The "Pathology" of the Natural Conjugate Prior Density in the Regression Model

Luc BAUWENS *

ABSTRACT. – In a Bayesian analysis of the linear regression model, one may have prior information on a subset of the regression coefficients, but one has usually no prior information on the error variance. If one incorporates this kind of information in a natural conjugate prior density, under certain conditions the posterior mean of the coefficients on which one is informative is equal to the prior mean, and the posterior mean of the coefficients on which one is not informative is equal to a constrained least squares estimator. The value of the posterior covariance matrix is also studied. We discuss and illustrate how to avoid getting posterior results too close to the "pathological" results summarized above.

La « pathologie » de la loi *a priori* naturelle conjuguée dans le modèle de régression

RÉSUMÉ. – Pour une analyse bayesienne du modèle de régression linéaire, on peut disposer d'information *a priori* sur une partie des coefficients de la régression, mais on ne dispose pas habituellement d'information *a priori* sur la variance des erreurs. Si l'on incorpore ce type d'information dans une loi *a priori* naturelle conjuguée, sous certaines conditions la moyenne *a posteriori* des coefficients sur lesquels on est informatif est égale à la moyenne *a priori*, et la moyenne *a posteriori* des coefficients sur lesquels on n'est pas informatif est égale à un estimateur de moindres carrés contraint. Nous étudions également la valeur de la matrice de covariance *a posteriori*. Nous proposons et illustrons une façon d'éviter d'obtenir des résultats *a posteriori* trop proches des résultats pathologiques résumés ci-dessus.

---

* L. BAUWENS : CORE, Université Catholique de Louvain, 34, Voie du Roman Pays, B-1348 Louvain-La-Neuve, Belgique. I thank participants of seminars at Montréal and Rotterdam for their suggestions or remarks when I presented this paper. Fredéric Aprahamian, Michel Lubrano, Mark Steel and two referees helped me with their detailed comments. None of them could be blamed for any error that could remain in this paper.

1 Introduction

In a Bayesian analysis of the linear regression model, one may have prior information on a subset of the regression coefficients, but one has usually no prior information on the variance of the error term. For example, in a dynamic model, one may have prior ideas, suggested by economic theories, on the values of long run elasticities, but it is more difficult and less important to have prior ideas about short run effects. Another example, as in LEAMER [1982], is when a subset of the exogenous variables are treated as doubtful and a prior mean equal to zero is assigned to their coefficients, while the coefficients of the other variables are endowed with a diffuse prior. The variance of the error term may be considered often as a nuisance parameter, implying that inference on the parameters of interest should be done marginally, and that it is not desirable to elicit an informative prior on it.

A convenient class of prior distributions for a Bayesian analysis of the regression model is the natural conjugate one. It can be defined as an inverted gamma density on the variance of the error term, times a normal density on the coefficients, conditional on the variance of the error term; the prior conditional covariance matrix of the coefficients is proportional to the variance of the error term, but their prior mean does not depend on it. The implied marginal prior density of the coefficients is then in the Student family.

It was shown by RICHARD [1973, pp. 180-182], that, for fixed *marginal* prior mean and variance on the regression coefficient of a model with a *single* regressor, if the prior inverted gamma density on the variance of the error term degenerates in a specified way to a diffuse limit, the posterior mean of the coefficient tends to the prior mean and the posterior variance tends to the prior variance times a constant smaller than one. At the limit, one does not learn anything from the data. The moral of this story is that one should be careful, when using a natural conjugate prior, not to be too diffuse on the variance of the error term when one is *marginally* informative on the coefficients. To our knowledge, this pitfall has not been mentioned in the literature; e. g. it is not found in the books of RAIFFA and SCHLAIFER [1961], ZELLNER [1971], LEAMER [1978], and JUDGE *et al.* [1985], although the last two contain a critical evaluation of the natural conjugate prior density. The main drawback of the natural conjugate prior density in the regression model is that it does not reveal easily a conflict of information between the prior and the sample. More generally, conjugate priors have been criticized because they are chosen for their easiness of computation (of posterior results), rather than for their ability to represent prior information, see e. g. BERGER [1985].

In this paper, we focus on the issue of the specification of the limiting process by which the prior inverted gamma density of the variance of the error term tends to a diffuse limit. We introduce a type of convergence that is more natural than Richard's, as it yields essentially Jeffreys' invariant prior. Moreover, we discuss directly the case of several regressors; it is then relevant to distinguish between the coefficients on which one has some prior information and the other coefficients on which one remains non-informative.

Our main result can be summarized as follows: with a natural conjugate prior density that is not informative on the variance of the error term, under certain conditions the coefficients on which one is informative become *a posteriori* equal to their prior mean *with probability equal to one.* This result is "pathological": although the prior information is not dogmatic, it dominates completely the sample information. It is peculiar to the natural conjugate prior density and does not occur for any prior with the same informative content (in terms of prior expectation and covariance matrix), provided the latter imposes prior independence between the variance of the error term and the regression coefficients.

The results are obtained by computing the limiting values of the posterior mean and covariance matrix when the prior density on the variance of the error term degenerates to a diffuse limit. An important reason to consider this problem is that Bayesians should be interested in the sensitivity of the posterior results with respect to the choice of the prior. In an empirical Bayesian analysis, care should be taken not to approach the limit inadvertantly.

In Section 2, we introduce the notation and state precisely the results and the conditions under which they are valid. The proofs are left to an Appendix. In Section 3, we propose a practical procedure to prevent getting posterior results that are too close to the "pathological" results summarized above. An empirical example illustrates the arguments.

2 The "Pathology"

We consider the linear regression model

$$y \mid X, \beta, \sigma^2 \sim N_T(X_1 \beta_1 + X_2 \beta_2, \sigma^2 I_T), \tag{1}$$

where y is $T \times 1$, $X = (X_1 X_2)$ is $T \times k$, $\beta = (\beta_1' \beta_2')'$ is $k \times 1$, X_i is $T \times k_i$, and β_i is $k_i \times 1$ ($i = 1$ or 2), with $k = k_1 + k_2$, and N_T denotes a T-variate normal distribution. We shall use the notation

$$\begin{gathered} S = X'X, \qquad S_i = X_i' X_i, \\ S_{21} = S_{12}' = X_2' X_1, \qquad S_{1.2} = S_1 - S_{12} S_2^{-1} S_{21} \\ \hat{b} = S^{-1} X' y = (\hat{b}_1' \hat{b}_2')', \\ r = y'y - \hat{b}' S \hat{b}, \qquad \bar{b}_2 = S_2^{-1} X_2' y. \end{gathered} \tag{2}$$

Regarding the prior information, the assumptions are:

(1) the prior is informative on β_1 and σ^2, and it is not informative on β_2;

(2) the prior information on β_1 has been elicited in terms of a location parameter b_1 and a spread matrix V_1, typically a mean vector and a covariance matrix;

(3) the prior distribution is selected in the natural conjugate family, *i. e.*

$$\begin{cases} \beta \mid \sigma^2 \sim N_k(b, \sigma^2 M^{-1}), \\ \sigma^2 \sim IG(s, \alpha), \end{cases} \quad (3)$$

[1]

where

$$b = \begin{bmatrix} b_1 \\ 0 \end{bmatrix}, \qquad M = \begin{bmatrix} e V_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad (4)$$

and

$$e = s/(\alpha - 2) = E(\sigma^2), \qquad \text{if} \quad \alpha > 2. \quad (5)$$

The second equality of (5) follows from the property that the marginal prior density of σ^2 is inverted gamma with positive parameters s and α, *i. e.* it has the kernel

$$f(\sigma^2) = (\sigma^2)^{-(\alpha+2)/2} \exp\left(-\frac{1}{2} s/\sigma^2\right). \quad (6)$$

The zero entries in M imply indeed that the prior on β_2 is flat over the parameter space of β_2. [2] The vector b_1 is equal to $E(\beta_1 \mid \sigma^2)$ and to $E(\beta_1)$. The matrix V_1 is equal to the unconditional covariance matrix $V(\beta_1)$, but $V(\beta_1 \mid \sigma^2)$ is equal to $\sigma^2 V_1/e$. [3]

The posterior parameters are given by the usual formula of convolution of prior parameters and sample sufficient statistics, reminded here for convenience:

$$\begin{aligned} P &= M + S, \\ \tilde{b} &= P^{-1}(M b + S \hat{b}), \\ \tilde{s} &= s + r + (b - \hat{b})' M P^{-1} S (b - \hat{b}), \\ \tilde{\alpha} &= \alpha + T. \end{aligned} \quad (7)$$

---

1. Meaning that σ^2 is marginally distributed as an inverted gamma random variable with scale parameter s and α degrees of freedom. The choice of s and α is discussed in Section 3.
2. If the vector b_2 of b, corresponding to β_2 of β, is assigned any value other than 0, the results given in this paper are not changed.
3. Remember that the prior density of β_1, *conditional* on σ^2, is normal, with mean and precision matrix given in (4), and that the *marginal* (or unconditional) prior density of β_1 is Student, with mean b_1, covariance matrix V_1, and α degrees of freedom.

We study the limiting behaviour of the posterior mean $\tilde{b}$, and of the posterior covariance matrix

$$\tilde{V} = \tilde{s} P^{-1}/(\tilde{\alpha} - 2), \tag{8}$$

when the prior density is the natural conjugate one described above, and when the prior density of σ^2 tends to become diffuse. More precisely, we assume

(i) that b_1 and V_1 are *held fixed, i. e.* the marginal prior mean and covariance matrix of β_1 are given (e. g. because they represent genuine prior information), and

(ii) that s tends to zero, and α tends to two (from above) in such a way that e tends to infinity, see (5). This is our specification of the limiting process by which the prior inverted gamma density of σ^2 tends to a diffuse limit, and deserves a detailed explanation.

From (6), if s tends to 0 and α tends to two, it is obvious that the kernel [4] of the prior density of σ^2 tends to $(1/\sigma^2)^2$. A usual diffuse prior for σ^2 is proportional to $1/\sigma^2$, and can be obtained by assuming that both s and α tend to zero. [5] We assume that α tends to two rather than zero because we have assumed above that the prior marginal covariance matrix of β_1 exists and this requires that $E(\sigma^2)$ exists, *i. e.* $\alpha > 2$. In Appendix 1, we discuss the case where *both* α *and* s *tend to zero* and where the marginal probability that β_1 belongs to a certain region is held fixed, instead of assuming fixed prior mean and covariance matrix of β_1.

Returning to the case where both s and $\alpha - 2$ tend to zero, given (4), it may happen that e tends to zero, remains constant, or tends to infinity, depending on the relative convergence speed of s and $\alpha - 2$. The *first case* is not interesting because it implies that M_1, being equal to $e V_1^{-1}$, tends to a null matrix as α approaches two, so that the prior information on β_1 becomes more and more diffuse in the natural conjugate prior density, contrary to what we assume. As the limit of α is reached, the scalar $e = s/(\alpha - 2)$ is of course no longer interpretable as $E(\sigma^2)$. In the *second case*, e remains constant as long as α is larger than two, and is arbitrarily large as α reaches its limit. In the *third case*, e tends smoothly to infinity. It is important to remember that as the mean of σ^2 increases, so does its variance. Therefore, this specification corresponds to decreasing prior knowledge about σ^2, but not to the prior idea that the model (1) becomes meaningless because one believes that its error term has a very large variance.

RICHARD [1973, pp. 180-182] assumes also that e tends smoothly to infinity, but by fixing α and increasing s. This specification of the process by which the prior density of σ^2 becomes diffuse is to some extent questionable

---

4. In order to obtain a diffuse prior by a limit argument it is necessary to take the limit of the kernel, not of the density, otherwise the limit may be equal to zero because the normalizing constant of the density tends to zero. It is usual anyway to drop from the prior density all the factors that do not depend on the parameters of the model.
5. This prior can be justified by Jeffreys' invariance principle, see e. g. ZELLNER [1971].

since it implies that the kernel (6) of the inverted gamma density tends to zero, rather than to a standard diffuse prior.

The results are given in the next propositions, where $\tilde{b}$ and $\tilde{V}$ are partitioned conformably with β in (1). Proposition 1 gives the limit of the posterior mean, which is the same whether e tends to infinity under our specification or under Richard's one.

PROPOSITION 1: *Given* (1)-(5), *for fixed* b_1 *and* V_1, *if* $e \to \infty$, *then*

(9) $$\tilde{b}_1 \to b_1,$$

(10) $$\tilde{b}_2 \to \hat{b}_2 - S_2^{-1} S_{21} b_1.$$

The limit of $\tilde{b}_2$ is easily recognized to be the least squares estimator of β_2 in (1) under the deterministic constraint that $\beta_1 = b_1$.

Proposition 2 gives the limit of the posterior covariance matrix, which depends on the way e tends to infinity.

PROPOSITION 2: Under the same conditions as in Proposition 1,

(i) if $s \to 0$ and $\alpha \to 2$ in such a way that $e \to \infty$, then

(11) $$\begin{cases} \tilde{V}_1 \to 0, \\ \tilde{V}_2 \to [r + (b_1 - \hat{b}_1)' S_{11.2} (b_1 - \hat{b}_1)] S_2^{-1}/T \quad \text{and} \quad \tilde{V}_{21} \to 0; \end{cases}$$

(ii) if $e \to \infty$ because $s \to \infty$ for fixed α (larger than 2), then

(12) $$\begin{cases} \tilde{V}_1 \to (\alpha - 2) V_1/(\tilde{\alpha} - 2), \\ \tilde{V}_2 \to \infty S_2^{-1} \quad \text{and} \quad \tilde{V}_{21} \to -S_2^{-1} S_{21} \tilde{V}_1. \end{cases}$$

The difference between the limits of $\tilde{V}$ occurs [6] because the prior parameter s appears in the posterior parameter $\tilde{s}$, see (7) and (8). The limit of $\tilde{V}_2$ in (11) is proportional to the constrained least-squates estimator of the covariance matrix of β_2, *i. e.* $(\tilde{u}' \tilde{u}) S_2^{-1}/(T - k_2)$, where

$$\tilde{u} = y - X_1 b_1 - X_2 (\hat{b}_2 - S_2^{-1} S_{21} b_1).$$ [7]

The proofs of the propositions are given in Appendix 2. Notice that it is assumed that e tends *smoothly* to infinity, as e. g. in the *third case* distinguished above ($\alpha - 2$ tending to zero faster than s). Clearly, in the *second case* (s and $\alpha - 2$ tending to zero at the same rate), as long as α is strictly larger than two, $e V_1^{-1}$ is a finite constant matrix. Therefore the posterior mean $\tilde{b}$ does not change, and the posterior covariance matrix $\tilde{V}$ changes as the scalar $\tilde{s}/(\tilde{\alpha} - 2)$, P being constant. At $\alpha = 2$, there is a switch of $\tilde{b}$ and of $\tilde{V}$ to their limits defined by (9), (10) and (11), respectively; this

---

6. If α tends to two and s tends to infinity, the limits of V_1 and V_{21} are equal to zero as in part (i) of Proposition 2.
7. The difference between $r + (b_1 - \hat{b}_1)' S_{1.2} (b_1 - \hat{b}_1)$ and $\tilde{u}' \tilde{u}$ is equal to $2 \hat{b}_1' S_{1.2} b_1$; it can be positive or negative.

switch can be checked numerically. Combining (9) and (11), we can conclude that β_1 becomes equal at the limit to its prior mean with probability one.

The reason of these "pathological" results can be explained intuitively as follows. We assume for simplicity that $\beta_2=0$ in (1), so that β_1 is conflated with β, and $M_1\,(=e\,V_1^{-1})$ with M. In the natural conjugate prior density, β and σ^2 are not independent in probability (although they are not correlated). This is revealed by the fact that $V(\beta)=E(\sigma^2).M^{-1}$, see (3) and (5). If the prior information on σ^2 becomes more and more vague, the *marginal* prior covariance matrix of β explodes. If we wish on the contrary to keep this covariance matrix fixed, we have to blow up M, the prior *conditional* precision, which then dominates the sample precision S, see (7). The domination of M on S explains the results.

It could be argued that it is silly to keep $V(\beta)$ fixed when one increases $E(\sigma^2)$: one could as well keep M fixed, but then become less and less informative on β. This is equivalent to recognize that if one knows very little *a priori* about the noise of the process, one knows very little about its systematic component. We find the latter viewpoint questionable: it is quite possible to have prior information on the first moment of a process, but not on the second moment.

The previous discussion and results do not imply that a natural conjugate prior of the type $p(\beta, \sigma^2) \propto (1/\sigma^2).p(\beta \mid \sigma^2)$, where $p(\beta \mid \sigma^2)$ is normal with mean b and covariance matrix $\sigma^2 M^{-1}$, is meaningless. For example, ZELLNER [1986] uses this type of prior for computing posterior odds ratios for selected pairs of hypotheses about the value of β. Formulas (7) are perfectly applicable when α and s are equal to zero. The question we raise in this case is: how can one select a value of the matrix M that reflects one's prior information? The choice of the order of magnitude of M becomes arbitrary, not to say untractable, if it cannot be made in connection with *marginal* prior information on β. The only solution we are aware of is ZELLNER'S [1986] g-prior (*i. e.* $M=gX'X$), which may be too restrictive.

3 "Remedies"

If one has no prior information on σ^2 but has prior information on β, a first solution to avoid obtaining posterior results which are too close to the "pathological" results (9)-(12) is to get out of the natural conjugate framework. A convenient prior, called the independent Student prior in the sequel, is then a Student density on β, independent of σ^2, times Jeffreys' diffuse prior on σ^2. It leads to a posterior 2-0 poly-t density of β; its posterior moments can be easily computed by univariate numerical integration, whatever the dimension of β is; for details, see DRÈZE [1977]. More generally, any prior that factorizes as an independent density on β times a prior on σ^2 (that can be diffuse) can be used, provided one is ready to organize the numerical integration required to compute posterior results, e. g. by Monte Carlo methods.

We propose a solution without getting out of the natural conjugate framework. Our motivation is the greater easiness to implement tests of linear restrictions on β when its posterior distribution is Student rather than 2-0 poly-t or other. We point out that tests of linear restrictions in auxiliary regressions can serve as diagnostics to detect errors of specification; see e. g. LUBRANO and MARIMOUTOU [1988], and BAUWENS and LUBRANO [1991]. Suppose that the null hypothesis H_0 is of the kind $R\beta = r$. The test procedure we refer to consists in rejecting H_0 if the null vector is outside the highest posterior density region (of pre-selected probability) of the marginal posterior distribution of $R\beta - r$. When the posterior distribution of β is Student, the test can be implemented with a table of the F distribution. [8] When the posterior of β is not in the Student family, there is no known and easy procedure of the same kind.

Our solution consists in eliciting a set of informative prior inverted gamma density on σ^2, and conducting a sensitivity analysis. The objective is to find a particular prior that satisfies two requirements. It should not be too diffuse. It should also be "neutral" for the inference on β, in the sense that the posterior mean and covariance matrix of β obtained with the natural conjugate prior are close to their values obtained with an independent Student prior. The set of prior on σ^2 we propose is defined by letting the prior mean e of σ^2 increase from zero to the empirical variance s_y^2 of the dependent variable of (1). This set is rather large in terms of possible values of the prior parameters s and α. For given α, it reduces to the set indexed by the values of s ranging from 0 to $(\alpha - 2)s_y^2$, see (5). In terms of a prior expected pseudo-R^2, defined as

$$R_p^2 = 1 - (e/s_y^2),$$

this set of prior corresponds to R_p^2 in the interval $[0, 1)$. Pure Bayesians can object to the dependence of the prior on the sample (through s_y^2). To avoid this twist to orthodoxy, a more general but also more difficult and costly procedure explained by RICHARD and STEEL [1988, Appendix D] can be used. Anyway, a value of α can be chosen as

$$\alpha = (2/cv) + 4, \tag{13}$$

where cv is the coefficient of variation of the distribution of σ^2 defined by (3). Formula (13) comes from the property that $V(\sigma^2) = 2e^2/(\alpha - 4)$. A sensible value of α should be neither too large – see (7) and (8) – nor too

---

8. If the distribution of β is Student with mean $\tilde{b}$, covariance matrix $\tilde{V}$ given by (8), and $\tilde{\alpha}$ degrees of freedom, the distribution of $R\beta - r$ is Student with mean $R\tilde{b} - r$, covariance matrix $R\tilde{V}R'$, and $\tilde{\alpha}$ degrees of freedom; then the distribution of $(\beta - \tilde{b})'R'(RP^{-1}R')^{-1}R(\beta - \tilde{b})\tilde{\alpha}/(m\tilde{s})$ is Fisher with m and $\tilde{\alpha}$ degrees of freedom (where m is the number of restrictions). H_0 is rejected if $F_* = (R\tilde{b} - r)'(RP^{-1}R')^{-1}(R\tilde{b} - r)\tilde{\alpha}/(m\tilde{s})$ is larger than a given quantile of the Fisher distribution. If the prior on β and σ^2 is diffuse [set $M = 0$, $s = 0$, and $\alpha = -k$ in (7)], the last quantity is equal to the classical F statistic.

small (notice that (13) ensures that α is greater than 4). In practice, a few values of α, say between 4 and 10, should be tried.

If one can find empirically a couple (α, s) that is neutral for the inference on β, one can consider that the corresponding natural conjugate prior is an acceptable substitute of the independent Student prior. Equivalently, the posterior Student density is then an acceptable approximation of the posterior poly-t. Then, if one wishes to perform tests of hypotheses on β or tests of misspecification, one can use the Bayesian F-test based on the posterior Student density as an approximate procedure for the case where the posterior is poly-t. Clearly, if there is a strong conflict of information between the prior and the sample, expressed by a bimodal posterior poly-t, the procedure we propose should not be used.

Let us illustrate our solution with a regression of the growth rate of real wages (GW) on the growth rate of the consumption price index (GPC), the lagged growth rate of labor productivity (GPR), and a rate of unemployment (U).[9] The equation is a simple Phillips curve. Table 1 reports

TABLE 1

Posterior means and standard deviations for a Belgian wage equation

Prior	β_C	β_R	β_U [a]	Intercept	$\tilde{e}$ [b]	e	R_p^2
DI	.41 (.15)	.28 (.10)	−.54 (.41)	.032 (.014)	.46	–	.41 [c]
STUD . . .	.24 (.12)	.33 (.11)	−.34 (.43)	.031 (.014)	68.	–	–
NAT1 . . .	.36 (.11)	.29 (.08)	−.49 (.34)	.032 (.011)	.30	.07	.90
NAT2 . . .	.22 (.11)	.33 (.10)	−.35 (.41)	.032 (.014)	.47	.41	.40
NAT3 . . .	.17 (.11)	.35 (.11)	−.29 (.45)	.031 (.015)	.59	.69	.00
NAT4 . . .	.02 (.09)	.59 (.15)	−.08 (.96)	.020 (.032)	2.8	7.6	−10
NAT5 . . .	.00 (.08)	.73 (.16)	−.02 (2.6)	.013 (.085)	21.	70.	-10^2
CONS . . .	.00	.75	−.01 (.61)	.012 (.020)	1.1	–	.00 [c]

[a] Coefficient of U.
[b] Posterior mean of σ^2. Entries for $\tilde{e}$ and e to be divided by 1,000.
[c] Value of the classical adjusted R^2.

posterior means and standard deviations of the parameters with a diffuse prior (row DI) and several informative prior densities. The latter have marginal prior moments of β_c and β_R (the coefficients of GPC and GPR) given by

$$(14) \qquad \begin{cases} E(\beta_c)=0. \;, & V(\beta_c)=0.15^2, \\ E(\beta_R)=0.75, & V(\beta_R)=0.30^2 \end{cases}$$

The prior mean of β_c reflects the idea that there is no money illusion, *i. e.* nominal wages do not increase faster than price inflation, ceteris paribus. The prior mean of β_R is based on the idea that some fraction of productivity increases are transmitted to real wages, the fraction being of

9. There are 22 annual observations on the Belgian economy, reported in Appendix 3.

the order of the share of wage income in production. A substantial dispersion is allowed around these mean values, as indicated by the variances given in (14).

In all cases, the prior on the other coefficients is diffuse. Row STUD of Table 1 gives the results when the prior is independent Student with ten degrees of freedom ($\alpha = 10$). The rows labelled NATi ($i = 1$ to 5) correspond to natural conjugate prior densities implying (14); the prior inverted gamma density of σ^2 has parameters $\alpha = 10$ and s such that e and R_P^2 take the values reported in the last columns. One can see that the prior of σ^2 becomes more and more diffuse, as s increases for fixed α, when i increases.

The results from NAT1 to NAT5 illustrate directly Propositions 1 and 2 (ii); the last row (CONS) of Table 1 reports the constrained posterior results, using a diffuse prior. A Bayesian F-test rejects the constraints with very high level of posterior probability; notice how the classical R^2 deteriorates (comparing the CONS and DI results).

The results of STUD and NAT2 are very close. There is indeed no conflict of information between prior and sample information.[10] If we wish to use the Student prior (STUD) rather than the natural conjugate one, and we wish to test the joint hypothesis H_0: $\beta_C = 0$, $\beta_R = 0$, and $\beta_U = 0$, the posterior F-value F_* (as defined in footnote 8) for the NAT2 results is equal to 5.41, which exceeds the .95 quantile of the Fisher distribution with 3 and 30 degrees of freedom: hence H_0 can be rejected at the 95 percent level of posterior confidence if the prior is natural conjugate, and this conclusion can be supposed to hold in the case of the corresponding Student prior. Similarly, a test for first order serial correlation of the error terms can be conducted as follows: regress GW on GPC, GPR, U, a constant, and the Bayesian residuals[11] lagged once, and test that the coefficient of the latter is null using a F-test in this augmented regression.[12] For doing this test, one should use the same prior information as in the original regression. In the case of NAT2, this augmented regression yields $F_* = 0.18$, which is well below the .95 quantile of a $F(1,29)$: first order autocorrelation of the errors can be rejected.[13]

To conclude this discussion, we can make a practical recommendation on the elicitation of the parameters of the natural conjugate prior density: we think it is wise to elicit the prior inverted gamma density of σ^2 in such a way that the prior expected pseudo-R^2 is positive, as this ensures that the prior expectation of σ^2 will at most be equal to the empirical variance of the dependent variable of the regression. At least, this recommendation provides a safeguard against excessively vague prior information on σ^2.

10. Compare (14) with the results in the row DI of Table 1. It is worth noticing that in this example, the prior expected pseudo-R^2 for which the results are close (.40 in the last column for NAT2) is almost equal to the classical R^2.
11. The Bayesian residuals are defined as $y - XE(\beta | y, X)$ when the model is $y = X\beta + u$. There properties are reviewed in BAUWENS and LUBRANO [1991].
12. For a detailed justification of this procedure, see LUBRANO and MARIMOUTOU [1988].
13. All the tests described above can be performed easily with the Bayesian Interactive Program (PC-BIP), available on request to the authors.

APPENDIX 1

In Section 2, we have assumed that the prior information on β_1 is available in the form of a mean vector and a covariance matrix, marginal with respect to σ^2. It is then necessary to assume that the degrees of freedom parameter α of the prior inverted gamma density of σ^2 is larger than two, and that α tends to two rather than to zero when the prior tends to a diffuse limit. In this Appendix, we define a different procedure that does not require that α be larger than two, and therefore allows to let α tend to zero. We show that the results of Propositions 1 and 2 remain true.

We suppose that the marginal prior information on β_1 is available in the form of a highest prior density interval of level π_j for each element β_{1j} of β_1, *i. e.*

$$\text{(15)} \qquad P[b_{1j}^{\text{inf}} \leqslant \beta_{1j} \leqslant b_{1j}^{\text{sup}}] = \pi_j,$$

where b_{1j}^{inf} and b_{1j}^{sup} are given constants. If the probability distribution of β_{1j} is univariate Student with parameters b_{1j} (the mode), m_{1j} (the precision parameter), s (the scale parameter) and α (the degrees of freedom), in short $\beta_{1j} \sim T(b_{1j}, m_{1j}, s, \alpha)$, then it is possible to recover b_{1j} and m_{1j} from the constants b_{1j}^{inf}, b_{1j}^{sup} and π_j by using a table of the standardized univariate Student distribution, *i. e.* a $T(0, 1, 1, \alpha)$. Indeed,

$$b_{1j} = 0.5\,(b_{1j}^{\text{sup}} + b_{1j}^{\text{inf}}),$$

by symmetry of the Student density, and

$$\text{(16)} \qquad m_{1j} = [2\,t/(b_{1j}^{\text{sup}} - b_{1j}^{\text{inf}})]^2\, s/\alpha,$$

where t is such that $P[-t \leqslant Z \leqslant t] = \pi_j$ if $Z \sim T(0, 1, 1, \alpha)$. This value of m_{1j} is deduced from (15) and the transformation needed to standardize a $T(b_{1j}, m_{1j}, s, \alpha)$ random variable, *i. e.* $Z = (\beta_{1j} - b_{1j})(\alpha m_{1j}/s)^{1/2}$, see e. g. ZELLNER [1971, p. 367].

Assuming that b_{1j} and m_{1j} have been obtained for each element of β_1, and that s and α are given (see Section 3), a joint Student distribution for β_1 with mode $b_1 = (b_{1j})$, precision matrix $M_1 = \text{diagonal}\,(m_{1j})$, scale parameter s, and degrees of freedom α, implies that $\beta_{1j} \sim T(b_{1j}, m_{1j}, s, \alpha)$, and therefore is consistent with the available prior information (15). It is important to notice that the diagonality of M_1 implies that the simple correlation coefficient between every pair of elements of β_1 is null (assuming these coefficients exist, *i. e.* $\alpha > 2$), but that the joint Student prior of β_1 is not equal to the product of the Student univariate marginal densities of the elements β_{1j}. In other words, the components of β_{1j} are linearly independent, but are not independent in probability. The construction of the joint Student density will thus imply some distorsion of prior information if it is believed *a priori* that the elements of β_1 are independent in probability.

What happens to m_{1j} as defined by (16) when s and α tend to zero? Excluding the uninteresting case where s/α tends to zero, as the fractile t appearing in (16) tends to infinity as α tends to zero, we conclude that m_{1j} tends to infinity, for every j. Again, if we want (15) to hold within a natural conjugate prior density and to let the prior on σ^2 degenerate to a diffuse limit, we have to inflate the diagonal elements of M_1. Hence we can write as in (4) that $M_1 = eV_1^{-1}$ for some e which tends to infinity and for a fixed diagonal matrix V_1 (which is not necessarily interpretable as a covariance matrix). It is immediate then that the proofs of Propositions 1 and 2 given in Appendix 2 remain valid.

Proof of proposition 1

In (7), let $Q = P^{-1}$ and partition Q as S in (2); given (4), partitioned inverse formulas give

$$(17)\qquad \begin{aligned} Q_1 &= P_1^{-1} + P_1^{-1} S_{12} P_{2.1}^{-1} S_{21} P_1^{-1}, \\ Q_{21} &= -S_2^{-1} S_{21} Q_1 = Q'_{12}, \\ Q_2 &= P_{2.1}^{-1}, \end{aligned}$$

where

$$P_{2.1} = S_2 - S_{21} P_1^{-1} S_{12} \qquad \text{and} \qquad P_1 = M_1 + S_1.$$

M_1 is the upper left block of M, equal to $e V_1^{-1}$, see (2.4).

The computation of $\tilde{b}_1$ gives

$$(18)\qquad \tilde{b}_1 = Q_1 M_1 b_1 + Q_1 (S_1 \hat{b}_1 + S_{12} \hat{b}_2) - Q_1 S_{12} S_2^{-1} (S_{21} \hat{b}_1 + S_2 \hat{b}_2).$$

If $e \to \infty$, since $M_1 = e V_1^{-1}$,

$$(19)\qquad \begin{gathered} P_1 \to \infty V_1^{-1}, \qquad P_1^{-1} \to 0, \qquad P_{21}^{-1} \to S_2^{-1}, \qquad Q_1 \to 0, \\ Q_1 M_1 \to I. \end{gathered}$$

The second and last term of (18) tend to a null vector, and (2.9) follows.

The computation of $\tilde{b}_2$ yields

$$\tilde{b}_2 = -S_2^{-1} S_{21} Q_1 (M_1 b_1 + S_1 \hat{b}_1 + S_{12} \hat{b}_2) + P_{2.1}^{-1} (S_{21} \hat{b}_1 + S_2 \hat{b}_2).$$

If $e \to \infty$, by (19),

$$\tilde{b}_2 \to -S_2^{-1} S_{21} b_1 + S_2^{-1} S_{21} \hat{b}_1 + \hat{b}_2,$$

which is equivalent to (10) given the well-known equality

$$\bar{b}_2 = \hat{b}_2 + S_2^{-1} S_{21} \hat{b}_1,$$

see (2).

Proof of proposition 2

Let us first compute the limit of

$$c = (b - \hat{b})' M P^{-1} S (b - \hat{b}),$$

the third term of the right hand side of $\tilde{s}$, see (7). It does not matter for this computation whether e tends to infinity because s does (α being fixed) or because s tends to zero more slowly than $\alpha-2$. Using (2) to (4) and (17), we get

$$(20)\quad c=(b_1-\hat{b}_1)'(\mathrm{M}_1\mathrm{Q}_1\mathrm{S}_1+\mathrm{M}_1 Q_{12}\mathrm{S}_{21})(b_1-\hat{b}_1) \\ +(b_1-\hat{b}_1)'(\mathrm{M}_1\mathrm{Q}_1\mathrm{S}_{12}+\mathrm{M}_1\mathrm{Q}_{12}\mathrm{S}_2)(b_2-\hat{b}_2).$$

By the last result in (19), the matrix sum in the second term of (20) tends to a null matrix, and

$$(21)\qquad c\rightarrow\bar{c}=(b_1-\hat{b}_1)'\mathrm{S}_{1.2}(b_1-\hat{b}_1).$$

If $e\rightarrow\infty$ when s tends to zero, then as in (19), $\mathrm{Q}_1\rightarrow 0$, so $\tilde{\mathrm{V}}_1\rightarrow 0$ and $\tilde{\mathrm{V}}_{21}\rightarrow 0$ since $\tilde{s}\rightarrow(r+\bar{c})/\mathrm{T}$, a finite limit; also $\mathrm{Q}_2\rightarrow\mathrm{S}_2^{-1}$, and $\tilde{\mathrm{V}}_2$ tends to the limit given in (11).

If $e\rightarrow\infty$ because s does, then $\tilde{s}$ tends to infinity because so does s. It follows directly that $\tilde{\mathrm{V}}_2$ blows up as indicated in (12). For $\tilde{\mathrm{V}}_1$, we have to compute the limit of $s\mathrm{Q}_1$ as $s\rightarrow\infty$. From (17),

$$(22)\qquad s\mathrm{Q}_1=s\mathrm{P}_1^{-1}+s\mathrm{P}_1^{-1}\mathrm{S}_{12}\mathrm{P}_{2.1}^{-1}\mathrm{S}_{21}\mathrm{P}_1^{-1}.$$

As $s\rightarrow\infty$, so does e, hence we can use the results in (19) and conclude that the second term of (22) tends to a null matrix. Since $\mathrm{P}_1=\mathrm{M}_1+\mathrm{S}_1$, the first term is equal to $s[s\mathrm{V}_1^{-1}/(\alpha-2)+\mathrm{S}_1]^{-1}$ and tends to $(\alpha-2)\mathrm{V}_1$ because S_1 is dominated. The limits of $\tilde{\mathrm{V}}_1$ and of $\tilde{\mathrm{V}}_{21}$ given in (12) follow directly.

APPENDIX 3

The data used in the example of Section 3 cover the period 1954-1976.

GW	GPC	GPR	U
0.39212470E-01	−0.75877310E-02	−0.11448390E-01	0.29225664E-01
0.24137500E-01	0.22743820E-01	0.30110470E-01	0.26762140E-01
0.11001940E-01	0.48010350E-01	−0.80171540E-01	0.27026780E-01
0.41454200E-01	−0.79797510E-02	0.15685990	0.42011230E-01
0.47026870E-01	0.29210090E-01	−0.14791190E-01	0.44906610E-01
0.33060550E-01	0.11146070E-02	0.40961060E-01	0.34165220E-01
0.44331550E-02	0.26462200E-01	0.41916230E-01	0.25642450E-01
0.46217200E-01	0.10608320E-01	0.49047480E-01	0.21154360E-01
0.38758750E-01	0.36883350E-01	0.30778090E-01	0.22698640E-01
0.49888850E-01	0.41466240E-01	0.25452150E-01	0.14898960E-01
0.44962530E-01	0.44930100E-01	0.91076540E-01	0.19593290E-01
0.35369520E-01	0.42580600E-01	0.50293440E-01	0.21318120E-01
0.45615320E-01	0.24965410E-01	0.38386310E-01	0.30221040E-01
0.31615140E-01	0.27841690E-01	0.49038480E-01	0.32811320E-01
0.39138080E-01	0.30160670E-01	0.60759160E-01	0.25526490E-01
0.65174220E-01	0.25125740E-01	0.61245710E-01	0.19723460E-01
0.64624910E-01	0.48300030E-01	0.58473870E-01	0.21670120E-01
0.66985250E-01	0.47848940E-01	0.45029590E-01	0.25474930E-01
0.10388710	0.57416800E-01	0.94748830E-01	0.25760460E-01
0.68831680E-01	0.12200300	0.75638920E-01	0.29627450E-01
0.95974680E-01	0.12077580	0.45573840E-01	0.56710310E-01
0.10483380E-01	0.77260730E-01	0.85494500E-01	0.62337970E-01

• References

BAUWENS, L., and LUBRANO, M. (1991). – "Bayesian Diagnostics for Heterogeneity", *Annales d'Economie et de statistiques*, 20/21, pp. 17-40.

BERGER, J. (1985). – *Statistical Decision Theory*, Berlin, Springer Verlag.

DRÈZE, J. H. (1977). – "Bayesian Regression Analysis Using Poly-*t* Densities", *Journal of Econometrics*, 6, pp. 329-354.

JUDGE, G. G., GRIFFITHS, W. E., CARTER HILL, R., LÜTKEPOHL, P. and TSOUNG-CHAO, L. (1985). – *The Theory and Practice of Econometrics*, New York, Wiley.

LEAMER, E. E. (1978). – *Specification Searches*, New York, Wiley.

LEAMER, E. E. (1982). – "Sets of Posterior Means with Bounded Variance Priors", *Econometrica*, 50, pp. 725-736.

LUBRANO, M., and MARIMOUTOU, V. (1988). – "Bayesian Specification Searches", *Cahiers du CERO*, 30, pp. 201-223.

RAIFFA, H., and SCHLAIFER, R. (1961). – *Applied Statistical Decision Theory*, Cambridge: MIT Press.

RICHARD, J.-F. (1973). – *Posterior and Predictive Densities for Simultaneous Equation Models*, Berlin, Springer Verlag.

RICHARD, J.-F., and STEEL, M. (1988). – "Bayesian Analysis of Systems of Seemingly Unrelated Regression Equations under a Recursive Extended Natural Conjugate Prior Density", *Journal of Econometrics*, 38, pp. 7-37.

ZELLNER, A. (1971). – *An Introduction to Bayesian Inference in Econometrics*, New York, Wiley.

ZELLNER, A. (1986). – "On Assessing Prior Distributions and Bayesian Regression Analysis with *g*-prior Distributions", in *Bayesian Inference and Decision Techniques*, P. GOEL and A. ZELLNER Eds., Amsterdam, Elsevier Science Publishers.