0.1903				
3S bootstrap	coef	0.9905	0.9921	1.0002	1.0136	1.0453	0.9947	3.7420	$< 10^{-4}$	$< 10^{-4}$
$\varepsilon = 0.5$, $g_0 = h_0 = 1/NT$	se	0.0375	0.0356	0.0628	0.2051	0.0799				
	rmse	0.0387	0.0365	0.0628	0.2055	0.0919				
3S bootstrap	coef	0.9944	0.9928	0.9995	0.9915	1.0429	0.9896	3.7437	$< 10^{-4}$	$< 10^{-4}$
$\varepsilon = 0.5$, $g_0 = h_0 = 0.1$	se	0.0372	0.0358	0.0627	0.2061	0.0789				
	rmse	0.0376	0.0365	0.0627	0.2063	0.0898				
3S bootstrap	coef	0.9854	0.9839	0.9976	0.9725	1.0084	0.9904	3.3833	1	1
$\varepsilon = 0$, $g_0 = h_0 = 1/NT$	se	0.0355	0.0341	0.0626	0.1943	0.0764				
	rmse	0.0384	0.0377	0.0627	0.1963	0.0765				
3S bootstrap	coef	0.9507	0.9492	0.9897	0.8968	0.8646	0.9967	3.3027	1	1
$\varepsilon = 0$, $g_0 = h_0 = 0.1$	se	0.0346	0.0334	0.0624	0.1540	0.0750				
	rmse	0.0602	0.0607	0.0632	0.1853	0.1548				

Table A6: Departure from Normality: The Skewed t-distribution

mean= 0, 5 degrees of freedom, shape=3

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

	β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ
MUNDLAK-TYPE FIXED EFFECTS WORLD†									
True	1	1	1			6.0051	6.3257		
Mundlak Coef	1.0008	1.0028	1.0018			7.0574	6.3826		
se	0.0366	0.0366	0.0938						
rmse	0.0361	0.0363	0.0929						
3S bootstrap coef	1.0002	1.0028	1.0021			7.0137	6.3452	0.0445	$< 10^{-4}$
se	0.0361	0.0365	0.0909						
rmse	0.0369	0.0374	0.0949						
CHAMBERLAIN-TYPE FIXED EFFECTS WORLD‡									
True	1	1	1			7.7044	17.1421		
MCS coef	1.0007	0.9998	1.0012			7.0824	16.0968		
se	0.0354	0.0353	0.0893						
rmse	0.0389	0.0397	0.0942						
3S bootstrap coef	1.0005	0.9979	1.0003			7.0942	16.8421	0.0003	$< 10^{-4}$
se	0.0382	0.0386	0.0907						
rmse	0.0392	0.0404	0.0954						
HAUSMAN-TAYLOR WORLD									
True	1	1	1	1	1	7.1890	32		
HT coef	0.9892	0.9874	1.0264	0.9726	1.0002	7.1341	48.7261		
se	0.1037	0.1040	0.1136	0.7053	0.6348				
rmse	0.1047	0.1031	0.1239	0.6616	0.6471				
3S bootstrap coef	0.9936	0.9899	1.0056	0.9590	1.0427	7.1768	33.5738	$< 10^{-4}$	$< 10^{-4}$
se	0.0988	0.0972	0.1619	0.6514	0.2008				
rmse	0.0990	0.0977	0.1620	0.6527	0.2053				

† The parameter π is omitted.

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

Table A7: Departure from Normality: The χ^2 distribution

2 degrees of freedom

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

	β_{11}	β_{12}	β_2	η_1	η_2	σ_u^2	σ_μ^2	λ_β	λ_μ
Mundlak-type Fixed Effects World†									
True	1	1	1			4.0093	6.2890		
Mundlak Coef	1.0021	1.0000	0.9999			3.9880	6.3339		
se	0.0312	0.0313	0.0707						
rmse	0.0307	0.0315	0.0701						
3S bootstrap coef	1.0017	0.9999	0.9988			3.9803	6.2969	0.01504	$< 10^{-4}$
se	0.0304	0.0305	0.0693						
rmse	0.0312	0.0317	0.0720						
Chamberlain-type Fixed Effects World‡									
True	1	1	1			4.0054	17.0676		
MCS coef	0.9983	0.9997	1.0036			4.0044	17.4551		
se	0.0309	0.0309	0.0679						
rmse	0.0323	0.0331	0.0737						
3S bootstrap coef	0.9972	0.9978	1.0025			3.9869	16.8611	$< 10^{-4}$	$< 10^{-4}$
se	0.0311	0.0317	0.0686						
rmse	0.0326	0.0326	0.0750						
Hausman-Taylor World									
True	1	1	1	1	1	4.0233	32		
HT coef	0.9874	0.9911	1.0172	0.9708	1.0244	3.9964	41.7417		
se	0.0802	0.0806	0.0859	0.6561	0.5448				
rmse	0.0849	0.0829	0.0955	0.6184	0.5117				
3S bootstrap coef	0.9908	0.9911	1.0005	0.9686	1.0480	3.9969	32.0724	$< 10^{-4}$	$< 10^{-4}$
se	0.0744	0.0738	0.1227	0.6045	0.1533				
rmse	0.07504	0.0744	0.1227	0.6052	0.1606				

† The parameter π is omitted.

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

B. Applications

The simulation results have shown that our proposed estimators behave as well if not better than the classical estimators. In what follows, we compare the results of fitting the same set of estimators to real data. We thus wish to investigate whether the inference that can be drawn from widely used data sets on important economics issues changes when we use our robust estimators. The first uses data drawn from the Panel Study of Income Dynamics (PSID) with $N = 595$ and $T = 7$ and focuses on the returns to education. The second investigates crime rates in North Carolina and uses data from $N = 90$ counties over a period of $T = 7$ years.

B.1. The Cornwell-Rupert earnings equations

In their paper, Cornwell and Rupert (1988) use data from the PSID over $1976-1982$ to regress the log wage on years of education (ED), weeks of work (WKS), years of full-time experience (EXP), occupation (OCC=1 if blue-collar), region of residence (SOUTH = 1), Urban area (SMSA = 1), industry (IND = 1 if manufacturing), marital status (MS = 1 if married), sex (FEM = 1), race (BLK = 1), and a union dummy (see also Baltagi and Khanti-Akom (1990)). As per our previous notation, we let $X_1 = (OCC, SOUTH,\ SMSA, IND)$, $X_2 = (EXP, EXP2, WKS, MS, UNION)$, $Z_1 = (FEM, BLK)$ and $Z_2 = ED$. For the Mundlak model, we omit Z_1 and Z_2 and consider that only the variables in X_2 are correlated with the individual effects.

The estimation results are reported in Table B1. Notice first that since we assume that only the X_2 variables are correlated with the individual effects, the standard Within estimator does not match the Mundlak-type FE. This is especially the case for the X_1 variables, *i.e* $OCC, SOUTH$, $SMSA, IND$. On the other hand, the Mundlak-type FE, our 3S estimators and the FB estimator yield qualitatively very similar results. This is true both for the X_1 and X_2 variables. Likewise, the estimates of $\pi_{exp}, \ldots, \pi_{union}$ and of σ^2_μ and σ^2_u are nearly identical across the different estimators.

We next focus on the Hausman-Taylor model. This specification requires that Z_1 and Z_2 be included in the model. Furthermore, the assumed correlation between the individual effects and X_2 and Z_2 justifies the use of an IV method whose instruments are given by $A_{HT} = [Q_W X_1, Q_W X_2, PX_1, Z_1]$, where $Q_W = I_{NT} - P$ is the within-transformation. Recall from our discussion of the HT estimator that the nature of the correlation is left unspecified. We have proposed to proxy it through the polynomial function $f[\cdot] = (\overline{x_{2,i}} - E_{\overline{x_2}})^2 \odot (Z_{2i} - E_{Z_2})^s$, where s need to be specified empirically.[1]

Table B2 reports the results of relevant estimators as well as those obtained from fitting the standard random effects model for the sake of completeness and despite the fact it is known to be biased. Contrary to the previous table, the classical (HT) and Bayesian estimators yield qualitatively different results. According to the former, workers living in the South and employed as blue collars have the same wages as elsewhere. The Bayesian estimators, on the other hand, find a statistically significant negative difference of about 4.5% and 6% in each case, respectively. Likewise, living in a SMSA has a negative impact on the wage rates according to the HT estimator but a positive one according to our robust estimators, which seems more reasonable. While *union* has no impact according to the IV estimator, we find a positive, reasonable and significant effect when using Bayesian estimators. The estimated coefficients of other time-varying variables *ind*, *exp*, exp^2, *wks* and *ms* are similar whatever the estimation method. Yet, the more interesting differences concern the parameter estimates associated with the time invariant variables, *i.e.* gender, race and education (Z_1 and Z_2). Indeed, according to the our robust estimators, women are more discriminated against than blacks which is the exact opposite found with the HT estimator.[2] The returns to education is also four percentage points lower according to the 3S estimators. As expected from the Monte Carlo simulations, the Bayesian estimators yield considerably more precise estimates. For instance, the 3S *bootstrap* 95% confidence interval of ED is $[8.06\%; 12.9\%]$ whereas that of the HT estimator is $[9.55\%; 18.03\%]$. Recall from Table 4 that the 3S *bootstrap* estimator yielded a slightly biased parameter estimate of the endogeneous time-invariant variable. Its smaller variance more than compensated for the bias as its RMSE was also much smaller.

---

[1] As suggested, we proxy the individual effects by $\widehat{\mu} = \left(Z'_\mu Z_\mu\right)^{-1} Z'_\mu \widehat{y}$, where $\widehat{y}$ are the fitted values of the pooling regression $y = X_1\beta_1 + X_2\beta_2 + Z_1\eta_1 + Z_2\eta$. We next compute the correlation between $\widehat{\mu}$ and Z_2. Since the estimated correlation is large (0.612), we have set $s = 1$.

[2] Baltagi and Bresson (2012) propose a robust HT estimator and compare it with the standard HT estimator with the same data. They get similar returns to education but very different gender and race effects. The magnitude of the parameter estimates of gender and race change in a similar fashion to ours.

B.2. The Cornwell-Trumbull crime model

Cornwell and Trumbull (1994) estimate an economic model of crime using panel data on 90 counties in North Carolina over the period 1981–1987. The model relates crime rates to a set of variables that proxy deterrent effects and returns to legal opportunities. All the variables are expressed in logarithms and include the probability of arrest (P_A), the probability of conviction conditional on being arrested (P_C), the probability of a prison sentence given conviction (P_P), policemen per capita as a measure of the county's ability to detect crime ($Police$), the population density ($Density$), the percentage of the minority groups at the county level ($Pctmin$), and regional dummies for western and central counties ($West$, $Central$). Opportunities in the legal sector are proxied by the weekly average wage rates at the county level in two distinct industries: transportation, utilities and communication ($Wtuc$), and manufacturing ($Wmfg$).

Table B3 shows that the MCS, the robust and the FB estimators yield qualitatively similar results.[3] Note however that the Bayesian estimators match more closely the Within estimates (FE) than the MCS does, although the confidence intervals of all estimators overlap. In addition, the 3S *bootstrap* and 3S *t-dist* yield smaller standard errors than the FB does. The estimated variances of the residuals, σ_u^2, are identical across estimators but the Within estimator yields a much larger value of the variance of the individual effects, σ_μ^2.

[3] We use an iterated ($iter = 100$) MCS procedure so as to match the results of Baltagi et al. (2009). The π_t coefficients are not reported to save on space. They are available upon request.

Table B1: Earnings Equation

Within, Mundlak-type FE, Robust Three-Stage and Full Bayesian, N=595, T=7

	Within		Mundlak-type FE		3S bootstrap		3S t-dist.			Full Bayesian		
	Coef.	se	Coef.	se	Coef.	se	Coef.	se	nse	Coef.	se	nse
occ	-0.0215	0.0138	-0.0576	0.0133	-0.0796	0.0184	-0.0551	0.0135	0.0002	-0.0611	0.0151	0.0002
south	-0.0019	0.0343	-0.0596	0.0273	-0.0989	0.0489	-0.0567	0.0282	0.0004	-0.0621	0.0299	0.0004
smsa	-0.0425	0.0194	0.0178	0.0179	0.0712	0.0282	0.0134	0.0178	0.0003	0.0224	0.0204	0.0003
ind	0.0192	0.0154	0.0054	0.0147	-0.0059	0.0232	0.0062	0.0146	0.0002	0.0042	0.0161	0.0002
exp	0.1132	0.0025	0.1131	0.0025	0.1121	0.0036	0.1131	0.0025	0.0000	0.1131	0.0028	0.0000
exp^2	-0.0004	0.0001	-0.0004	0.0001	-0.0004	0.0001	-0.0004	0.0001	0.0000	-0.0004	0.0001	0.0000
wks	0.0008	0.0006	0.0009	0.0006	0.0011	0.0007	0.0009	0.0006	0.0000	0.0009	0.0007	0.0000
ms	-0.0297	0.0190	-0.0340	0.0192	-0.0418	0.0260	-0.0344	0.0191	0.0003	-0.0343	0.0211	0.0003
union	0.0328	0.0149	0.0376	0.0151	0.0404	0.0218	0.0372	0.0150	0.0002	0.0383	0.0169	0.0002
π_{exp}			-0.0509	0.0087	-0.0516	0.0098	-0.0499	0.0095	0.0001	-0.0509	0.0092	0.0001
π_{exp^2}			-0.0008	0.0002	-0.0008	0.0002	-0.0008	0.0002	0.0000	-0.0008	0.0002	0.0000
π_{wks}			0.1212	0.0019	0.1213	0.0023	0.1212	0.0020	0.0000	0.1212	0.0020	0.0000
π_{ms}			0.3537	0.0589	0.3469	0.0508	0.3439	0.0572	0.0008	0.3554	0.0620	0.0009
π_{union}			0.2019	0.0466	0.1957	0.0483	0.1916	0.0503	0.0007	0.2017	0.0493	0.0007
σ_u^2	0.0230		0.0232		0.0234		0.0231			0.0232		
σ_μ^2	1.0688		1.2759		1.2621		1.2677			1.2751		

Table B2: Earnings Equation

RE, Hausman-Taylor, Robust three-stage and Full Bayesian, N=595, T=7

	RE		Hausman-Taylor		3S bootstrap		3S t-dist.			Full Bayesian		
	Coef.	se	Coef.	se	Coef.	se	Coef.	se	nse	Coef.	se	nse
occ	-0.0501	0.0166	-0.0207	0.0138	-0.0458	0.0165	-0.0445	0.0128	0.0002	-0.0472	0.0138	0.0002
south	-0.0166	0.0265	0.0074	0.0320	-0.0616	0.0337	-0.0543	0.0201	0.0003	-0.0581	0.0218	0.0003
smsa	-0.0138	0.0200	-0.0418	0.0190	0.0580	0.0180	0.0443	0.0154	0.0002	0.0521	0.0170	0.0002
ind	0.0037	0.0173	0.0136	0.0152	0.0228	0.0193	0.0239	0.0135	0.0002	0.0239	0.0143	0.0002
exp	0.0821	0.0028	0.1131	0.0025	0.1122	0.0041	0.1133	0.0025	0.0000	0.1133	0.0027	0.0000
$\exp^2$	-0.0008	0.0001	-0.0004	0.0001	-0.0004	0.0001	-0.0004	0.0001	0.0000	-0.0004	0.0001	0.0000
wks	0.0010	0.0008	0.0008	0.0006	0.0010	0.0008	0.0009	0.0006	0.0000	0.0009	0.0007	0.0000
ms	-0.0746	0.0230	-0.0299	0.0190	-0.0404	0.0298	-0.0368	0.0188	0.0003	-0.0368	0.0207	0.0003
union	0.0632	0.0171	0.0328	0.0149	0.0454	0.0270	0.0351	0.0151	0.0002	0.0356	0.0168	0.0002
intercept	4.2637	0.0977	2.9127	0.2837	3.3046	0.1778	3.2652	0.1338	0.0019	3.2663	0.1441	0.0020
fem	-0.3392	0.0513	-0.1309	0.1267	-0.2922	0.0630	-0.2934	0.0565	0.0008	-0.2912	0.0559	0.0008
blk	-0.2103	0.0580	-0.2857	0.1557	-0.1070	0.0487	-0.1187	0.0488	0.0007	-0.1168	0.0471	0.0007
ed	0.0997	0.0057	0.1379	0.0212	0.1048	0.0121	0.1076	0.0095	0.0001	0.1073	0.0102	0.0001
σ_u^2	0.0231		0.0230		0.0233		0.0232			0.0232		
σ_μ^2	0.0690		0.8870		0.8986		0.9044			0.9059		

Table B3: Crime Model

Within, Chamberlain, Robust three-stage and Full Bayesian, N=90, T=7

	Within		Chamberlain MCS		3S bootstrap		3S t-dist.			Full Bayesian		
	Coef.	se	Coef.	se	Coef.	se	Coef.	se	nse	Coef.	se	nse
Intercept			-5.4441	0.9156	-5.6025	0.7828	-5.0497	1.7254	0.0244	-5.1360	0.8767	0.0124
pctmin			0.2042	0.0357	0.2204	0.0320	0.2166	0.0650	0.0009	0.2214	0.0704	0.0010
west			-0.1979	0.0909	-0.1379	0.0896	-0.1724	0.1503	0.0021	-0.1787	0.1768	0.0025
central			-0.0515	0.0458	-0.0786	0.0449	-0.0382	0.0807	0.0011	-0.0405	0.0911	0.0013
P_A	-0.3942	0.0328	-0.3452	0.0283	-0.3908	0.0514	-0.3942	0.0321	0.0005	-0.3955	0.0611	0.0009
P_C	-0.3108	0.0214	-0.2548	0.0172	-0.3038	0.0342	-0.3104	0.0211	0.0003	-0.3111	0.0395	0.0006
P_P	-0.2041	0.0327	-0.1763	0.0248	-0.2039	0.0490	-0.2039	0.0324	0.0005	-0.2043	0.0608	0.0009
Police	0.4203	0.0270	0.4341	0.0261	0.4027	0.0650	0.4202	0.0272	0.0004	0.4213	0.0506	0.0007
Density	0.4917	0.2743	0.3941	0.3228	0.4575	0.3702	0.4831	0.2781	0.0039	0.4809	0.4455	0.0063
wtuc	0.0259	0.0179	0.0125	0.0121	0.0213	0.0188	0.0257	0.0183	0.0003	0.0257	0.0341	0.0005
wmfg	-0.3362	0.0647	-0.2929	0.0680	-0.3174	0.0838	-0.3344	0.0650	0.0009	-0.3332	0.1124	0.0016
σ_u^2	0.0202		0.0206		0.0202		0.0202			0.0202		
σ_μ^2	0.1297		0.0505		0.0662		0.0737			0.0706		

C. The first step of the robust Bayesian estimator in the two-stage hierarchy

C.1. Derivation of eq.(8)

As $y = X\beta + Wb + u$, $u \sim N(0, \Sigma)$ 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 $v(b) = (y^* - X\widehat{\beta}(b))'(y^* - X\widehat{\beta}(b))$, where $\Lambda_X = X'X$. Then

$$\begin{aligned}
\Lambda_X \widehat{\beta}(b) = X'y^* \text{ and } \widehat{\beta}'(b) &= y^{*\prime}X\Lambda_X^{-1} \\
(y^* - X\beta)'(y^* - X\beta) &= y^{*\prime}y^* - \widehat{\beta}'(b)\Lambda_X\beta - \beta'\Lambda_X\widehat{\beta}(b) + \beta'\Lambda_X\beta \\
&= y^{*\prime}y^* - \widehat{\beta}'(b)\Lambda_X\beta - \beta'\Lambda_X\widehat{\beta}(b) + \beta'\Lambda_X\beta \\
&\quad + \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) - \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) \\
&\quad + \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) - \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) \\
(y^* - X\beta)'(y^* - X\beta) &= y^{*\prime}y^* + (\beta - \widehat{\beta}(b))'\Lambda_X\left(\beta - \widehat{\beta}(b)\right) \\
&\quad - \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) + \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b) \\
&\quad - \widehat{\beta}'(b)\Lambda_X\widehat{\beta}(b).
\end{aligned}$$

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:

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

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

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

$$m\left(y^{*}\mid\pi_0,b,g_0\right)=\int\limits_0^{\infty}\int\limits_{\mathbb{R}^{K_1}}\left\{\begin{array}{c}\left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}}g_0^{\frac{K_1}{2}}\left(\tau\right)^{\frac{NT+K_1}{2}-1}\left|\Lambda_X\right|^{1/2}\\ \times\exp\left(-\frac{\tau}{2}\left\{v\left(\beta\right)\right\}-\frac{\tau}{2}\left\{\begin{array}{c}(\beta-\widehat{\beta}\left(b\right))'\Lambda_X(\beta-\widehat{\beta}\left(b\right)\\ +g_0(\beta-\beta_0\iota_{K_1})'\Lambda_X(\beta-\beta_0\iota_{K_{11}})\end{array}\right\}\right)\end{array}\right\}d\beta\ d\tau.$$

We can simplify the expression inside the exponentiation:

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

As the Bayes estimate of β is given by[4] (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

$$\begin{aligned}m\left(y^{*}\mid\pi_0,b,g_0\right)&=\int\limits_0^{\infty}\int\limits_{\mathbb{R}^{K_1}}\left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}}g_0^{\frac{K_1}{2}}\left(\tau\right)^{\frac{NT+K_1}{2}-1}\left|\Lambda_X\right|^{1/2}\\&\quad\times\exp\left[\begin{array}{c}-\frac{\tau}{2}\left\{v\left(\beta\right)\right\}\\ -\frac{\tau}{2}\left\{\begin{array}{c}\left(g_0+1\right)\left(\beta-\beta_*(b\mid g_0)\right)'\Lambda_X(\beta-\beta_*(b\mid g_0))\\ +\left(\frac{g_0}{g_0+1}\right)\left(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1}\right)\end{array}\right\}\end{array}\right]d\beta\ d\tau\\&=\int\limits_0^{\infty}\left\{\int\limits_{\mathbb{R}^{K_1}}\exp\left[-\frac{\tau}{2}\left(g_0+1\right)\left(\beta-\beta_*(b\mid g_0)\right)'\Lambda_X(\beta-\beta_*(b\mid g_0))\right]d\beta\right\}\\&\quad\times\left(\frac{1}{2\pi}\right)^{\frac{NT+K_1}{2}}g_0^{\frac{K_1}{2}}\left(\tau\right)^{\frac{NT+K_1}{2}-1}\left|\Lambda_X\right|^{1/2}\\&\quad\times\left\{\exp\left[-\frac{\tau}{2}\left\{v\left(b\right)+\left(\frac{g_0}{g_0+1}\right)\left(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}\left(b\right)-\beta_0\iota_{K_1}\right)\right\}\right]d\tau\right\}.\end{aligned}$$

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

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

[4]Derivation of this estimator is presented below.

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

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

and using the Gauss integral formula

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

we get

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

Hence we can write

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

$$m(y \mid \pi_0, b, g_0) = (2\pi)^{-NT/2} \left(\frac{g_0}{g_0+1}\right)^{K_1/2} \int_0^{\infty} \tau^{\frac{NT}{2}-1} \exp\left(-\frac{\tau}{2} v(b) \left\{ \begin{array}{c} 1+ \\ \left(\frac{g_0}{g_0+1}\right) \left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}}\right) \end{array} \right\} \right) d\tau,$$

where

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

Using a property of the Gamma function, *i.e.*,

$$\int_0^{\infty} t^b \exp(-at) \; dt = \frac{\Gamma(b+1)}{a^{b+1}}$$

we thus get

$$\begin{aligned} m(y^\star \mid \pi_0, b, g_0) &= \widetilde{H} \left(\frac{g_0}{g_0+1}\right)^{K_1/2} \left(1 + \left(\frac{g_0}{g_0+1}\right) \frac{(\widehat{\beta}(b) - \beta_0 \iota_{K_1})' \Lambda_X (\widehat{\beta}(b) - \beta_0 \iota_{K_1})}{v(b)} \right)^{-\frac{NT}{2}} \\ &= \widetilde{H} \left(\frac{g_0}{g_0+1}\right)^{K_1/2} \left(1 + \left(\frac{g_0}{g_0+1}\right) \left(\frac{R^2_{\beta_0}}{1-R^2_{\beta_0}} \right) \right)^{-\frac{NT}{2}} \end{aligned} \tag{8}$$

with

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

Q.E.D

Following a similar line of reasoning, it trivially follows that

$$m(\widetilde{y} \mid \pi_0, \beta, h_0) \;\; = \;\; \widetilde{H} \left(\frac{h_0}{h_0+1}\right)^{K_2/2} \left(1 + \left(\frac{h_0}{h_0+1}\right) \left(\frac{R^2_{b_0}}{1-R^2_{b_0}} \right) \right)^{-\frac{NT}{2}}$$

where

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

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

C.2. Derivation of eq.(10) and eq.(11)

The maximization of $m(y^* \mid q, b, g_0)$ is equivalent to maximizing $\log m(y^* \mid q, b, g_0)$. Write:

$$\begin{aligned}\log m(y^* \mid q, b, g_0) &= \log \widetilde{H} + \frac{K_1}{2}\log\left(\frac{g_q}{g_q+1}\right) \\ &\quad -\frac{NT}{2}\log\left(1+\left(\frac{g_q}{g_q+1}\right)\frac{(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_q\iota_{K_1})}{v(b)}\right).\end{aligned}$$

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

$$\frac{\partial \log m(y^* \mid q, b, g_0)}{\partial \beta_q} = 0 \text{ and } \frac{\partial \log m(y^* \mid q, b, g_0)}{\partial g_q} = 0$$

The first term, $(\partial \log m(y^* \mid q, b)/\partial \beta_q)$, leads to

$$\begin{aligned}\frac{\partial \log m(y^* \mid q, b, g_0)}{\partial \beta_q} &= -\left(\frac{NT}{2}\right)\frac{\partial}{\partial \beta_q}\left\{\log\left(\begin{array}{c}1+ \\ \left(\frac{g_q}{g_q+1}\right)\frac{(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_q\iota_{K_1})}{v(b)}\end{array}\right)\right\} \\ &= -\left(\frac{NT}{2}\right)\cdot\left[\frac{1}{1+\left(\frac{g_q}{g_q+1}\right)\cdot\frac{(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_q\iota_{K_1})}{v(\beta)}}\right] \\ &\quad \times\left(\frac{g_q}{g_q+1}\right)(-2)(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X\iota_{K_1} = 0.\end{aligned}$$

Since

$$\left[1+\left(\frac{g_q}{g_q+1}\right)\frac{(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_q\iota_{K_1})}{v(\beta)}\right]^{-1} \neq 0 \text{ and finite}$$

it follows that

$$(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X\iota_{K_1} = 0.$$

Thus

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

The second term of the first order conditions is

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

This implies

$$\begin{aligned}\frac{\partial \log m(y^* \mid q, b, g_0)}{\partial g_q} &= \frac{\partial}{\partial g_q}\left\{\frac{K_1}{2}\log\left(\frac{g_q}{g_q+1}\right)\right\} \\ &\quad -\left(\frac{NT}{2}\right)\frac{\partial}{\partial g_q}\left\{\log\left(\begin{array}{c}1+ \\ \left(\frac{g_q}{g_q+1}\right)\frac{(\widehat{\beta}(b)-\beta_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_q\iota_{K_1})}{v(b)}\end{array}\right)\right\} \\ &= \frac{K_1}{2}\left[\frac{1}{g_q(g_q+1)}\right]-\left(\frac{NT}{2}\right)\left[\frac{\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}{(g_q+1)+g_q\left(\frac{R^2_{\beta_q}}{1-R^2_{\beta_q}}\right)}\frac{1}{g_q+1}\right] = 0,\end{aligned}$$

with

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

Therefore

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

or equivalently

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

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

Let $A = R^2_{\beta_q}/(1-R^2_{\beta_q})$. Hence

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

It follows that

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

Q.E.D

C.3. Derivation of eq.(12), eq.(13) and eq.(15)

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

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

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

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

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

$$= \left[1+\frac{\varepsilon}{1-\varepsilon}\left(\frac{\frac{\widehat{g}}{\widehat{g}+1}}{\frac{g_0}{g_0+1}}\right)^{K_1/2}\left(\frac{1+\left(\frac{g_0}{g_0+1}\right)\frac{(\widehat{\beta}(b)-\beta_0\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\beta_0\iota_{K_1})}{v(b)}}{1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})'\Lambda_X(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1})}{v(b)}}\right)^{\frac{NT}{2}}\right]^{-1}$$

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

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 :

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

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

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

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

$$\begin{aligned}\pi_0^*(\beta \mid g_0) &= \frac{\Gamma\left(\frac{NT+K_1}{2}\right)|M_{0,\beta}|^{1/2}}{\pi^{K_1/2}\Gamma\left(\frac{NT}{2}\right)}(\xi_{0,\beta})^{NT/2}\left[(\beta-\beta_*(b))'M_{0,\beta}(\beta-\beta_*(b))+\xi_{0,\beta}\right]^{-\frac{NT+K_1}{2}}, \\ \text{with } \xi_{0,\beta} &= 1+\left(\frac{g_0}{g_0+1}\right)\frac{\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\beta_0\iota_{K_1}\right)}{v(b)}. \qquad (13)\end{aligned}$$

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

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

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 $\left(\frac{\xi_{q,\beta}M_{q,\beta}^{-1}}{NT-2}\right)$ and degrees of freedom (NT) with

$$\xi_{q,\beta} = 1+\left(\frac{\widehat{g}}{\widehat{g}+1}\right)\frac{\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)'\Lambda_X\left(\widehat{\beta}(b)-\widehat{\beta}_q\iota_{K_1}\right)}{v(b)} \text{ and } M_{q,\beta} = \left(\frac{(\widehat{g}+1)}{v(\beta)}\right)\Lambda_X. \qquad (15)$$

Q.E.D

C.4. Derivation of eq.(14) and eq.(16).

To prove equation (16), 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):

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

We can write

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

Let $D = (\tau\widehat{g}\Lambda_X + \tau X'X)^{-1}$. 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

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

So

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

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

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

Since

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

Q_2^* is a constant relative to $p(\beta|y^*)$. 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

$$\begin{aligned}
\widehat{\beta}_{EB}(b \mid g_0) &= D\left(\tau X'y^* + \tau\widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right) = \left((\widehat{g}+1)\Lambda_X\right)^{-1}\left(X'y^* + \widehat{g}\widehat{\beta}_q\Lambda_X\iota_{K_1}\right) \\
&= \left((\widehat{g}+1)\right)^{-1}\left(\Lambda_X^{-1}X'y^* + \widehat{g}\widehat{\beta}_q\iota_{K_1}\right) \\
&= \frac{\widehat{\beta}(b) + \widehat{g}\widehat{\beta}_q\iota_{K_1}}{\widehat{g}+1} = \widehat{\beta}(b) - \frac{\widehat{g}}{\widehat{g}+1}\left(\widehat{\beta}(b) - \widehat{\beta}_q\iota_{K_1}\right). \qquad (16)
\end{aligned}$$

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 $\pi_0(\beta, \tau \mid g_0)$:

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

Q.E.D

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

D.1. Derivation of eq.(20) and eq.(21)

In the second step of the two-stage hierarchy, we have derived the predictive density corresponding to the base prior conditional on h_0 (See section C.1):

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

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

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

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

$$\begin{aligned}
m(\widetilde{y} \mid \pi_0, \beta) &= \frac{\widetilde{H}}{B(c,d)}\int_0^1 (\varphi)^{\frac{K_2}{2}+c+1}(1-\varphi)^{d-1}\left(1 + \varphi\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right)^{-\frac{NT}{2}} d\varphi \qquad (20) \\
&= \frac{B(d, \frac{K_2}{2}+c)}{B(c,d)}\widetilde{H} \times {}_2F_1\left(\frac{NT}{2}; \frac{K_2}{2}+c; \frac{K_2}{2}+c+d; -\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right),
\end{aligned}$$

where ${}_2F_1$ is the Gaussian hypergeometric function.[5] 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:

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

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

$$\begin{aligned} \int_0^1 (t)^{a_2-1}\left(1-t\right)^{a_3-a_2-1}\left(1-zt\right)^{-a_1} dt &= B(a_2, a_3-a_2) \times {}_2F_1\left(a_1;a_2;a_3;z\right) \\ &= \frac{\Gamma\left(a_2\right)\Gamma\left(a_3-a_2\right)}{\Gamma\left(a_3\right)} \times {}_2F_1\left(a_1;a_2;a_3;z\right) \end{aligned}$$

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

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

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.[6] The second incomplete Gaussian hypergeometric function can be written as:

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

[6]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:

$$m\left(\widetilde{y}\mid\widehat{q},\beta\right)= \tag{21}$$
$$=\frac{\widetilde{H}}{B\left(c,d\right)}\left\{\begin{array}{c}\frac{2.\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}+c}}{K_2+2c}\\ \times F_1\left(\frac{K_2}{2}+c;1-d;\frac{NT}{2};\frac{K_2}{2}+c+1;\frac{h^*}{h^*+1};-\frac{h^*}{h^*+1}\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)\\ +\left\{\begin{array}{c}\left[\left(\frac{h^*}{h^*+1}\right)^{\frac{K_2}{2}}\left(1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R^2_{b_q}}{1-R^2_{b_q}}\right)\right)^{-\frac{NT}{2}}\right]\\ \times\left[\begin{array}{c}B\left(c,d\right)-\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c}\\ \times{}_2F_1\left(c;d-1;c+1;\frac{h^*}{h^*+1}\right)\end{array}\right]\end{array}\right\}\end{array}\right\}$$

D.2. Derivation of eq.(23)

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

$$\begin{aligned}\int_0^{\frac{h^*}{h^*+1}}\left(\varphi\right)^{c-1}\left(1-\varphi\right)^d d\eta&=\int_0^1\left((\frac{h^*}{h^*+1})t\right)^{c-1}\left(1-(\frac{h^*}{h^*+1})t\right)^d(\frac{h^*}{h^*+1})dt\\ &=\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c}\times_2F_1\left(c;-d;c+1;\frac{h^*}{h^*+1}\right).\end{aligned}$$

The solution of the second one is:

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

and the solution of the third one is:

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

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

$$\widehat{b}_{EB}(\beta) = \frac{1}{B(c,d)} \begin{bmatrix} \widehat{b}(\beta) \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \times_2 F_1 \left(c; -d; c+1; \frac{h^*}{h^*+1} \right) \\ +\widehat{b}_q \iota_{K_2} \frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1} \times_2 F_1 \left(c+1; 1-d; c+2; \frac{h^*}{h^*+1} \right) \\ + \left\{ \widehat{b}(\beta) \left(\frac{1}{h^*+1} \right) + \widehat{b}_q \iota_{K_2} \left(\frac{h^*}{h^*+1} \right) \right\} \\ \times \begin{bmatrix} B(c,d) - \frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2 F_1 \left(c; d-1; c+1; \frac{h^*}{h^*+1} \right) \end{bmatrix} \end{bmatrix}. \tag{23}$$

E. Variance-covariance matrices of the 2S and 3S robust Bayesian estimators

Following Berger (1985)(p.207), the ML-II posterior variance-covariance matrix of β of the first step of the 2S robust Bayesian estimator is given by

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

The ML-II posterior variance-covariance matrix of b of the second step of the 2S robust Bayesian estimator can be derived in a similar fashion to that of $\hat{\beta}_{ML-II}$.

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

The ML-II posterior variance-covariance matrix of b of the 3S robust Bayesian estimator can be derived as in the two-stage hierarchy model:

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

The main differences with the latter 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 $\widehat{\lambda}_{b,h_0}$).

F. Positive or negative semidefinite differences of MSE

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

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

with

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

The ML-II posterior variance-covariance matrix of β is given by (see above):

$$\begin{aligned} Var\left(\widehat{\beta}_{ML-II}\right) &= \lambda_{\beta,g_0} Var\left[\pi_0^*(\beta \mid g_0)\right] + \left(1-\lambda_{\beta,g_0}\right) Var\left[q^*(\beta \mid g_0)\right] \\ &+ \lambda_{\beta,g_0}\left(1-\lambda_{\beta,g_0}\right)\left(\beta_*(b \mid g_0)-\beta_{EB}(b \mid g_0)\right)\left(\beta_*(b \mid g_0)-\beta_{EB}(b \mid g_0)\right)' \end{aligned}$$

Under the null, $H_0 : \varepsilon = 0$, the weight is $\lambda_{\beta,g_0} = 1$ and the restricted estimate is

$$\widehat{\beta}_{rest} = \beta_*(b \mid g_0)$$

while under $H_1 : \varepsilon \neq 0$, the unrestricted estimate is $\widehat{\beta}_{un} = \lambda_{\beta,g_0}\beta_*(b \mid g_0) + \left(1-\widehat{\lambda}_{\beta,g_0}\right)\beta_{EB}(b \mid g_0)$. Following Magnus and Durbin (1999), we can write the joint distribution of β_{rest} and β_{un} as:

$$\begin{pmatrix} \widehat{\beta}_{rest} \\ \widehat{\beta}_{un} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_{un}+Q \\ \beta_{un} \end{pmatrix}, \begin{pmatrix} Var\left[\pi_0^*(\beta \mid g_0)\right] & Var\left[\pi_0^*(\beta \mid g_0)\right] \\ Var\left[\pi_0^*(\beta \mid g_0)\right] & Var\left[\pi_0^*(\beta \mid g_0)\right]+D \end{pmatrix}\right)$$

where

$$Q = \left(1-\lambda_{\beta,g_0}\right)\left(\beta_*(b \mid g_0)-\beta_{EB}(b \mid g_0)\right)$$

and

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

Applying Theorem 1 of Magnus and Durbin (1999) on this joint distribution, we get the conditions under which $MSE\left(\widehat{\beta}_{rest}\right) - MSE\left(\widehat{\beta}_{un}\right)$ are positive or negative semidefinite:

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

and hence

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

and

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

D^{-} denotes the generalized inverse of D. If $MSE\left(\widehat{\beta}_{rest}\right)$ and $MSE\left(\widehat{\beta}_{un}\right)$ are two positive semidefinite matrices, the notation $MSE\left(\widehat{\beta}_{rest}\right) \leq MSE\left(\widehat{\beta}_{un}\right)$ means that $MSE\left(\widehat{\beta}_{un}\right) - MSE\left(\widehat{\beta}_{rest}\right)$ is positive semidefinite.

G. Full Bayesian estimators

We also derive the full Bayesian estimators for RE world, the Mundlak world, the Chamberlain world and the Hausman-Taylor world.[7]

G.1. Gibbs sampling for the RE world

We run full Bayesian estimates (Gibbs sampling) on the RE world following the works of Chib and Carlin (1999), Koop (2003), Chib (2008), and Greenberg (2008) to mention a few. They have proposed algorithms for the three-stage hierarchical models in a standard RE world. Our initial specification $y_i = X_i\beta + W_i b_i + u_i$ could be written as the following 3S hierarchy:

$$\begin{cases} \text{First stage}: & y_i = X_i\beta + \iota_T\mu_i + u_i, u_i \sim N(0, \tau^{-1}I_T) \\ \text{Second stage}: & \beta \sim N(\beta_0, B_0) \text{ and } \mu_i \sim N\left(0, \sigma_\mu^2\right) \\ \text{Third stage}: & \tau \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \text{and } \sigma_\mu^{-2} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right). \end{cases}$$

where y_i is $(T \times 1)$, X_i is $(T \times K_1)$, u_i is $(T \times 1)$. The $(T \times K_2)$ matrix W_i of our original model has been replaced by the $(T \times 1)$ vector of ones, ι_T, and the $(K_2 \times 1)$ vector b_i is replaced by the scalar μ_i which expresses the time-invariant individual effect. $G(.)$ is the Gamma distribution.

We can define the conditional posterior distributions within the Gibbs sampler of the previous model for the RE world[8].

1. We choose diffuse priors with the following hyperparameters $\beta_0 = 0_{K_1}$, $B_0 = 10^2 I_{K_1}$, $\alpha_0 = 2$, $\delta_0 = 200$, $\gamma_0 = 2$, $\eta_0 = 200$ such that the means of the precision τ and σ_μ^{-2} are $E[\tau] = E\left[\sigma_\mu^{-2}\right] = 10^{-2}$ and their variances[9] are $Var[\tau] = Var\left[\sigma_\mu^{-2}\right] = 10^{-4}$.
2. We draw initial values of:

$$\begin{array}{ll} \beta^{(0)} \sim N(\beta_0, B_0) & , \tau^{(0)} \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \\ \sigma_\mu^{-2(0)} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right) & , \mu_i^{(0)} \sim N\left(0, \sigma_\mu^{2\,(0)}\right) \end{array}$$

3. At the d^{th} (for $d = 1, ..., D$) draw, we sample:

$$\begin{aligned} \tau^{(d)} &\sim G\left(\frac{\alpha_1}{2}, \frac{{\delta_1}^{(d)}}{2}\right) \\ \sigma_\mu^{-2(d)} &\sim G\left(\frac{\gamma_1}{2}, \frac{{\eta_1}^{(d)}}{2}\right) \\ {\mu_i}^{(d)} &\sim N\left({\bar{\mu_i}}^{(d)}, {D_{1i}}^{(d)}\right) \\ \beta^{(d)} &\sim N\left(\bar{\beta}^{(d)}, {B_1}^{(d)}\right) \end{aligned}$$

[7] We thank two referees for this suggestion.
[8] See below for the derivations.
[9] A random variable x follows a Gamma distribution $G(\alpha, \beta)$ with shape α and scale β if its pdf can be written as

$$p(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

Its mean and variance are given by $E[x] = \frac{\alpha}{\beta}$ and $Var[x] = \frac{\alpha}{\beta^2}$.

where

$$
\begin{aligned}
\delta_1{}^{(d)} &= \delta_0 + \sum_{i=1}^{N} \left(y_i - X_i \beta^{(d-1)} - \iota_T \mu_i{}^{(d-1)} \right)' \times \left(y_i - X_i \beta^{(d-1)} - \iota_T \mu_i{}^{(d-1)} \right) \\
\eta_1{}^{(d)} &= \eta_0 + \sum_{i=1}^{N} \mu_i^{2(d-1)} \\
D_{1i}{}^{(d)} &= \left[T\tau^{(d)} + \ \sigma_\mu^{-2(d)} \right]^{-1} \\
\bar{\mu_i}^{(d)} &= D_{1i}{}^{(d)} \left[\tau^{(d)} \iota_T' \tilde{y}_i^{(d-1)} \right] \\
\tilde{y}_i^{(d-1)} &= y_i - X_i \beta^{(d-1)} \\
B_{1i}{}^{(d)} &= \sigma_\nu^{2(d)} J_T + \tau^{-1(d)} I_T \\
B_1{}^{(d)} &= \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1(d)} X_i \right) + B_0^{-1} \right]^{-1} \\
\bar{\beta}^{(d)} &= B_1{}^{(d)} \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1(d)} y_i \right) + B_0^{-1} \beta_0 \right] \\
\alpha_1 &= \alpha_0 + NT \\
\gamma_1 &= \gamma_0 + N
\end{aligned}
$$

For the Gibbs sampling, we run $D = 1,000$ draws and we burn the $D_{burn} = 500$ first draws. We store all the vectors β and $\{\mu_i\}_{i=1,\ldots,N}$ and the scalars σ_ϵ^2 and σ_μ^2 for the $D^*(= D - D_{burn})$ draws. When the D draws are completed, we compute their posterior means and standard errors on the D^* last draws.

G.2. Gibbs sampling for the Mundlak world, the Chamberlain world and the Hausman-Taylor world

The assumption $E[\mu_i] = 0$ assumes exchangeability and independence of the μ_i and the covariates in X_i. This assumption is unrealistic in the Mundlak world, the Chamberlain world or in the Hausman-Taylor world. Dependence between μ_i and covariates A_i (possibly including covariates in X_i) may be introduced in a hierarchical fashion (see Greenberg (2008)) by assuming:

$$\mu_i \sim N\left(A_i\gamma, \sigma_\nu^2\right) \text{ with } \gamma \sim N\left(\gamma_0, G_0\right).$$

A_i is a $(1 \times r)$ vector of covariates and γ a $(r \times 1)$ vector of parameters. γ_0 and G_0 are known hyperparameters. Then, $\mu_i = A_i\gamma + \nu_i$ with $\nu_i \sim N(0, \sigma_\nu^2)$. Moreover, the nature of A_i changes, depending on the specified world in our Monte-Carlo simulation study:

- for the Mundlak world, $A_i = \bar{x}_{2i}$ and $\gamma = \pi$.
- for the Chamberlain world, $A_i = (x_{2i1}, x_{2i2}, ..., x_{2iT})$ and $\gamma = (\pi_1, \pi_2, ..., \pi_T)'$.
- for the Hausman-Taylor world, $A_i = f\left(\bar{x}_{2i}, Z_{2i}\right)$.

This approach — with new hyperparameters γ_0 and G_0 — could be rather complicated especially for the Hausman-Taylor world. But we can use the 3S hierarchy with few changes in the set explanatory variables, the coefficients vector and in the specific effects. As $\mu_i = A_i\gamma + \nu_i$ with $\nu_i \sim N(0, \sigma_\nu^2)$, we can write $y_i = X_i\beta + \iota_T\mu_i + u_i = (X_i\beta + \iota_T A_i\gamma) + \iota_T\nu_i + u_i$ and our initial 3S hierarchy becomes

$$
\left\{
\begin{array}{cc}
\text{First stage :} & y_i = X_i^*\beta^* + \iota_T\nu_i + u_i, u_i \sim N(0, \tau^{-1} I_T) \\
\text{Second stage :} & \beta^* \sim N\left(\beta_0, B_0\right) \text{ and } \nu_i \sim N\left(0, \sigma_\nu^2\right) \\
\text{Third stage :} & \tau \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \text{and } \sigma_\nu^{-2} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right).
\end{array}
\right.
$$

where $X_i^* = [X_i, A_i]$, $\beta^{*\prime} = (\beta', \gamma')'$ and $\sigma_\mu^2 = Var[A_i\gamma] + \sigma_\nu^2$. $E[\nu_i] = 0$ assumes exchangeability and independence of the ν_i and the covariates in X_i^*. For the Hausman-Taylor world, $y_i = X_i\beta + Z_i\eta + \iota_T\mu_i + u_i$, we use our proposed strategy: $\mu_i = \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right)\theta_X + \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right)^2 \odot \left(Z_{2i} - E_{Z_2}\right)\theta_Z + \nu_i$.
Then, $X_i^* = [X_i, Z_i, A_i] = \left[X_i, Z_i, \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right), \left(\overline{x_{2,i}} - E_{\overline{x_2}}\right)^2 \odot \left(Z_{2i} - E_{Z_2}\right)\right]$ and $\beta^{*\prime} = (\beta', \eta', \gamma')' = (\beta', \eta', \theta_X', \theta_Z')'$.

G.3. Derivation of the posterior densities of the Gibbs sampling for the RE world

Our initial specification $y_i = X_i\beta + W_i b_i + u_i$ is written as the following 3S hierarchy:

$$\begin{cases} \text{First stage :} & y_i = X_i\beta + \iota_T\mu_i + u_i, u_i \sim N(0, \Sigma_u) \text{ with } \Sigma_u = \tau^{-1} I_T \\ \text{Second stage :} & \beta \sim N(\beta_0, B_0) \text{ and } \mu_i \sim N(0, \Sigma_\mu) \text{ with } \Sigma_\mu = \sigma_\mu^2 \\ \text{Third stage :} & \tau \sim G\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right) \text{ and } \sigma_\mu^{-2} \sim G\left(\frac{\gamma_0}{2}, \frac{\eta_0}{2}\right). \end{cases}$$

where y_i is $(T \times 1)$, X_i is $(T \times K_1)$, u_i is $(T \times 1)$. The $(T \times K_2)$ matrix W_i has been replaced by the $(T \times 1)$ vector of ones ι_T and the $(K_2 \times 1)$ vector b_i is replaced by the scalar μ_i which expresses the time-invariant individual effect and $\mu = (\mu_1, \mu_2, ..., \mu_N)'$ is the $(N \times 1)$ vector of time-invariant effects. $G(.)$ is the Gamma distribution.

The known hyperparameters are: β_0, B_0, α_0, δ_0, γ_0 and η_0. The posterior distribution is proportional to:

$$\begin{aligned} \pi\left(\beta, \mu, \tau, \sigma_\mu^2 \mid y, X\right) \quad \propto \quad & |\Sigma_u|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\beta - \iota_T\mu_i)' \Sigma_u^{-1}(y_i - X_i\beta - \iota_T\mu_i)\right] \\ & \times \exp\left[-\frac{1}{2}(\beta - \beta_0)' B_0^{-1}(\beta - \beta_0)\right] \\ & \times \tau^{\frac{\alpha_0}{2}-1} \exp\left[-\frac{\tau\delta_0}{2}\right] \\ & \times \sigma_\mu^{\frac{-2N}{2}} \exp\left[-\frac{\sigma_\mu^{-2}\mu'\mu}{2}\right] \\ & \times \sigma_\mu^{-2\left[\frac{\gamma_0}{2}-1\right]} \exp\left[-\frac{\sigma_\mu^{-2}\eta_0}{2}\right] \end{aligned}$$

The posterior distribution of the precision τ is given by:

$$\begin{aligned} & |\Sigma_u|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(y_i - X_i\beta - \iota_T\mu_i)' \Sigma_u^{-1}(y_i - X_i\beta - \iota_T\mu_i)\right] \\ & \times \tau^{\frac{\alpha_0}{2}-1} \exp\left[-\frac{\tau\delta_0}{2}\right] \\ = \quad & \tau^{\frac{\alpha_0+NT}{2}-1} \exp\left[-\frac{\tau}{2}\left\{\delta_0 + \sum_{i=1}^{N}(y_i - X_i\beta - \iota_T\mu_i)'(y_i - X_i\beta - \iota_T\mu_i)\right\}\right] \end{aligned}$$

then

$$\begin{aligned} \tau \quad \sim \quad & G\left(\frac{\alpha_1}{2}, \frac{\delta_1}{2}\right) \\ & \text{with } \alpha_1 = \alpha_0 + NT \\ & \text{and } \delta_1 = \delta_0 + \sum_{i=1}^{N}(y_i - X_i\beta - \iota_T\mu_i)'(y_i - X_i\beta - \iota_T\mu_i) \end{aligned}$$

The posterior distribution of the precision σ_μ^{-2} is given by:

$$\begin{aligned} & \sigma_\mu^{\frac{-2N}{2}} \exp\left[-\frac{\sigma_\mu^{-2}\mu'\mu}{2}\right] \times \sigma_\mu^{-2\left[\frac{\gamma_0}{2}-1\right]} \exp\left[-\frac{\sigma_\mu^{-2}\eta_0}{2}\right] \\ = \quad & \sigma_\mu^{-2\left[\frac{N+\gamma_0}{2}-1\right]} \exp\left[-\frac{\sigma_\mu^{-2}}{2}(\eta_0 + \mu'\mu)\right] \end{aligned}$$

then

$$\begin{aligned} \sigma_\mu^{-2} \quad \sim \quad & G\left(\frac{\gamma_1}{2}, \frac{\eta_1}{2}\right) \\ & \text{with } \gamma_1 = \gamma_0 + N \\ & \text{and } \eta_1 = \eta_0 + \mu'\mu \end{aligned}$$

Following Chib and Carlin (1999) and Greenberg (2008), it is preferable to sample β and μ in one block as $\pi\left(\beta,\mu \mid y,X,\tau,\sigma_\mu^2\right)$ rather than in two blocks $\pi\left(\beta \mid y,X,\mu,\tau,\sigma_\mu^2\right)$ and $\pi\left(\mu \mid y,X,\beta,\tau,\sigma_\mu^2\right)$, because of potential correlation between the two. This is done by using:

$$\begin{aligned}
\pi\left(\beta,\mu \mid y,X,\tau,\sigma_\mu^2\right) &= \pi\left(\beta \mid y,X,\mu,\tau,\sigma_\mu^2\right) \times \pi\left(\mu \mid y,X,\beta,\tau,\sigma_\mu^2\right) \\
&= \pi\left(\beta \mid y,X,\mu,\tau,\sigma_\mu^2\right) \times \prod_{i=1}^{N} \pi\left(\mu_i \mid y_i,X_i,\beta,\tau,\sigma_\mu^2\right)
\end{aligned}$$

The first terms on the right-hand side is obtained by integrating out the μ_i from $\pi\left(\beta,\mu \mid y,X,\tau,\sigma_\mu^2\right)$. For the second term, set $\tilde{y}_i = y_i - X_i\beta$ and complete the square in μ_i.

$$\begin{aligned}
\pi\left(\mu_i \mid y_i,X_i,\beta,\tau,\sigma_\mu^2\right) &\propto \exp\left[-\frac{1}{2}\left(\tilde{y}_i - \iota_T\mu_i\right)' \Sigma_u^{-1}\left(\tilde{y}_i - \iota_T\mu_i\right)\right] \\
&\quad \times \exp\left[-\frac{1}{2}\left(\mu_i\right)' \Sigma_\mu^{-1}\left(\mu_i\right)\right]
\end{aligned}$$

Let us consider the expressions in the exponentiations, ignoring the $\left(-\frac{1}{2}\right)$ terms:

$$\begin{aligned}
&\left(\tilde{y}_i - \iota_T\mu_i\right)' \Sigma_u^{-1}\left(\tilde{y}_i - \iota_T\mu_i\right) + \left(\mu_i\right)' \Sigma_\mu^{-1}\left(\mu_i\right) \\
= \;& \tilde{y}_i'\Sigma_u^{-1}\tilde{y}_i - \tilde{y}_i'\Sigma_u^{-1}\iota_T\mu_i - \mu_i'\iota_T'\Sigma_u^{-1}\tilde{y}_i + \mu_i'\iota_T'\Sigma_u^{-1}\iota_T\mu_i + \mu_i'\Sigma_\mu^{-1}\mu_i \\
= \;& \mu_i'\left[\iota_T'\Sigma_u^{-1}\iota_T + \Sigma_\mu^{-1}\right]\mu_i - 2\mu_i'\left[\iota_T'\Sigma_u^{-1}\tilde{y}_i\right] + \tilde{y}_i'\Sigma_u^{-1}\tilde{y}_i \\
= \;& \mu_i\left[\tau\iota_T'\iota_T + \sigma_\mu^{-2}\right]\mu_i - 2\mu_i\left[\tau\iota_T'\tilde{y}_i\right] + \tau\tilde{y}_i'\tilde{y}_i \\
= \;& \mu_i\left[T\tau + \sigma_\mu^{-2}\right]\mu_i - 2\mu_i\left[\tau\iota_T'\tilde{y}_i\right] + \tau\tilde{y}_i'\tilde{y}_i
\end{aligned}$$

Since we are only concerned with the distribution of μ_i, and as $\Sigma_u (= \tau^{-1}I_T)$ is assumed to be known, terms that do not involve μ_i are all absorbed into the proportionality constant. Applying this idea to the expressions between brackets, then, the posterior distribution of the time-invariant specific effect μ_i is given by:

$$\begin{aligned}
\mu_i &\sim N\left(\overline{\mu}_i, D_{1i}\right) \\
&\quad \text{with } D_{1i} = \left[\iota_T'\Sigma_u^{-1}\iota_T + \Sigma_\mu^{-1}\right]^{-1} \\
&\quad \text{and } \overline{\mu}_i = D_{1i}\left[\iota_T'\Sigma_u^{-1}\tilde{y}_i\right]
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
\mu_i &\sim N\left(\overline{\mu}_i, D_{1i}\right) \\
&\quad \text{with } D_{1i} = \left[T\tau + \sigma_\mu^{-2}\right]^{-1} \\
&\quad \text{and } \overline{\mu}_i = D_{1i}\left[\tau\iota_T'\tilde{y}_i\right]
\end{aligned}$$

To find the posterior distribution of β, we write:

$$y_i = X_i\beta + \left(\iota_T\mu_i + u_i\right) = X_i\beta + \psi_i$$

and integrate out μ_i and u_i. Then

$$\begin{aligned}
E\left[\psi_i\psi_i'\right] &= E\left[\left(\iota_T\mu_i + u_i\right)\left(\iota_T\mu_i + u_i\right)_i'\right] \\
&= \iota_T E\left[\mu_i\mu_i'\right]\iota_T' + \left[u_i u_i'\right]
\end{aligned}$$

As $u_i \sim N(0,\Sigma_u)$ with $\Sigma_u = \tau^{-1}I_T$ and $\mu_i \sim N\left(0,\Sigma_\mu\right)$ with $\Sigma_\mu = \sigma_\mu^2$, then

$$E\left[\psi_i\psi_i'\right] = \iota_T\Sigma_\mu\iota_T' + \Sigma_u = \sigma_\mu^2 J_T + \tau^{-1}I_T = B_{1i}$$

which implies $y_i \sim N\left(X_i\beta, B_{1i}\right)$. It follows that

$$\begin{aligned}
\pi\left(\beta \mid y,X,\mu,\tau,\sigma_\mu^2\right) &\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(y_i - X_i\beta\right)' B_{1i}^{-1}\left(y_i - X_i\beta\right)\right] \\
&\quad \times \exp\left[-\frac{1}{2}\left(\beta - \beta_0\right)' B_0^{-1}\left(\beta - \beta_0\right)\right]
\end{aligned}$$

Completing the expressions between brackets, we get:

$$\begin{aligned}
& \sum_{i=1}^{N} (y_i - X_i\beta)' B_{1i}^{-1} (y_i - X_i\beta) + (\beta - \beta_0)' B_0^{-1} (\beta - \beta_0) \\
= \; & \beta' \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1} X_i \right) + B_0^{-1} \right] \beta - 2\beta' \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1} y_i \right) + B_0^{-1} \beta_0 \right] \\
& + \sum_{i=1}^{N} y_i' B_{1i}^{-1} y_i + \beta_0' B_0^{-1} \beta_0
\end{aligned}$$

from which we have

$$\begin{aligned}
\beta \;\; \sim \;\; & N\left(\overline{\beta}, B_1\right) \\
& \text{with } B_1 = \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1} X_i \right) + B_0^{-1} \right]^{-1} \\
& \text{with } B_{1i} = \sigma_\mu^2 J_T + \tau^{-1} I_T, \, \forall i \\
& \text{and } \overline{\beta} = B_1 \left[\sum_{i=1}^{N} \left(X_i' B_{1i}^{-1} y_i \right) + B_0^{-1} \beta_0 \right]
\end{aligned}$$

H. Laplace approximations

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

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

$$\begin{aligned}
m\left(\widetilde{y} \mid \pi_0, \beta\right) &= \int_0^\infty m\left(\widetilde{y} \mid \pi_0, \beta, h_0\right) p(h_0) dh_0 \\
&= \frac{\widetilde{H}}{B(c,d)} \times \int_0^\infty \left\{ \begin{array}{c} \left(\frac{h_0}{h_0+1}\right)^{K_2/2} \left(1 + \left(\frac{h_0}{h_0+1}\right)\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right)\right)^{-\frac{NT}{2}} \\ \times \, h_0^{c-1} \left(\frac{1}{1+h_0}\right)^{c+d} \end{array} \right\} dh_0 \\
&= \frac{B(d, \frac{K_2}{2}+c)}{B(c,d)} \widetilde{H} \times {}_2F_1 \left(\frac{NT}{2}; \frac{K_2}{2}+c; \frac{K_2}{2}+c+d; -\left(\frac{R_{b_0}^2}{1-R_{b_0}^2}\right) \right).
\end{aligned}$$

As shown by Liang et al. (2008), numerical overflow is problematic for moderate to large NT and large $R_{b_0}^2$ 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:

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

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

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

The Laplace approximation is given by:

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

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

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

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

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

The roots are given by:

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

with

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

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

$$\exp\left[\widehat{\phi}\right] = \frac{C_1 + \sqrt{\Delta}}{C_2}. \tag{H.1}$$

The corresponding variance is

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

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

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

H.2. Laplace approximation of the predictive density based on the contaminated prior

Recall that

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

with

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

Consequently, we can write

$$\begin{aligned} m\left(\widetilde{y}\mid\widehat{q},\beta\right) &= \int\limits_0^\infty m\left(\widetilde{y}\mid\widehat{q},\beta,h_0\right)p(h_0)dh_0 \\ &= \frac{\widetilde{H}}{B\left(c,d\right)}\int\limits_0^{h^*}\left\{\begin{array}{c}\left(\frac{h_0}{h_0+1}\right)^{K_2/2}\left[1+\left(\frac{h_0}{h_0+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}}\\ \times\, h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d}\end{array}\right\}dh_0 \\ &\quad+\frac{\widetilde{H}}{B\left(c,d\right)}\left(\frac{h^*}{h^*+1}\right)^{K_2/2}\left[1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}} \\ &\quad\times\int\limits_{h^*}^\infty h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d}dh_0 \\ &= \frac{\widetilde{H}}{B\left(c,d\right)}\left[I_1+\left(\frac{h^*}{h^*+1}\right)^{K_2/2}\left[1+\left(\frac{h^*}{h^*+1}\right)\left(\frac{R_{b_q}^2}{1-R_{b_q}^2}\right)\right]^{-\frac{NT}{2}}I_2\right]. \end{aligned}$$

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

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

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

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

As

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

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

$$l_2\left(\phi\right)=c\phi+\left(d-1\right)\log\left(1-\exp\left[\phi\right]\right).$$

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

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

and the root is given by:

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

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

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

and the Laplace approximation of I_2 is

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

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

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

H.3. 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:

$$\begin{aligned}
\widehat{b}_{EB}(\beta) &= \int_0^{\infty}\widehat{b}_{EB}(\beta \mid h_0)\, p(h_0) dh_0 \\
&= \frac{1}{B(c,d)}\left[\begin{array}{c}\widehat{b}(\beta)\int_0^{h^*} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_q \iota_{K_2}\int_0^{h^*} h_0^{c}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ +\left\{\widehat{b}(\beta)\left(\frac{1}{g^*+1}\right)+\widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\}\int_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0\end{array}\right] \\
&= \frac{1}{B(c,d)}\left[\begin{array}{c}\widehat{b}(\beta)\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c}\times_2 F_1\left(c;-d;c+1;\frac{h^*}{h^*+1}\right) \\ +\widehat{b}_q\iota_{K_2}\frac{\left(\frac{h^*}{h^*+1}\right)^{c+1}}{c+1}\times_2 F_1\left(c+1;1-d;c+2;\frac{h^*}{h^*+1}\right) \\ +\left\{\widehat{b}(\beta)\left(\frac{1}{h^*+1}\right)+\widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\} \\ \times\left[\begin{array}{c}B(c,d)-\frac{\left(\frac{h^*}{h^*+1}\right)^c}{c} \\ \times_2F_1\left(c;d-1;c+1;\frac{h^*}{h^*+1}\right)\end{array}\right]\end{array}\right].
\end{aligned}$$

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

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

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

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

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:

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

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:

$$\begin{aligned}
\widehat{b}_{EB}(\beta) &= \frac{1}{B(c,d)}\left[\begin{array}{c} \widehat{b}(\beta)\int\limits_0^{h^*} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 + \widehat{b}_q\iota_{K_2}\int\limits_0^{h^*} h_0^{c}\left(\frac{1}{1+h_0}\right)^{c+d+1} dh_0 \\ +\left\{\widehat{b}(\beta)\left(\frac{1}{g^*+1}\right)+\widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\}\int\limits_{h^*}^{\infty} h_0^{c-1}\left(\frac{1}{1+h_0}\right)^{c+d} dh_0 \end{array}\right] \\
&\simeq \frac{\widehat{\beta}(b)}{B(c,d)}.\sqrt{2\pi}\widehat{\sigma}^2_{l_{D_1}} \exp\left[l_{D_1}\left(\widehat{\phi}\right)\right] + \frac{\widehat{b}_q\iota_{K_2}}{B(c,d)}\sqrt{2\pi}.\widehat{\sigma}^2_{l_{D_2}} \exp\left[l_{D_2}\left(\widehat{\phi}\right)\right] \\
&+ \frac{\left\{\widehat{b}(\beta)\left(\frac{1}{g^*+1}\right)+\widehat{b}_q\iota_{K_2}\left(\frac{h^*}{h^*+1}\right)\right\}}{B(c,d)}\sqrt{2\pi}\widehat{\sigma}_{l_2} \exp\left[l_2\left(\widehat{\phi}\right)\right].
\end{aligned}$$

I. References

Abramovitz, M., Stegun, I.A., 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 3, 173–216.

Baltagi, B., Khanti-Akom, S., 1990. On efficient estimation with panel data: an empirical comparison of instrumental variables estimators. Journal of Applied Econometrics 5, 401–406.

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

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

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

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

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

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

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

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

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

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

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

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

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

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