Central European Journal of Economic Modelling and Econometrics

Bayesian Variations on the Frisch and Waugh Theme

Jacek Osiewalski*

Submitted: 19.11.2011, Accepted: 9.12.2011

Abstract

The paper is devoted to discussing consequences of the so-called Frisch-Waugh Theorem to posterior inference and Bayesian model comparison. We adopt a generalised normal linear regression framework and weaken its assumptions in order to cover non-normal, jointly elliptical sampling distributions, autoregressive specifications, additional nuisance parameters and multi-equation SURE or VAR models. The main result is that inference based on the original full Bayesian model can be obtained using transformed data and reduced parameter spaces, provided the prior density for scale or precision parameters is appropriately modified.

Keywords: Bayesian inference, regression models, SURE models, VAR processes, data transformations

JEL Classification: C11, C51, C52.

*Cracow University of Economics, e-mail: eeosiewa@cyf-kr.edu.pl

1 Introduction

In their 1933 *Econometrica* paper Ragnar Frisch and Frederic V. Waugh showed that in a time-series context the same OLS estimates were obtained whether a regression with a time trend was fitted to original variables or the "de-trended" data were used in a regression without trend variable. This simple result is a consequence of purely algebraic properties of the OLS estimator. 70 years later it was formulated in a more general way; Greene (2003, p.27) stated the "Frisch-Waugh Theorem" in the following form (slightly modified here):

In the linear least squares regression of vector y on two sets of variables, X_1 and X_2, the sub-vector b_1 [of the OLS estimator b] *is the set of coefficients obtained when the residuals from a regression of y on X_2 alone are regressed on the set of residuals obtained when each column of X_1 is regressed on X_2.*

Intuitively, and quite obviously, this result is applicable in the Bayesian context when posterior calculations coincide with the OLS ones. In particular, in the normal linear regression framework with the Jeffreys prior density we can obtain the marginal posterior distribution of the sub-vector β_1 (of the regression parameter vector β) using only the OLS residuals defined above. The same holds for linear autoregressive models under an improper uniform prior on the coefficients. This justifies building the Bayesian model in terms of appropriately transformed observables (de-trended, de-seasonalised, ...) and β_1 alone (instead of β). Care should be taken, however, about the form of the prior density in such reduced Bayesian models in order to guarantee coherence with the original model.

The aim of this paper is to formally discuss such coherence issues within a generalized normal linear regression model and to weaken its assumptions in order to cover non-normal elliptical sampling distributions, autoregressive specifications, additional nuisance parameters and multi-equation framework. We consider not only posterior inference on parameters of interest (β_1), but also the calculation of Bayes factors for competing specifications. The main result is that using transformed data and reduced parameter spaces is fully coherent with the Bayesian model for original observables and parameters, provided the prior density for the scale or precision parameters is appropriately modified.

The motivation for such an elementary Bayesian study came mainly from Kleibergen and Paap (2002). In that paper the VAR(1) model was built for highly transformed economic time series in order to exclusively focus on long-run properties. However, the prior density was not modified accordingly. Although the appropriate correction of the prior can be less important in practice, there is still some need for a basic theoretical Bayesian study focused on coherence of specifications for original and transformed variables.

The structure of the paper is as follows. In the next section the basic model framework is presented and the simplest Bayesian result is formulated. Important generalizations are discussed in sections 3 and 4.

2 The basic multiple regression framework

We consider the generalized regression model (with non-random regressors)

$$\underset{(n\times 1)}{y} = \underset{(n\times K)}{X}\underset{(K\times 1)}{\beta} + \varepsilon = \underset{(n\times K_1)}{X_1}\underset{(K_1\times 1)}{\beta_1} + \underset{(n\times K_2)}{X_2}\underset{(K_2\times 1)}{\beta_2} + \varepsilon, \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \tau^{-1}\Omega, \tag{1}$$

where $K = K_1 + K_2 < n$, X has full column rank K, Ω is symmetric positive definite and β_2 as well as $\tau > 0$ are nuisance parameters. The K_1-dimensional sub-vector (corresponding to β_1) of the GLS estimator of β can be written as $b_1 = \left(\widetilde{X}_1'\Omega^{-1}\widetilde{X}_1\right)^{-1}\widetilde{X}_1'\Omega^{-1}\widetilde{y}$ where $\widetilde{y} = My$, $\widetilde{X}_1 = MX_1$, $M = I_n - X_2\left(X_2'\Omega^{-1}X_2\right)^{-1}X_2'\Omega^{-1}$. That is, b_1 can be obtained using GLS directly to the appropriately transformed variables, as if

$$\underset{(n\times 1)}{\widetilde{y}} = \underset{(n\times K_1)}{\widetilde{X}_1}\underset{(K_1\times 1)}{\beta_1} + \varepsilon^*, \qquad E(\varepsilon^*) = 0, \quad V(\varepsilon^*) = \tau^{-1}\Omega. \tag{2}$$

We call (2) a quasi-model, because the true relationship is

$$\widetilde{y} = \widetilde{X}_1\beta_1 + \widetilde{\varepsilon}, \qquad \widetilde{\varepsilon} = M\varepsilon,\ E(\widetilde{\varepsilon}) = 0,\ V(\widetilde{\varepsilon}) = \tau^{-1}M\Omega M',$$

with a singular covariance matrix as M is singular. An obvious Bayesian application of these consequences of purely algebraic properties of the GLS estimator can be stated as follows:

Lemma 1. *Let the error term in (1) be n-variate Normal, let X be non-random and let Ω be fully known. Assume the prior density $p(\beta,\tau) = p(\beta_1,\tau)\,p(\beta_2) = c_2 p(\beta_1,\tau)$, $\beta_2 \in \mathbb{R}^{K_2}$ where c_2 is an arbitrary positive constant (i.e., the prior density for β_2 is improper uniform). Then $p(y,\beta_1,\tau) \propto \tau^{-\frac{K_2}{2}}p(\beta_1,\tau)\,f_N^n\left(\widetilde{y}|\widetilde{X}_1\beta_1,\tau^{-1}\Omega\right)$ i.e., the marginal density $p(y,\beta_1,\tau)$, and thus the posterior density for (β_1,τ), can be obtained as if the transformed data followed the auxiliary quasi-model (2) and we took $p_*(\beta_1,\tau) \propto \tau^{-\frac{K_2}{2}}p(\beta_1,\tau)$ as the prior.*

Remark that $f_N^n(w|a,D) = (2\pi)^{-\frac{n}{2}}(\det D)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(w-a)'D^{-1}(w-a)\right]$ is here used to denote not only the probability density function of the Normal distribution. When applied to $\widetilde{y} = My$, with singular M, it gives the functional form of the improper density of y (given β_1 and τ) that appears after integrating out β_2 from the original Bayesian model (which is a σ-finite measure, but not a probability measure).

Proof: Our assumptions lead to the Bayesian model characterised by the joint density function $p(y,\beta_1,\beta_2,\tau) = c_2 p(\beta_1,\tau) f_N^n\left(y|X_1\beta_1+X_2\beta_2,\tau^{-1}\Omega\right)$, which can also be written as

$$\begin{aligned} p(y,\beta_1,\beta_2,\tau) \quad = \quad & c_2 (2\pi)^{\frac{K_2}{2}} \left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}} \tau^{-\frac{K_2}{2}} p(\beta_1,\tau)\cdot \\ & \cdot f_N^n\left(\widetilde{y}|\widehat{X}_1\beta_1,\tau^{-1}\Omega\right)\cdot \\ & \cdot f_N^{K_2}\left(\beta_2|\widehat{\beta}_2(\beta_1),\tau^{-1}\left(X_2'\Omega^{-1}X_2\right)^{-1}\right), \end{aligned} \tag{3}$$

where $\widehat{\beta}_2(\beta_1) = \left(X_2'\Omega^{-1}X_2\right)^{-1} X_2'\Omega^{-1}(y - X_1\beta_1)$ is the conditional GLS estimate of β_2 (given β_1). Integrating out the nuisance parameter vector β_2 we obtain the desired result:

$$p(y,\beta_1,\tau) = c_2 (2\pi)^{\frac{K_2}{2}} \left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}} \tau^{-\frac{K_2}{2}} p(\beta_1,\tau) f_N^n\left(\widetilde{y}|\widetilde{X}_1\beta_1,\tau^{-1}\Omega\right), \tag{4}$$

where all positive constants can be omitted for the purpose of posterior inference.

3 Useful generalisations

The assumptions of Lemma 1 are very strong and they unnecessarily restrict the scope of Bayesian applications of the Frisch-Waugh theorem. So we consider important extensions: random regressors depending on y, non-Normal elliptical distributions of ε and, finally, unknown Ω (in the next section). We would like to cover autoregressive models first. Their crucial feature is that the transformation from ε to y in (1) is one to one with unitary Jacobian. Note that in the case of autoregressive models we condition on some initial observations without making this conditioning explicit in our notation.

Lemma 2. *Let the error term in (1) be n-variate Normal with fully known positive definite matrix Ω, let X be such a function of y that the transformation from ε to y is 1 - 1 with unitary Jacobian (and X is of full column rank with probability 1). Then, under the prior $p(\beta,\tau) = p(\beta_1,\tau)\,p(\beta_2) = c_2 p(\beta_1,\tau)$, one obtains the following marginal density*

$$p(y,\beta_1,\tau) \propto \left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}} \tau^{-\frac{K_2}{2}} p(\beta_1,\tau) f_N^n\left(\widetilde{y}|\widetilde{X}_1\beta_1,\tau^{-1}\Omega\right).$$

Proof: Due to the assumption of unitary Jacobian, the Bayesian model is now represented by the joint density function of exactly the same form as in (3) and, thus, (4) holds. The only difference is that now X_1 and X_2 are functions of y, so we cannot omit the determinant in (4).

Hence, despite random regressors depending on y, the posterior density for (β_1, τ) can be obtained using the auxiliary quasi-model (2) and $p_*(\beta_1, \tau) \propto \tau^{-\frac{K_2}{2}} p(\beta_1, \tau)$ as the prior.
The results presented in Lemma 1 and 2 are important not only for posterior inference on the parameter of interest (β_1) in one particular regression or autoregressive model, but also for Bayesian model comparison through Bayes factors and posterior odds, provided that the competing models differ only by the part involving parameters of interest. Assume that $y = X_1^{(j)}\beta_1^{(j)} + X_2\beta_2 + \varepsilon$ (with $\beta_1^{(j)}$ of dimension $K_1^{(j)}$; $j = 1, 2$) represent alternative models for the same y. These specifications fulfil all assumptions of Lemma 2, with $p^{(j)}\left(\beta_1^{(j)}, \tau\right)$ as priors for τ and model specific parameters. Using (4) for each model separately, we can write the Bayes factor $B_{12} = \dfrac{p^{(1)}(y)}{p^{(2)}(y)}$ as the ratio of two integrals, A_1 and A_2, where

$$A_j = \int\int \tau^{-\frac{K_2}{2}} p\left(\beta_1^{(j)}, \tau\right) f_N^n\left(\widetilde{y}|\widetilde{X}_1^{(j)}\beta_1^{(j)}, \tau^{-1}\Omega\right) d\tau\, d\beta_1^{(j)} \quad (j = 1, 2).$$

Thus, the fully correct Bayes factor can be obtained using two auxiliary Bayesian models for transformed data (that is, two quasi-models of the form (2) together with appropriately modified priors). The formal explanation is that $p^{(1)}(y)$ and $p^{(2)}(y)$ have common part equal to $c\,(2\pi)^{\frac{K_2}{2}}\left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}}$ (see (4)), which cancels when the Bayes factor is calculated.
In most cases the value of the Bayes factor cannot be obtained using the analytical approach. But if $p\left(\beta_1^{(j)}, \tau\right) = f_N^{K_1^{(j)}}\left(\beta_1^{(j)}|a^{(j)}, \tau^{-1}D^{(j)}\right) f_G\left(\tau|\frac{n_0}{2}, \frac{s_0}{2}\right)$, i.e. the priors are of the natural conjugate Normal-Gamma form, then the Bayes factor B_{12} can be expressed as $B_{12} = \frac{A_1}{A_2} = \left[\frac{\det\left(\Omega+\widetilde{X}_1^{(2)}D^{(2)}\widetilde{X}_1^{(2)\prime}\right)}{\det\left(\Omega+\widetilde{X}_1^{(1)}D^{(1)}\widetilde{X}_1^{(1)\prime}\right)}\right]^{\frac{1}{2}}\left(\frac{S_2}{S_1}\right)^{\frac{n+n_0-K_2}{2}}$, where $S_j = s_0 + \left(\widetilde{y} - \widetilde{X}_1^{(j)}a^{(j)}\right)'\left(\Omega + \widetilde{X}_1^{(j)}D^{(j)}\widetilde{X}_1^{(j)\prime}\right)^{-1}\left(\widetilde{y} - \widetilde{X}_1^{(j)}a^{(j)}\right)$. Note that in the prior structure assumed here the marginal Gamma prior for τ (with mean $\frac{n_0}{s_0}$ and variance $2\frac{n_0}{(s_0)^2}$) is common, only the conditional Normal priors for model-specific parameters differ. This leads to so simple analytical results. The factor $\left(\frac{S_2}{S_1}\right)^{-\frac{K_2}{2}}$ in B_{12} corresponds to the term $\tau^{-\frac{K_2}{2}}$ in the integrands of A_1 and A_2. Thus, the lack of the prior correction term $\tau^{-\frac{K_2}{2}}$ would result in an incorrect value $B_{12}^* = B_{12}\left(\frac{S_2}{S_1}\right)^{\frac{K_2}{2}}$ instead of the true Bayes factor B_{12} for comparing specifications $y = X_1^{(j)}\beta_1^{(j)} + X_2\beta_2 + \varepsilon$ $(j = 1, 2)$ under an improper uniform prior of common β_2.
As yet, we have assumed a general prior structure for (β_1, τ). But Osiewalski and Steel (1993a,b) showed that the Jeffreys prior of the precision parameter τ ensures perfect robustness of the form of $p(y, \beta)$ with respect to changes of the distribution of ε within

the class of all n-variate continuous ellipsoidal distributions. This immediately leads to:

Lemma 3. *Let the error term of the relation* $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ *be n-variate ellipsoidal with the density function* $p(\varepsilon|\beta_1, \beta_2, \tau) = \left[\det\left(\tau^{-1}\Omega\right)\right]^{-\frac{1}{2}} g\left(\tau \cdot \varepsilon'\Omega^{-1}\varepsilon\right)$, *where g is a non-negative function such that* $\int_0^\infty g(z)z^{\frac{n-2}{2}}\,dz = \pi^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)$. *Let Ω be fully known and let $X = [X_1\ X_2]$ be such a function of y that the transformation from ε to y is 1 - 1 with unitary Jacobian (and X is of full column rank with probability 1). Assume an improper prior* $p(\beta,\tau) = p(\beta_1)\,p(\beta_2)\,p(\tau) = c\tau^{-1}p(\beta_1)$, $\beta_2 \in \mathbb{R}^{K_2}$, $\tau \in \mathbb{R}^1_+$. *Then the true marginal density function of y and β_1,* $p(y,\beta_1)$, *can be obtained by integrating out τ from the auxiliary function*

$$p_N(y,\beta_1,\tau) = c\,(2\pi)^{\frac{K_2}{2}}\left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}}\tau^{-\frac{K_2+2}{2}}p(\beta_1)\,f_N^n\left(\widetilde{y}|\widetilde{X}_1\beta_1,\tau^{-1}\Omega\right). \quad (5)$$

Proof: The Bayesian model is now represented by the joint density

$$\begin{aligned} p(y,\beta,\tau) &= c\tau^{-1}p(\beta_1)\left[\det\left(\tau^{-1}\Omega\right)\right]^{-\frac{1}{2}} g\left(\tau\varepsilon'\Omega^{-1}\varepsilon\right) \\ &= c\tau^{\frac{n-2}{2}}p(\beta_1)\det(\Omega)^{-\frac{1}{2}} g\left(\tau\varepsilon'\Omega^{-1}\varepsilon\right). \end{aligned}$$

where $\varepsilon = y - (X_1\beta_1 + X_2\beta_2)$. Using the basic property of the nonnegative function g we easily integrate τ out and obtain $p(y,\beta) = c\,(\pi)^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)(\det\Omega)^{-\frac{1}{2}}p(\beta_1)\left(\varepsilon'\Omega^{-1}\varepsilon\right)^{-\frac{n}{2}}$, which does not depend on the particular form of g and, thus, is the same as under Normality of the error vector, i.e., as for $g(z) = (2\pi)^{-\frac{n}{2}}\exp\left(-\frac{z}{2}\right)$; see Osiewalski and Steel (1993a,b). Hence, in order to derive $p(y,\beta_1)$, we can work with the Bayesian model

$$p_N(y,\beta_1,\beta_2,\tau) = c\tau^{-1}p(\beta_1)\,f_N^n\left(y|X_1\beta_1 + X_2\beta_2,\tau^{-1}\Omega\right),$$

which is marginally equivalent to the original specification. Using Lemmas 1 and 2 we get (5) and, thus, $p(y,\beta_1) = p_N(y,\beta_1) = \int_0^\infty p_N(y,\beta_1,\tau)\,d\tau$.

Lemma 3 explains that the marginal posterior distribution of β_1 alone (but not of τ) can be obtained using transformed data, provisionally assuming for them the normal quasi-model (2) and the prior $p_*(\beta_1,\tau) \propto \tau^{-\frac{K_2+2}{2}}p(\beta_1)$. The fact that posterior inference on τ is sensitive to the particular form of the ellipsoidal distribution of the error vector is of little importance as τ is a nuisance parameter (as well as β_2).

Under an improper uniform prior for β_1, the additional factor $\tau^{-\frac{K_2}{2}}$ in the prior specification for the auxiliary quasi-model (2) has an obvious interpretation. It guarantees the correct degrees of freedom ($n - K_1 - K_2$ instead of $n - K_1$) of the Student t marginal posterior distribution of β_1. Working with transformed data means integrating out nuisance regression parameters first. This has to be reflected in the Bayesian quasi-model in order to assure its coherence in terms of $p(y, \beta_1)$ with the original Bayesian model for untransformed data.

4 Unknown Ω and multi-equation specifications

If the symmetric positive definite matrix Ω is not fully known, but it is only a known function of some additional unknown parameters, say λ, then things may significantly complicate. Let us keep all the assumptions of Lemma 3, except that now $\Omega = \Omega(\lambda)$ and the prior is defined on a larger space: $p(\beta, \lambda, \tau) = p(\beta_1, \lambda)\, p(\beta_2)\, p(\tau) = c\tau^{-1} p(\beta_1, \lambda)$. The same reasoning as in two preceding sections leads to the conclusion that the marginal density function of y, β_1 and λ, $p(y, \beta_1, \lambda)$, can be obtained by integrating out τ from the function

$$p_N(y, \beta_1, \lambda, \tau) = c\,(2\pi)^{\frac{K_2}{2}} \left[\det\left(X_2'\Omega^{-1}X_2\right)\right]^{-\frac{1}{2}} \tau^{-\frac{K_2+2}{2}} p(\beta_1, \lambda)\, f_N^n\left(\widetilde{y}|\widetilde{X}_1\beta_1, \tau^{-1}\Omega\right), \tag{6}$$

where both the determinant and, more importantly, the data transformation depend on λ through Ω. Such dependence makes (6) practically useless.

There is a model structure, however, where - despite the presence of unknown λ - the data transformation does not involve parameters and, additionally, the determinant in (6) is multiplicatively separable in the data and unknown λ. This occurs in multi-equation SURE or VAR models due to their Kronecker product structure. Let $n = mT$ (n has been used to denote the dimension of y; now y is divided into m blocks, each of dimension T) and assume $X_2 = I_m \otimes Z_2$, $\tau^{-1}\Omega = \Sigma \otimes I_T = \tau^{-1}\left(C^{-1} \otimes I_T\right)$ where Z_2 is $T \times k_2$ and it has full rank k_2 (the column dimension of X_2 is $K_2 = k_2 m$), Σ is an unknown $m \times m$ symmetric, positive definite matrix, and (τ, C) are such that $\tau C = \Sigma^{-1}$ and $c_{11} = 1$ in C. In this case one easily obtains $X_2'\Omega^{-1}X_2 = C \otimes (Z_2'Z_2)$ and

$$M = I_m \otimes \left[I_T - Z_2\left(Z_2'Z_2\right)^{-1} Z_2'\right]. \tag{7}$$

Thus, the transformation of y and X_1, appearing in the last term of (6), solely depends on Z_2 and the determinant in (6) has the required product structure. The form of the model with so simple algebraic properties is well known; its i-th equation ($i = 1, \ldots, m$) can be written as

$$\underset{(T\times 1)}{y_{(i)}} = \underset{(T\times k_{1(i)})}{Z_{1(i)}} \underset{(k_{1(i)}\times 1)}{\beta_{1(i)}} + \underset{(T\times k_2)}{Z_2} \underset{(k_2\times 1)}{\beta_{2(i)}} + \varepsilon_{(i)}. \tag{8}$$

where the first part of the structure allows for possibly different variables in each equation and cross-equation restrictions, while the second part assumes the same k_2 regressors in each equation and completely free nuisance coefficients (with no cross-equation restrictions). In order to put (8) into the framework of the initial specification

(1) we write

$$y = \begin{bmatrix} y_{(1)} \\ \cdots \\ y_{(m)} \end{bmatrix},\ X_1 = \begin{bmatrix} Z_{1(1)} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & Z_{1(m)} \end{bmatrix},\ \beta_1 = \begin{bmatrix} \beta_{1(1)} \\ \cdots \\ \beta_{1(m)} \end{bmatrix},$$

$$X_2 = I_m \otimes Z_2,\ \beta_2 = \begin{bmatrix} \beta_{2(1)} \\ \cdots \\ \beta_{2(m)} \end{bmatrix},\ \varepsilon = \begin{bmatrix} \varepsilon_{(1)} \\ \cdots \\ \varepsilon_{(m)} \end{bmatrix}$$

and stress that only X_2 need to have the Kronecker product structure, although both X_1 and X_2 are block-diagonal (the column dimension and rank of X_1 is $K_1 = \sum_{i=1}^{m} k_{1(i)}$, the same as the row dimension of β_1). Finally, let λ be the column vector of dimension $L = \frac{(m+1)m-2}{2}$ that consists of all distinct unknown entries of C; formally vech(C)=$[1\ \lambda']'$.

Theorem. *Let ε, the vector grouping the error terms of the whole system (8), be mT-variate ellipsoidal with the density function $p\left(\varepsilon|\beta_1,\beta_2,\Sigma^{-1}\right) = \left[\det\left(\Sigma^{-1}\otimes I_T\right)\right]^{\frac{1}{2}} g\left[\varepsilon'\left(\Sigma^{-1}\otimes I_T\right)\varepsilon\right]$, where g is a non-negative function such that $\int_0^\infty g(z) z^{\frac{n-2}{2}}\, dz = \pi^{-\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)$. Let X_1 and X_2 be such functions of y that the transformation from ε to y is 1 - 1 with unitary Jacobian. Assume an improper prior $p\left(\beta,\Sigma^{-1}\right) = p\left(\beta_1\right)p\left(\beta_2\right)p\left(\Sigma^{-1}\right) = cp\left(\beta_1\right)\left[\det\left(\Sigma^{-1}\right)\right]^{-\frac{m+1}{2}}$. Then the true marginal density function of y and β_1, $p(y,\beta_1)$, can be obtained by integrating out Σ^{-1} from the auxiliary function*

$$\begin{aligned} p_N\left(y,\beta_1,\Sigma^{-1}\right) &= c\,(2\pi)^{\frac{mk_2}{2}}\left[\det\left(Z_2'Z_2\right)\right]^{-\frac{m}{2}}\left[\det\left(\Sigma^{-1}\right)\right]^{-\frac{k_2+m+1}{2}} \\ &\cdot p\left(\beta_1\right) f_N^{mT}\left(\widetilde{y}|\widetilde{X}_1\beta_1,\Sigma\otimes I_T\right). \end{aligned} \tag{9}$$

Proof: Remind that $\Sigma^{-1} = \tau C$ and vech(C) =$[1\ \lambda']'$. Thus, vech$\left(\Sigma^{-1}\right) = \tau[1\ \lambda']'$, the Jacobian of the transformation from Σ^{-1} to (τ, C) is τ^L and, in terms of $(\lambda,\ \tau)$, the prior is

$$p\left(\beta,\tau,\lambda\right) = cp\left(\beta_1\right)\left[\det\left(\tau C\right)\right]^{-\frac{m+1}{2}}\tau^{\frac{m(m+1)-2}{2}} = cp\left(\beta_1\right)\det\left(C\right)^{-\frac{m+1}{2}}\tau^{-1},$$

i.e. of the form $p\left(\beta,\tau,\lambda\right) = c\tau^{-1}p\left(\beta_1,\lambda\right)$, which, for any g, guarantees perfect marginal equivalence of our Bayesian model $p\left(y,\beta_1,\beta_2,\tau,\lambda\right)$ and the auxiliary Bayesian model $p_N\left(y,\beta_1,\beta_2,\tau,\lambda\right)$ based on the Normality assumption for the error vector. That is, $p\left(y,\beta_1,\beta_2,\lambda\right)$ and $p_N\left(y,\beta_1,\beta_2,\lambda\right)$ coincide, and so do their respective marginal density functions $p(y,\beta_1)$ and $p_N(y,\beta_1)$. But, in order to derive the latter, we can first integrate β_2 from $p_N(y,\beta_1,\beta_2,\tau,\lambda)$; this leads to (6), which in the case of model (8) specialises to

$$\begin{aligned} p_N(y,\beta_1,\tau,\lambda) &= c\,(2\pi)^{\frac{mk_2}{2}}\left[\det(Z_2'Z_2)\right]^{-\frac{m}{2}}(\det C)^{-\frac{k_2+m+1}{2}}\,\tau^{-\frac{mk_2+2}{2}} \\ &\quad\cdot p(\beta_1)\,f_N^{mT}\left(\widetilde{y}|\widetilde{X}_1\beta_1,\tau^{-1}\left(C^{-1}\otimes I_T\right)\right), \end{aligned} \tag{10}$$

where the data transformation matrix is given in (7). The change of parameterisation in (10), from (τ, C) back to Σ^{-1} (with Jacobian equal to τ^{-L}), leads to $p_N\left(y,\beta_1,\Sigma^{-1}\right)$ in (9), which yields the required marginal density function $p_N(y,\beta_1)=p(y,\beta_1)$.
In order to perform valid posterior inference on $\beta_{1(i)}$'s in the m-equation system (8), we can use the auxiliary quasi-model for transformed data, i.e. the last factor in (9), but we have to multiply the original prior by $\left[\det\left(\Sigma^{-1}\right)\right]^{-\frac{k_2}{2}}$. In the specific case of a m-variate regression model, i.e. for $k_{1(i)}=k_1$ $(i=1,\ldots,m)$ and $X_1=I_m\otimes Z_1$, with the improper uniform prior for the K_1–(i.e., mk_1–)dimensional vector β_1 (grouping all $\beta_{1(i)}$'s), the marginal posterior distribution of $\beta_{1(i)}$ in equation i is k_1-variate Student t with $T-k_1-k_2-m+1$ degrees of freedom (see Zellner 1971), not $T-k_1-m+1$. For systems with large number of nuisance coefficients (k_2) and small sample size T (short time series) the degrees of freedom correction can be practically important.
As explained in Section 3, our results are important not only for posterior inference on β_1. They also justify using auxiliary quasi-models for transformed data together with appropriately modified priors when one wants to make the formal Bayesian comparison of different m-equation systems (8) which differ only in the part involving $\beta_{1(i)}$'s.

References

[1] Frisch R., Waugh F.V., 1933, Partial time regressions as compared with individual trends, *Econometrica* 1, 387–401.

[2] Greene W.H., 2003, *Econometric Analysis*, Fifth Edition, Prentice Hall, Upper Saddle River, NJ.

[3] Kleibergen F., Paap R., 2002, Priors, posteriors and Bayes factors for Bayesian analysis of cointegration, *Journal of Econometrics* 111, 223–249.

[4] Osiewalski J., Steel M.F.J., 1993a, Robust Bayesian inference in elliptical regression models, *Journal of Econometrics* 57, 345–363.

[5] Osiewalski J., Steel M.F.J., 1993b, Bayesian marginal equivalence of elliptical regression models, *Journal of Econometrics* 59, 391–403.

[6] Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, J. Wiley, New York.