

# BVAR MODELS IN THE CONTEXT OF COINTEGRATION: A MONTE CARLO EXPERIMENT

Luis J. Álvarez (\*)  
and Fernando C. Ballabriga (\*\*)

(\*) Banco de España.

(\*\*) ESADE and Banco de España.

We are grateful to Juan, J. Dolado and Enrique Sentana for their comments and suggestions.

Banco de España - Servicio de Estudios  
Documento de Trabajo nº 9405

In publishing this series the Banco de España seeks to disseminate studies of interest that will help acquaint readers better with the Spanish economy.

The analyses, opinions and findings of these papers represent the views of their authors; they are not necessarily those of the Banco de España.

ISBN: 84-7793-288-3

Depósito legal: M-7052-1994

Imprenta del Banco de España

## ABSTRACT

The kind of prior typically employed in Bayesian vector autoregression (BVAR) analysis has aroused widespread suspicion about the ability of these models to capture long-run patterns. This paper specifies a bivariate cointegrated stochastic process and conducts a Monte Carlo experiment to assess the small sample performance of two classical and two Bayesian estimation methods commonly applied to VAR models. In addition, a proposal to introduce a new dimension to the prior information in order to allow for explicit account of long-run restrictions is suggested and evaluated in the light of the experiment. The results of the experiment show that: (i) the Minnesota-type prior with hyperparameter search performs well, suggesting that the prevalent suspicion about the inability of this prior to capture long-run patterns is not well-grounded; (ii) the fine-tuning of the prior is crucial; and (iii) adding long-run restrictions to the prior does not provide improvements in the case analyzed.



## 1. INTRODUCTION

For many years it has been recognized that most economic time series exhibit trending behavior, implying that they are non-stationary. However, economic theory rarely suggests equilibria that are not stationary functions of the variables involved. As a consequence, a widely held view is that there should be fundamental economic forces, such as a market mechanism or government intervention, which make some economic variables move stochastically together. Standard examples include the relationship between real wages and productivity, long and short-run interest rates and nominal exchange rates and relative prices.

This interest in models that capture the belief of close relations in the long-run has led to extensive research in the econometric and statistical literature. Cointegration has been studied, *inter alia*, by Granger (1981), Granger and Weiss (1983), Engle and Granger (1987) and Johansen (1988, 1991a)<sup>1</sup>. Box and Tiao (1977) introduced canonical analysis whereas Peña and Box (1987) proposed dynamic factor analysis. The methods of reduced rank regression were originally applied to the study of non-stationary series by Velu, Reinsel and Wichern (1986) and Ahn and Reinsel (1990).

However, this focus on common non-stationary trends among economic time series has received so far surprisingly little attention by econometricians following the work of Litterman (1980, 1984), Doan, Litterman and Sims (1984) and Sims (1989) despite its interest having been recently pointed out by Sims (1991a).

A reason for this neglect may lie in some recent research which casts doubt on the practical and theoretical usefulness of unit root econometrics. Specifically, Sims (1988) and Sims and Uhlig (1991) point out the discrepancy between classical and Bayesian inference in time series models with possible unit roots, drawing attention to the fact that inference based on the likelihood principle (e.g. Bayesian) is robust to whether or not the data are

---

<sup>1</sup> See Banerjee, Dolado, Galbraith and Hendry (1993) for a thorough treatment of the literature on cointegration.

stationary. They take the position that Bayesian inference is more reasonable and simple to implement<sup>2</sup>. Sims, Stock and Watson (1990) show that even if one takes the classical approach to inference, standard asymptotic theory is valid more often than expected. Christiano and Eichenbaum (1990) point out that it is impossible to discriminate on the basis of finite samples between difference stationary and trend stationary stochastic processes local to each other; Campbell and Perron (1991) emphasize that the same principle applies when comparing the classes of cointegrated and non-cointegrated processes<sup>3</sup>. Finally, Quah (1992) shows that there is no relationship between unit roots and the relative importance of permanent and transitory economic disturbances, which is what matters from the economic theory point of view.

Taken together, one implication of the above literature is that in doing classical econometric analysis in a non-stationary framework, the analyst is using a controversial approach to inference to discriminate among alternatives that cannot be consistently distinguished on the basis of available information; in particular, if the search for cointegration is part of the analysis, the analyst may end up imposing with *certainty* some false long-run restrictions on the behavior of the time series vector. Furthermore, the effort does not provide knowledge about the importance of permanent economic driving forces.

On the other hand, the usual type of prior in BVAR models is characterized by taking coefficients across variables and equations to be independent of each other, so no explicit prior account of long-run relations among variables is considered. This fact, coupled with the random walk type prior mean used in earlier applications, tends to raise the suspicion that the resulting estimation will be biased towards systems of univariate AR models. Clements and Mizon (1991), Lütkepohl (1991) and Phillips (1991b) unfavorable criticism of BVAR models goes in this direction.

---

<sup>2</sup> Phillips (1991a), also with a Bayesian approach, challenges the methods, the assertions and the conclusions of these articles on the Bayesian analysis of unit roots. Nevertheless, see also the comments by eight discussants in the same issue of the *Journal of Applied Econometrics*.

<sup>3</sup> The fact that the determination of the number of cointegrating vectors is not straightforward using only formal testing procedures is often recognized in applied work. See, *inter alia*, Johansen (1992a), Johansen and Juselius (1992), Juselius (1992), and Reinsel and Ahn (1992).

This paper tries to assess the validity of the above mentioned criticism, frequently leveled at BVAR models when it is suspected that there may be close relations in the long-run among the variables considered. The paper also explores the addition of a new dimension to the usual type of prior in order to take into account explicitly the possibility of there being stochastic long-run patterns in the Bayesian prior. To this end, we proceed in two stages. First, we implement the maximum-likelihood procedure proposed by Johansen. Second, we use the estimation results of the first stage to incorporate a set of stochastic linear restrictions into the prior of the coefficients of our model in order to introduce flexibility in the specification and allow the data to depart from the plausible cointegration restrictions.

The remainder of the paper is organized as follows. In section 2 we describe the statistical model. In section 3 we discuss how to combine the cointegration restrictions with other available prior information on the coefficients. In section 4 we compare alternative estimation methods, included the one proposed, through a bivariate Monte Carlo experiment. Section 5 states the conclusions.

## 2. A REVIEW OF BVAR MODELS

We assume that our set of economic time series is a realization of a  $n$ -vector stochastic process,  $Y$ , which for all  $t$  satisfies the model

$$Y(t) = B(L)Y(t) + DX(t) + \varepsilon(t) \quad t = 1, \dots, T \quad (1)$$

where  $B(L) = B_1L + \dots + B_mL^m$  is a  $nxn$  matrix polynomial in the lag operator  $L$  [ $L^k Y(t) \equiv Y(t-k)$ ],  $m$  is the number of lags allowed,  $X(t)$  is a  $dx1$  vector of deterministic variables,  $D$  is a  $nxd$  matrix of coefficients for the deterministic variables and  $\varepsilon(t)$  is a  $nx1$  vector Gaussian white noise process

$$\varepsilon(t) \sim N(0, \Sigma_\varepsilon) \quad (2)$$

The  $i$ -th equation of model (1), abstracting from deterministic components, is

$$\begin{aligned} Y_i(t) = & b_{11}^i Y_1(t-1) + \dots + b_{1m}^i Y_1(t-m) + \\ & + \dots + \\ & + b_{n1}^i Y_n(t-1) + \dots + b_{nm}^i Y_n(t-m) + \\ & + \varepsilon_i(t) \end{aligned} \quad (3)$$

To simplify the presentation we stack up all the coefficients of the  $n$  equations of the VAR in a vector  $b$  with individual elements  $b_{js}^i$  and  $d_l^i$ , where  $i$  refers to the equation,  $j$  to the variable,  $s$  to the lag and  $l$  to the deterministic component ( $i = 1, \dots, n; j = 1, \dots, n; s = 1, \dots, m; l = 1, \dots, d$ ). The first  $m$  coefficients of the vector correspond to the first equation and first variable, the following  $m$  coefficients to the first equation and the second variable, the coefficients from  $nm+1$  to  $nm+d$  to the deterministic variables of the first equation and so on. Therefore  $b$  is a  $n(nm+d)x1$  vector<sup>4</sup>.

With  $n(nm+d)$  parameters the exhaustion of degrees of freedom is an important consideration. To overcome the problems associated with overparametrization, Litterman

---

<sup>4</sup> Note that the vector  $b$  is not indexed by  $t$ . For notational convenience we do not consider time-varying parameters as it is customary in the BVAR literature. See, for example, Sims (1991a) and Canova (1993).



(1980) and Doan, Litterman and Sims (1984) suggested the use of a Bayesian estimation approach in which the sample information is combined with some a priori information on the coefficients to derive a posterior distribution. Their suggestion of using a prior comes from the fact that equations with too many free parameters tend to pick up excess noise, whereas equations with too few parameters fail to pick up the signal. The specification of a prior then provides a flexible format through which one can confront the trade-off between overfitting the data and increasing signal extraction capabilities. The approach can also be thought of as imposing "fuzzy" restrictions on the coefficients of the model rather than employing exclusion restrictions.

To this end, vector  $b$  is given a multivariate normal prior density function with mean  $\bar{b}$  and variance-covariance matrix  $\Sigma_b$

$$b \sim N(\bar{b}, \Sigma_b) \quad (4)$$

This prior neither attempts to reflect personal knowledge nor be based on economic theory. Instead, it intends to capture empirical features that can be widely accepted by many researchers. Furthermore, to reflect the uncertainty about these features, the prior distribution is made to depend upon a low-dimensional vector  $\tau$  of hyperparameters.

It has to be recognized, however, that from a strict Bayesian standpoint our prior information should not contain unknown parameters (the elements of  $\tau$ ); we should also attach a prior distribution to them. A full Bayesian implementation would require to specify that distribution and then go through the appropriate integration process to obtain the posterior distribution. However, two shortcuts to this computationally demanding procedure are usual practice in the BVAR literature. The first is based on what we will refer to as "standard prior" and consists in using the posterior associated with a specific setting of  $\tau$  reflecting some empirical rules of thumb concerning time series behavior. The approach actually amounts to assuming that  $\tau$  is a degenerate random vector with probability weight one on the specific choice. The second, suggested by Doan, Litterman and Sims (1984) and based on what we will call "optimal prior", is to employ the posterior associated with the hyperparameter setting that maximizes the conditional sample pdf. As explained by these authors, under the assumption that the density of the hyperparameter vector is flat in some

hypercube, the suggested approximation is good if for the set of hyperparameter vectors for which the conditional sample pdf is large, the associated posterior does not change much.

The basic idea of this second approach is to specify a relatively unrestricted vector autoregression and a prior that can be varied along several dimensions (the elements of  $\tau$ ) affecting the trade-off between overparametrization and oversimplification. The consideration of the likelihood as a function of movement in these dimensions is then used to find the optimal balance. In effect, these hyperparameters are used to fine-tune the prior, which then acts as a filter to extract as much information from the data as possible.

Turning to the specifics of the prior, it has been a common practice to incorporate into it the fact that most economic time series can be roughly approximated by a simple random walk model with drift. This has led many researchers, especially in earlier applications [see, e.g., Litterman (1986)], to consider the prior mean for all the coefficients as zero, except for the first lag of the dependent variable which is taken to be one; i.e.

$$\bar{b}_{js}^i = \begin{cases} 1 & i=j, s=1 \\ 0 & \textit{otherwise} \end{cases} \quad (5)$$

or, more generally, an unknown component of the hyperparameter vector  $\tau$  [see, e.g., Ballabriga (1988)]

$$\bar{b}_{js}^i = \begin{cases} \tau_0 & i=j, s=1 \\ 0 & \textit{otherwise} \end{cases} \quad (6)$$

Quite recently, Sims (1989) has recognised that many BVAR models include variables determined in auction markets, such as exchange rates, that should be the object of a special treatment. Instead of using a discrete time random walk prior, Sims has proposed the discrete specification resulting from the time averaging of a random walk process operating on a shorter interval. If the true random walk is defined over an hourly interval but the estimation is carried out with monthly averages, then we get an ARIMA (0, 1, 1) with moving average parameter  $\Theta = 2 - \sqrt{3}$ . This process can be represented as an infinite autoregressive process

which can be truncated at the longest lag allowed to obtain a new prior mean<sup>5</sup>.

The initial prior usually makes the individual components of the vector  $b$  independent of each other<sup>6</sup>; i.e. it makes the covariance matrix  $\Sigma_b$  diagonal and sets the diagonal elements according to

$$\sigma_{ij}^2(s) = \frac{\tau_1}{s^{\tau_3}} \sigma_i^2 \quad i = j \quad \begin{matrix} i = 1, \dots, n \\ s = 1, \dots, m \end{matrix} \quad (7)$$

$$\sigma_{ij}^2(s) = \frac{\tau_1 \tau_2}{s^{\tau_3}} \frac{\sigma_i^2}{\sigma_j^2} \quad i \neq j \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, n \\ s = 1, \dots, m \end{matrix} \quad (8)$$

$$\sigma_{ij}^2(s) = \tau_1 \tau_4 \sigma_i^2 \quad \begin{matrix} i = 1, \dots, n \\ l = 1, \dots, d \end{matrix} \quad (9)$$

where  $i$  refers to the equation,  $j$  to the explanatory variable and  $s$  to the lag. Here,  $\sigma_i^2$  and  $\sigma_j^2$  are parameters measuring the scale of fluctuations in variables  $i$  and  $j$ , usually taken in practice as the squared residual standard error from a univariate AR( $m$ ) model.  $\tau_1$  represents the prior overall tightness.  $\tau_2$  controls the tightness of own lags relative to the tightness of lags of the other variables in the equation. Thus, if it is below one, it allows the imposition of a tighter prior on lags of the other variables in the equation. In the limiting case, if  $\tau_2$  equals zero, we are considering univariate AR( $m$ ) processes. It should also be noted that the ratio of variances is present in (8) to take into account the units of measurement of the data. The hyperparameter  $\tau_3$  controls for the lag-decay in the prior variance. If it is greater than or equal to one it tightens the standard error as the lag increases, thus making coefficients on more distant lags less likely to be large. Finally,  $\tau_4$  controls the deterministic variables

---

<sup>5</sup> If nonseasonally adjusted macroeconomic series are available then the extensions of Canova (1988, 1992) and Raynauld and Simonato (1993) should be considered.

<sup>6</sup> Sometimes the prior is modified to allow for some kind of dependence among own lag coefficients. See, e.g., Sims (1989), Canova (1992) and Ballabriga (1988). Sims (1991a) also suggests a modification to allow for dependence across variables as a way to "leave more room for long term cross-variable relationships". We are not aware, however, of any application allowing for prior cross-equation restrictions.

tightness. In practice, however, more than one hyperparameter is employed to control the deterministic variables. For example, the constant term may have one associated hyperparameter and a linear trend a different one.

Having specified the model, prior and sample information are to be combined according to Bayes rule to obtain a posterior distribution. The Kalman filter provides a convenient device for this purpose. As the prior depends on the particular setting of the hyperparameter vector  $\tau$  so will the posterior. As has been argued above, the analyst may search for the hyperparameter vector that maximizes the likelihood. For estimation purposes two things should be noted. First, whereas ordinary least squares is fully asymptotically efficient for an unrestricted VAR because the same explanatory variables appear on the right hand side of each equation, this is no longer the case with the type of prior information being considered. Therefore, there are gains from treating the equations of the system jointly, and a multivariate version of the Kalman filter should be employed. Second, the hyperparameter searches for likelihood maximization are usually handled with a nonstandard hillclimbing routine, BAYESMTH described in Sims (1986). The program fits a surface to the observed likelihood values to generate a guess for the location of the function's peak. This hyperparameter vector is then used to compute its likelihood through the Kalman filter. The algorithm updates the likelihood surface and then proceeds iteratively until convergence is obtained. A convenient feature of this routine is that it not only solves the maximization problem but also provides an approximation to the likelihood's shape, which is useful to assess the quality of the posterior approximation.

### 3. INCORPORATING COINTEGRATION RESTRICTIONS INTO THE PRIOR

In the previous section we have described a prior which placed no emphasis on the possibility of there existing stable long-run relationships among the levels of the non-stationary variables considered. In this section we propose a way to modify that prior in order to allow for explicit prior account of potential long-run restrictions.

We start by re-arranging model (1) to yield the representation

$$\Delta Y(t) = \Gamma_1 \Delta Y(t-1) + \dots + \Gamma_{m-1} \Delta Y(t-m+1) - \Pi Y(t-m) + \varepsilon(t) \quad (10)$$

where  $\Delta$  is the difference operator [ $\Delta Y(t) \equiv Y(t) - Y(t-1)$ ] and the deterministic components  $X(t)$  have been omitted for presentation purposes. We also have that

$$\Gamma_s = - \left( I - \sum_{i=1}^s B_i \right) \quad s = 1, \dots, m-1 \quad (11)$$

and

$$\Pi = I - \sum_{i=1}^m B_i = I - B(1) \quad (12)$$

Model (10) is expressed as a traditional first difference VAR model except for the term  $\Pi Y(t-m)$  which may contain information about the long-run relationships among the variables in  $Y(t)$ . If we assume that  $Y(t)$  is integrated of order one,  $I(1)$  for short, then  $\Delta Y(t)$  is  $I(0)$ , and for the system to be balanced (in the sense that the right hand side variables be of the same order of integration as the left hand side variables) it has to be that either  $\Pi = 0$ , in which case there is no long-run relationship between the variables and we are left with a traditional first difference VAR model, or  $\Pi Y(t-m)$  is a stationary variable. The latter case applies when the variables in  $Y(t)$  are cointegrated, and in turn implies [see Johansen (1988)] that the rank  $r$  of the  $\Pi$  matrix is less than the number of variables in  $Y(t)$ . Specifically, this rank, also referred to as the order of cointegration, is equal to the number of distinct cointegration vectors linking the variables in  $Y(t)$ . The hypothesis of cointegration is thus the hypothesis of reduced-rank of the long-run impact matrix  $\Pi$ .

As it is clear from (12), the entries of  $\Pi$  are linear combinations of the coefficients of the VAR. This suggests the following two-stage estimation procedure for our model when it is suspected that long-run relationships may characterize the stochastic process under analysis:

- (i) Estimate  $\Pi$  using standard classical cointegration techniques.
- (ii) Use the estimated  $\Pi$  to define a set of linear stochastic restrictions on the coefficients of the VAR and combine them with (4) to obtain a modified prior.

As we have already mentioned in the introduction, the idea of the procedure is to allow explicitly for stable long-run relationship among the levels of certain economic variables while taking into account the uncertainty that surrounds the specification of the model.

Among the several methods existing in the literature to estimate cointegrating relations, we opt for Johansen (1988, 1991a) procedure to implement stage (i). There are several reasons for this choice. First, Phillips (1991c) has theoretically shown that the best way to proceed in the estimation of cointegrated systems is full system estimation by maximum likelihood incorporating all knowledge about the presence of unit roots. Johansen's is a FIML approach to cointegration that proceeds along these lines. The procedure is based on a full specification of the vector time series and gives a joint description of both the short-run and the long-run dynamics of the system. Second, when more than two series are being considered more than one stable linear relationship can exist. Therefore, it is important, as does the Johansen's procedure, to relax the assumption that the cointegration vector is unique. Finally, it appears that, in general, the procedure performs better in finite samples than other procedures, offering good results even when the errors are non-Gaussian or when the dynamics are unknown [see Gonzalo (1991)].

As for the implementation of stage (ii), notice first that the long-run impact matrix has the form

$$\Pi = \begin{pmatrix} 1 - \sum_{s=1}^m b_{1s}^1 & - \sum_{s=1}^m b_{2s}^1 & \dots & - \sum_{s=1}^m b_{ns}^1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ - \sum_{s=1}^m b_{1s}^n & - \sum_{s=1}^m b_{2s}^n & \dots & 1 - \sum_{s=1}^m b_{ns}^n \end{pmatrix} \quad (13)$$

So if we employ Johansen's estimate of  $\Pi$  ( $\Pi^J$  in what follows) under the hypothesis that there are  $r$  cointegration vectors, we have a rank  $r$  matrix defining  $nxn$  linear restrictions on the coefficients of the VAR which can be written in the form

$$Qb = q \quad (14)$$

with  $Q$  a  $n^2 \times k$  matrix,  $k = (mn+d)n$ , and  $q$  a  $n^2 \times 1$  vector. The form of the  $Q$  matrix is

$$Q = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & A_n \end{pmatrix} \quad (15)$$

where  $A_1 = A_2 = \dots = A_n = A$  is a  $n(nm+d)$  matrix given by

$$A = \begin{pmatrix} -1 & -1 & \dots & -1_{1m} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0_d \\ 0 & 0 & \dots & 0_{2m} & -1 & -1 & \dots & -1_{2m2} & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0_d \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0_{nm} & 0 & 0 & \dots & 0_{nm2} & \dots & -1 & -1 & \dots & -1_{nmn} & 0 & \dots & 0_d \end{pmatrix} \quad (16)$$

As for  $q$  it takes the form

$$q = \left( \Pi'_{11}-1 \ \Pi'_{12} \ \dots \ \Pi'_{1n} \ \Pi'_{21} \ \Pi'_{22}-1 \ \dots \ \Pi'_{2n} \ \dots \ \Pi'_{n1} \ \dots \ \Pi'_{nn}-1 \right)' \quad (17)$$

$\Pi'_{ij}$  being the  $ij$ -th element of  $\Pi'$ . Vector  $b$  is as defined in section 2; i.e.

$$b = \left( b^1_{11} \ b^1_{12} \dots b^1_{1m} \ b^1_{21} \dots b^1_{2m} \ \dots \ d^1_1 \dots d^1_d \dots b^n_{11} \dots b^n_{1m} \dots d^n_1 \dots d^n_d \right)' \quad (18)$$

To introduce flexibility in the specification and allow the data to depart from the rank restriction imposed by Johansen's procedure we take (14) to be true on average. Specifically, we make the restriction stochastic by adding a  $n^2 \times 1$  random vector  $v$

$$q = Qb + v \quad (19)$$

and we model the uncertainty on the fulfillment of the restriction as a mean zero Gaussian stochastic process

$$v \sim N(0, \Sigma_v) \quad (20)$$

where  $\Sigma_v$  is taken to be proportional to a diagonal matrix with elements equal to the elements of the main diagonal of the covariance matrix of  $\Pi'$ . The proportionality factor is  $\tau_5$ , a component of the hyperparameter vector  $\tau$ .

We then have two sets of prior information concerning the distribution of  $b$ . On the one hand, the information given by (19) - (20) which intends to capture the belief that close relations among the levels of certain variables may be present. On the other hand, the information in (4), the type of prior used in the BVAR literature that we can rewrite as

$$\begin{aligned} b &= \bar{b} + e \\ e &\sim N(0, \Sigma_b) \end{aligned} \quad (21)$$



To combine (19), (20) and (21) we use Theil's mixed estimation technique [see Theil and Goldberger (1961) and Theil (1963)] and take the resulting estimator and its variance as the mean and variance of the new normal prior. Specifically, we write

$$\begin{pmatrix} \bar{b} \\ k \times l \\ q \\ n^2 \times l \end{pmatrix} = \begin{pmatrix} I \\ Q \\ n^2 \times k \end{pmatrix} b + \begin{pmatrix} -e \\ k \times l \\ v \\ n^2 \times l \end{pmatrix} \quad (22)$$

with

$$\text{var} \begin{pmatrix} -e \\ v \end{pmatrix} = \begin{pmatrix} \Sigma_b & 0 \\ k \times k & k \times n^2 \\ 0 & \Sigma_v \\ n^2 \times k & n^2 \times n^2 \end{pmatrix} \equiv \begin{matrix} \Sigma \\ (n^2 + k) \times (n^2 + k) \end{matrix} \quad (23)$$

Theil's mixed estimator for  $b$  is then

$$b^* = \left( (I Q') \Sigma^{-1} (I Q')' \right)^{-1} (I Q') \Sigma^{-1} \begin{pmatrix} \bar{b} \\ q \end{pmatrix} \quad (24)$$

and its variance

$$\begin{aligned} \text{var}(b^*) &= \left( (I Q') \Sigma^{-1} (I Q')' \right)^{-1} \\ &= \left( \Sigma_b^{-1} + Q' \Sigma_v^{-1} Q \right)^{-1} \end{aligned} \quad (25)$$

For an easier interpretation, rewrite (24) as (see Appendix 1)

$$b^* = \bar{b} + \Sigma_b Q' \left( Q \Sigma_b Q' + \Sigma_v \right)^{-1} (q - Q \bar{b}) \quad (26)$$

which clearly shows that the degree of modification of the prior mean  $\bar{b}$  will depend upon its degree of compatibility with the information incorporated into (19)-(20). On the other hand, as can be seen from (25), the prior variance modification will be a function of the degree of uncertainty associated with the cointegration restrictions; a dimension of the prior controlled through the hyperparameter  $\tau_5$ . In one extreme, if there is a very high degree of uncertainty concerning the fulfillment of the restrictions ( $\tau_5 \rightarrow \infty$ ), the prior piece of information (21) is not modified. In the other extreme ( $\tau_5 = 0$ ), it is fully modified.

#### 4. A BIVARIATE MONTE CARLO EXPERIMENT

Several methods for estimating long-run equilibrium relationships have been proposed in the literature. Sims, Stock and Watson (1990) show that consistent estimates can be obtained from unrestricted VAR models specified in levels, so they suggest employing the OLS estimation procedure. Further, Park and Phillips (1989) and Ahn and Reinsel (1990) show that the OLS estimator has the same asymptotic properties as the maximum likelihood estimator which observes the cointegration restrictions. However, from a classical standpoint, the most widely used procedure in the estimation of cointegrated systems is the maximum-likelihood method proposed by Johansen (1988, 1991a).

On the other hand, the two methods outlined in section 2 coexist in the BVAR literature: the method based on the "standard prior" (where the elements of  $\tau$  are set equal to empirically successful values) and the one based on the "optimal prior" (where  $\tau$  is chosen so as to maximize the sample likelihood). The latter is more likely to provide a better approximation to the posterior, which suggests that it should deliver more efficient results than the former. However, the search for the optimal  $\tau$  may get to be a very time-consuming process, a fact that might sometimes explain the sticking to the "standard prior". Nevertheless, as we have already argued, neither the "optimal prior" nor the "standard prior" take explicitly into account the possibility of there being long-run restrictions among the time series analyzed. This has provided a partial focus for the macroeconomic debate in the context of unit roots: to some macroeconometricians such a characteristic of the prior just means that the analysis based on it will almost certainly be wrong [see the criticisms of Clement and Mizon (1991), Lütkepohl (1991) and Phillips (1991b)]. To others this characteristic is not determinant because the superconvergence property of the unit roots and cointegration aspects of the data means that these aspects of the model estimates are quite insensitive to the prior [see Sims (1991c)]. The previous section tries to bridge positions by suggesting the addition of a dimension to the usual type of prior aimed at taking explicitly into account possible long-run equilibrium relationships; we will refer to this modified prior as the "cointegrated prior".

The goal of this section is to compare the small sample performance of the aforementioned five methods (see Chart 1) through a bivariate Monte Carlo experiment, focussing our attention on BVAR models.

To this end, the following bivariate data generating process (DGP), of the type used by Banerjee, Dolado, Hendry and Smith (1986) and Engle and Granger (1987), is considered:

$$\begin{aligned}\Delta Y_1(t) &= \alpha \Delta Y_2(t) + \beta [Y_1(t-1) - Y_2(t-1)] + \eta_1(t) \\ \Delta Y_2(t) &= \gamma [Y_1(t-1) - Y_2(t-1)] + \eta_2(t)\end{aligned}\quad (27)$$

$$\begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix} \sim iid N(0, \Sigma_\eta)$$

Both series are  $I(1)$ , but there is a linear combination of  $Y_1(t)$  and  $Y_2(t)$  given by the cointegrating vector  $\delta = (1 \ -1)'$  that is  $I(0)$ .

The DGP can also be written in the form of model (10) (as a vector error correction model)

$$\begin{pmatrix} \Delta Y_1(t) \\ \Delta Y_2(t) \end{pmatrix} = \begin{pmatrix} \beta + \alpha\gamma & -(\beta + \alpha\gamma) \\ \gamma & \gamma \end{pmatrix} \begin{pmatrix} Y_1(t-1) \\ Y_2(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}\quad (28)$$

where

$$\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix}\quad (29)$$

And in the vector autoregressive form

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \begin{pmatrix} 1 + \beta + \alpha\gamma & -(\beta + \alpha\gamma) \\ \gamma & 1 - \gamma \end{pmatrix} \begin{pmatrix} Y_1(t-1) \\ Y_2(t-1) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}\quad (30)$$

It should be noted that in this bivariate DGP no variable is weakly exogeneous with respect to the cointegrating relation. Therefore, there would be a loss in efficiency if we used single-equation analysis [see Johansen (1992b) and Dolado (1992)].

In the simulations we have arbitrarily set  $\alpha = 0.5$ ,  $\beta = -0.8$  and  $\gamma = 0.8$ , so the VAR representation of our DGP is given by

$$\begin{aligned} Y_1(t) &= 0.6 Y_1(t-1) + 0.4 Y_2(t-1) + \varepsilon_1(t) \\ Y_2(t) &= 0.8 Y_1(t-1) + 0.2 Y_2(t-1) + \varepsilon_2(t) \end{aligned} \quad (31)$$

The specification is completed with a white noise Gaussian assumption for  $\varepsilon(t)$  of the form

$$\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix} \sim iid \mathbf{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.004 & 0.0004 \\ 0.0004 & 0.004 \end{pmatrix} \right] \quad (32)$$

With this DGP we have generated 100 series of length 80 and have discarded the initial 50 values of each. Our results are therefore based on the analysis of 100 small-sized (30 observations) samples. As we have already mentioned, fine-tuning the priors is a costly process. This has led us to limit to 100 the number of experiments.

Before proceeding to the discussion of the results, several comments are in order:

- (i) For all the methods considered we have fitted a VAR(2) with intercept term.
- (ii) When we display results on Johansen's method we have always imposed the (valid) restriction that the cointegration rank is one. As the  $\lambda$ -maximum and trace statistics performed rather well, we do not believe that this may have (strongly) favoured the performance of this procedure<sup>7</sup>.

---

<sup>7</sup> Our major concern was not the behaviour of Johansen's method. Interested readers in the trace and  $\lambda$ -maximum tests as well as short and long-run parameters in cointegrated systems should see Eitrheim

- (iii) The hyperparameter vector which corresponds to our "standard prior" appears in table 1.
- (iv) The fine-tuning to obtain the "optimal " and "cointegrated" priors has been done with the nonstandard hillclimbing routine described in Sims (1986). We have used an average of 50 function evaluations to obtain convergence<sup>8</sup>. In table 1 we present, as descriptive statistics, the mean and standard deviation of the different hyperparameters across the 100 samples.
- (v) We have also examined the behaviour of a "cointegrated prior" in which we used the theoretical  $\Pi$  matrix instead of the ones estimated by Johansen's procedure. We did achieve some gains in terms of the likelihood. However, as the posterior was the same up to three decimal places as when we estimated the  $\Pi$  matrix, we omit the presentation of these results.
- (vi) All the Bayesian procedures have been implemented using the multivariate version of the Kalman Filter.

To gauge estimator performance (see the discussion in Appendix 2) we consider the root mean square error and the probability of concentration<sup>9</sup> defined as  $Pr(|\hat{\beta} - \beta| \leq 0.1)$ . Tables 2 to 5 present results on the precision of the different estimators. We observe the following:

- (i) From a classical standpoint, examination of the results referring to unconstrained OLS and Johansen does not present a clear winner<sup>10</sup>.

---

(1992), Gonzalo (1991), Johansen (1991b) or Reimers (1992).

<sup>8</sup> Sims (1989) indicates that this number allows to "obtain a very rough convergence". In empirical work we have often found 50 iterations to be insufficient. Therefore, we believe that this fact may have disfavoured (slightly) these procedures.

<sup>9</sup> To check the robustness of this measure we have also considered as cut-off value 0.15. The relative performance among different methods did not change.

<sup>10</sup> However, if we observe tables 6 to 9 in Appendix 2, the OLS estimator of own lag coefficients has a substantial downward bias in mean and median. Furthermore, it should also be mentioned that the long-run elasticity (not presented) is far better estimated using Johansen's procedure (which imposes

- (ii) The performance of the "standard prior" is very poor. The posterior mean (see tables 6-7 in Appendix 2) is not far away from the prior mean, so we end up with (roughly) a bivariate random walk process. Although this prior is not employed by experienced researchers, its posterior agrees quite closely with the criticisms directed towards BVAR models [see Clements and Mizon (1991), Lütkepohl (1991) and Phillips (1991c)]. These results point out the dangers of the mechanical implementation of VAR priors in which "too many" unit roots are imposed.
  
- (iii) The performance of the "optimal prior" is satisfactory. If we compare its mean hyperparameter vector with the one used in the "standard prior" (table 1), we see that they are very distant. On observing the mean square error and probability of concentration associated with their respective posterior we find that the expensive hyperparameter search is worthwhile. Comparing the statistics obtained with the "optimal prior" to those of Johansen's procedure we observe that there is no clear winner<sup>11</sup> with respect to the coefficients of the first lag, and that the coefficients of the second lag, which are zero in the DGP, are estimated much more precisely with the "optimal prior" approach<sup>12</sup>. We claim, therefore, that the assertion that BVAR models are severely misspecified for cointegrated processes has no basis. According to our results, this widespread belief appears to be right only in the "standard prior" approach; therefore, the fine-tuning of the prior is strongly recommended.
  
- (iv) The performance of the "cointegrated prior" is also satisfactory. As can be seen in Tables 2 to 5, there are practically no differences with the "optimal prior", which is a reflection of the fact that the hyperparameter value

---

the number of unit roots in the system) than OLS (which estimates them).

<sup>11</sup> We should probably emphasize that our comparison is based on measures of the "closeness" of estimates to the true values. If big weight is attached to unbiasedness, Johansen's estimates should be preferred.

<sup>12</sup> The long-run elasticity (not presented) is also slightly better estimated with the "optimal prior".

associated with the cointegration restriction is very high (its mean is 21.58) and therefore it only modifies slightly the prior. We think that this similar performance is explained by Sims (1991c) view that the sample is highly informative with respect to cointegrating vectors, which are estimated more sharply and converge more quickly with sample size than other aspects of the model<sup>13</sup>. This should be specially true in the context of the specified DGP, characterized by a matrix of "large" adjustment coefficients and therefore expected to generate series which are not too frequently away from the equilibrium relationship.

---

<sup>13</sup> The same reason explains the close results obtained when employing the theoretical or estimated  $\Pi$ .



## 5. CONCLUSIONS

BVAR models have the strongly desirable characteristic of allowing the analyst to treat systematically during the specification process the uncertainty about the true nature of the data generating mechanism. However, the type of prior information generally used to deal with that uncertainty has led to the unfavorable criticism that the approach may be unable to capture potential long-run restrictions among the levels of the variables analyzed.

This paper conducts a bivariate Monte Carlo experiment to assess the validity of the above criticism and to explore a modification to the usual type of prior to explicitly incorporate long-run restrictions. The experiment provides two remarkable outcomes.

First, the usual type of prior performs well if a hyperparameter search is conducted (optimal prior), but provides misleading results if the prior is set according to empirical rules of thumb (standard prior). We interpret this result as a piece of evidence suggesting that the widespread suspicion about the inability of this type of prior to capture long-run patterns is not well-grounded, but also implying that fine-tuning the prior is very important to get a good approximation to the posterior. The result, therefore, provides only partial support to Sims (1991c) view that in practice the choice of the aspects of the prior concerning unit roots and cointegrating vectors is not very important because these aspects of the model converge quickly to their true values.

Second, the good performance of the "optimal prior" implies in turn that adding long-run restrictions to the prior does not provide improvements in the case analyzed in this paper. Our intuition, however, is that in set-ups where the series are more frequently away from the equilibrium relationship, the combination proposed should allow to come out with a higher quality reduced form model.

Certainly, our conclusions are tentative. Gaining further support for them and confirming the above intuition requires further exploration of the DGP parametric space.



## APPENDIX 1

### ALTERNATIVE EXPRESSION FOR THE MODIFIED PRIOR MEAN

Using Theil's mixed estimation technique we have seen that the modified prior mean is

$$b^* = [(I Q') \Sigma^{-1} (I Q')']^{-1} (I Q')^{-1} \Sigma^{-1} \begin{pmatrix} \bar{b} \\ q \end{pmatrix} \quad (A.1)$$

To obtain the expression (26) in the text we use the following identities:

$$(A + BDB')^{-1} \equiv A^{-1} - A^{-1} BEB'A^{-1} + A^{-1} BE(E+D)^{-1} EBA'^{-1} \quad (A.2)$$

$$(A + B)^{-1} \equiv A^{-1} (A^{-1} + B^{-1})^{-1} B^{-1} \quad (A.3)$$

$$B(B'A^{-1}B + D^{-1})^{-1} B' \equiv BEB^1 + BE(E + D)^{-1} EB \quad (A.4)$$

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

Inverting the block diagonal matrix in (A.1) and rearranging gives

$$b^* = (\Sigma_b^{-1} + Q' \Sigma_v^{-1} Q)^{-1} (\Sigma_b^{-1} \bar{b} + Q' \Sigma_v^{-1} q) \quad (A.5)$$

Considering then (A.2)

$$\begin{aligned} b^* &= \bar{b} + \Sigma_b Q' (Q \Sigma_b Q')^{-1} Q \bar{b} + \\ &+ \Sigma_b Q' (Q \Sigma_b Q')^{-1} (Q \Sigma_b Q + \Sigma_v^{-1})^{-1} (Q \Sigma_b Q')^{-1} Q' \bar{b} + \\ &+ \Sigma_b Q' \Sigma_v^{-1} q - \Sigma_b Q' (Q \Sigma_b Q')^{-1} Q \Sigma_b Q' \Sigma_v^{-1} q + \\ &+ \Sigma_b Q' (Q \Sigma_b Q')^{-1} (Q' \Sigma_b Q + \Sigma_v^{-1})^{-1} (Q \Sigma_b Q')^{-1} Q' \Sigma_b Q' \Sigma_v^{-1} q \end{aligned} \quad (A.6)$$

And using (A.3) and (A.4) we have

$$b^* = \bar{b} + \Sigma_b Q' (Q \Sigma_b Q' + \Sigma_v)^{-1} (q - Q \bar{b}) \quad (A.7)$$

which is the expression sought.

## APPENDIX 2

### OTHER MEASURES OF PERFORMANCE

In section 4 we have presented results on the root mean square error and the probability of concentration around the true values of the coefficients of the estimated models. This appendix includes information on two measures of location (mean and median) as well as two measures of dispersion (standard deviation and interquartile range).

We do believe that unbiased estimators are highly desirable when there exists the possibility of carrying out controlled experiments: as more samples are collected, the average value of sample estimates tends toward the value of unknown parameters. However, economists are seldom able to perform controlled experiments. Thus, we prefer estimators that miss the mark on average (i.e. they are biased) if this brings gains in terms of reduced variance. This explains our choice of measures of "closeness" to gauge estimator performance. Nevertheless, for those interested, performance in terms of mean values is presented in Tables 6 and 7.

Measures in terms of medians and interquartile ranges are useful because they are quite robust to the existence of outliers. We present them in Tables 8-9 and 10-11. Finally, Tables 12 and 13 contain the results in terms of standard deviations.

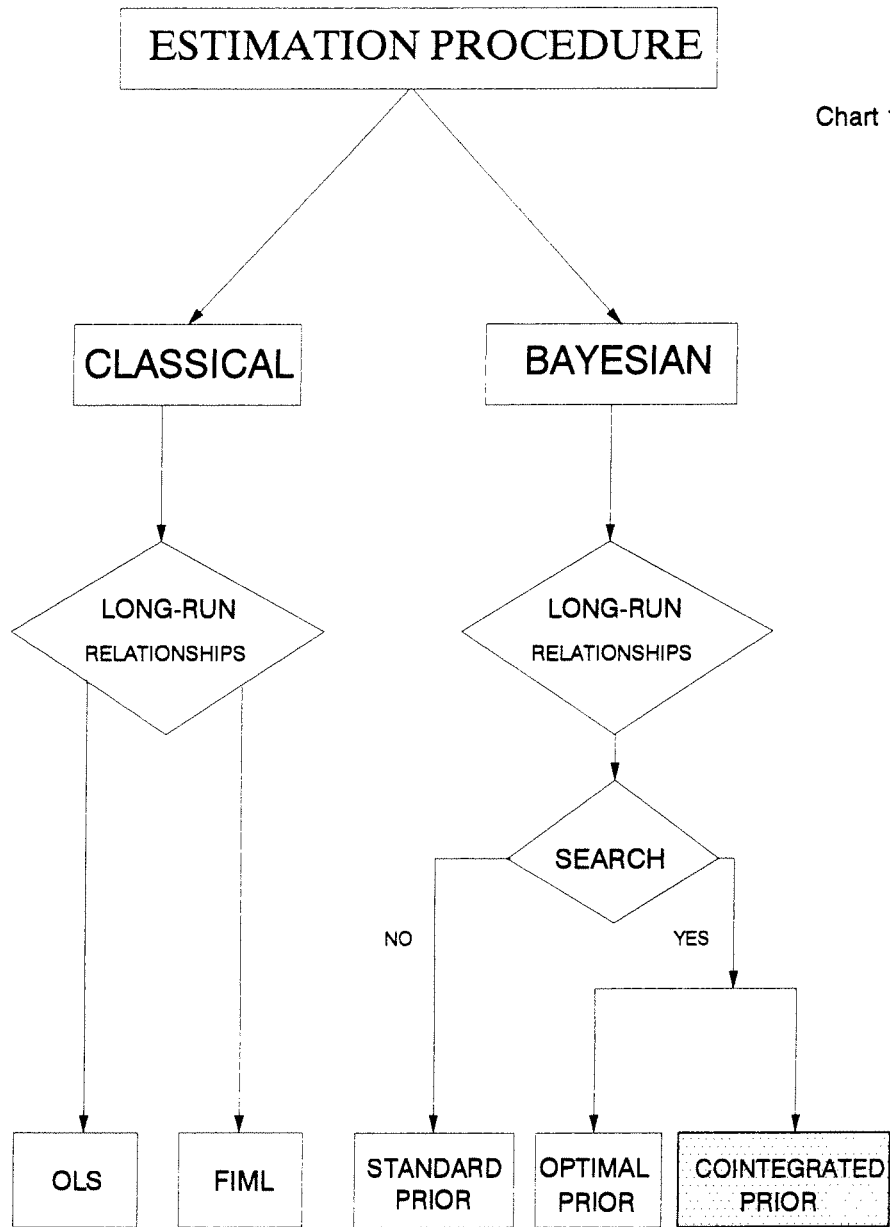


Chart 1

**HYPERPARAMETER VALUES**

**Table 1**

|          | <u>Standard prior</u> | <u>Optimal prior</u> | <u>Cointegrated prior</u> |
|----------|-----------------------|----------------------|---------------------------|
| $\tau_0$ | 1.0000                | 0.4672<br>(0.1184)   | 0.4611<br>(0.1212)        |
| $\tau_1$ | 0.2000                | 0.3050<br>(0.2034)   | 0.2876<br>(0.1380)        |
| $\tau_2$ | 0.5000                | 0.8921<br>(0.3414)   | 1.0359<br>(0.3158)        |
| $\tau_3$ | 1.0000                | 2.3160<br>(1.1224)   | 2.8989<br>(0.8572)        |
| $\tau_4$ | 1.0000                | 8.9731<br>(10.7429)  | 10.1494<br>(5.7320)       |
| $\tau_5$ | -----                 | -----                | 21.5849<br>(6.6580)       |

Note: Mean values across 100 samples. For optimal and cointegrated priors, standard deviations in parenthesis.

**ESTIMATED COEFFICIENTS: ROOT MEAN SQUARE ERROR**  
**DEPENDENT VARIABLE:  $Y_1$**

**Table 2**

|                    | $b_{11}^1$ | $b_{12}^1$ | $b_{21}^1$ | $b_{22}^1$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.2186     | 0.2399     | 0.1840     | 0.1582     | 0.1416   |
| Johansen           | 0.1885     | 0.2348     | 0.1977     | 0.1585     | 0.0461   |
| Standard prior     | 0.3963     | 0.0007     | 0.2460     | 0.1788     | 0.0000   |
| Optimal prior      | 0.1774     | 0.0003     | 0.1121     | 0.1052     | 0.0000   |
| Cointegrated prior | 0.1838     | 0.0002     | 0.1145     | 0.0947     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: ROOT MEAN SQUARE ERROR**  
**DEPENDENT VARIABLE:  $Y_2$**

**Table 3**

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.2349     | 0.2289     | 0.2053     | 0.2044     | 0.0985   |
| Johansen           | 0.2363     | 0.2360     | 0.2006     | 0.2032     | 0.0643   |
| Standard prior     | 0.3094     | 0.5061     | 0.7929     | 0.0010     | 0.0000   |
| Optimal prior      | 0.2506     | 0.1601     | 0.2913     | 0.0004     | 0.0000   |
| Cointegrated prior | 0.2421     | 0.1488     | 0.2868     | 0.0002     | 0.0000   |

Sample size: 30 observations



**ESTIMATED COEFFICIENTS: PROBABILITY OF CONCENTRATION**  
**DEPENDENT VARIABLE:  $Y_1$**

**Table 4**

|                    | $b_{11}^i$ | $b_{12}^i$ | $b_{21}^i$ | $b_{22}^i$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.3700     | 0.3100     | 0.4400     | 0.6000     | 0.5500   |
| Johansen           | 0.4000     | 0.3700     | 0.4000     | 0.4900     | 0.9500   |
| Standard prior     | 0.0000     | 1.0000     | 0.0200     | 0.1900     | 1.0000   |
| Optimal prior      | 0.3200     | 1.0000     | 0.6800     | 0.6800     | 1.0000   |
| Cointegrated prior | 0.2800     | 1.0000     | 0.6400     | 0.7400     | 1.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: PROBABILITY OF CONCENTRATION**  
**DEPENDENT VARIABLE:  $Y_2$**

**Table 5**

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.2600     | 0.3400     | 0.4700     | 0.3800     | 0.7200   |
| Johansen           | 0.3300     | 0.2900     | 0.4700     | 0.3800     | 0.9000   |
| Standard prior     | 0.0100     | 0.0000     | 0.0000     | 1.0000     | 1.0000   |
| Optimal prior      | 0.2800     | 0.5700     | 0.0600     | 1.0000     | 1.0000   |
| Cointegrated prior | 0.2600     | 0.5900     | 0.0800     | 1.0000     | 1.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: MEAN VALUE**  
DEPENDENT VARIABLE:  $Y_1$

Table 6

|                    | $b_{11}^1$ | $b_{12}^1$ | $b_{21}^1$ | $b_{22}^1$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.5035     | -0.0880    | 0.3947     | -0.0228    | 0.1062   |
| Johansen           | 0.5730     | -0.0410    | 0.4300     | 0.0131     | 0.0125   |
| Standard prior     | 0.9963     | -0.0006    | 0.1643     | -0.1643    | 0.0000   |
| Optimal prior      | 0.4679     | 0.0000     | 0.4154     | 0.0663     | 0.0000   |
| Cointegrated prior | 0.4618     | 0.0001     | 0.4286     | 0.0540     | 0.0000   |
| Theoretical value  | 0.6000     | 0.0000     | 0.4000     | 0.0000     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: MEAN VALUE**  
DEPENDENT VARIABLE:  $Y_2$

Table 7

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.7890     | 0.0584     | 0.1281     | -0.0996    | 0.0618   |
| Johansen           | 0.8284     | 0.0815     | 0.1599     | -0.0710    | 0.0010   |
| Standard prior     | 0.5051     | -0.4972    | 0.9929     | -0.0007    | 0.0000   |
| Optimal prior      | 0.6045     | -0.8075    | 0.4661     | -0.0001    | 0.0000   |
| Cointegrated prior | 0.6050     | -0.0829    | 0.4600     | -0.0001    | 0.0000   |
| Theoretical value  | 0.8000     | 0.0000     | 0.2000     | 0.0000     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: MEDIAN VALUE**  
**DEPENDENT VARIABLE:  $Y_1$**

**Table 8**

|                    | $b_{11}^1$ | $b_{12}^1$ | $b_{21}^1$ | $b_{22}^1$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.5036     | -0.0997    | 0.3935     | -0.0443    | 0.0921   |
| Johansen           | 0.5748     | -0.0514    | 0.4193     | -0.0018    | 0.0054   |
| Standard prior     | 0.9964     | -0.0006    | 0.1601     | -0.1634    | 0.0000   |
| Optimal prior      | 0.4589     | 0.0000     | 0.4149     | -0.0525    | 0.0000   |
| Cointegrated prior | 0.4601     | 0.0000     | 0.4361     | -0.0446    | 0.0000   |
| Theoretical value  | 0.6000     | 0.0000     | 0.4000     | 0.0000     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: MEDIAN VALUE**  
**DEPENDENT VARIABLE:  $Y_2$**

**Table 9**

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.7624     | 0.0631     | 0.1360     | -0.1028    | 0.0446   |
| Johansen           | 0.8166     | 0.0773     | 0.1806     | -0.0546    | -0.0053  |
| Standard prior     | 0.4956     | -0.4930    | 0.9931     | -0.0008    | 0.0000   |
| Optimal prior      | 0.5984     | -0.0667    | 0.4581     | 0.0000     | 0.0000   |
| Cointegrated prior | 0.6111     | -0.0629    | 0.4583     | 0.0000     | 0.0000   |
| Theoretical value  | 0.8000     | 0.0000     | 0.2000     | 0.0000     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: INTERQUARTILE RANGE**  
**DEPENDENT VARIABLE:  $Y_1$**

Table 10

|                    | $b_{11}^1$ | $b_{12}^1$ | $b_{21}^1$ | $b_{22}^1$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.2434     | 0.2660     | 0.2376     | 0.1757     | 0.1029   |
| Johansen           | 0.2524     | 0.3013     | 0.2361     | 0.2035     | 0.0370   |
| Standard prior     | 0.0011     | 0.0004     | 0.0821     | 0.0849     | 0.0000   |
| Optimal prior      | 0.1351     | 0.0001     | 0.1467     | 0.0920     | 0.0000   |
| Cointegrated prior | 0.1459     | 0.0001     | 0.1227     | 0.0732     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: INTERQUARTILE RANGE**  
**DEPENDENT VARIABLE:  $Y_2$**

Table 11

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.3190     | 0.2764     | 0.2595     | 0.2359     | 0.1041   |
| Johansen           | 0.3230     | 0.3125     | 0.2489     | 0.2526     | 0.0847   |
| Standard prior     | 0.1081     | 0.1102     | 0.0013     | 0.0008     | 0.0000   |
| Optimal prior      | 0.2342     | 0.1809     | 0.1354     | 0.0001     | 0.0000   |
| Cointegrated prior | 0.2188     | 0.1652     | 0.1464     | 0.0001     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: STANDARD DEVIATION**  
**DEPENDENT VARIABLE:  $Y_1$**

**Table 12**

|                    | $b_{11}^1$ | $b_{12}^1$ | $b_{21}^1$ | $b_{22}^1$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.1972     | 0.2243     | 0.1848     | 0.1574     | 0.0941   |
| Johansen           | 0.1875     | 0.2324     | 0.1964     | 0.1587     | 0.0446   |
| Standard prior     | 0.0009     | 0.0003     | 0.0708     | 0.0708     | 0.0000   |
| Optimal prior      | 0.1190     | 0.0003     | 0.1116     | 0.0820     | 0.0000   |
| Cointegrated prior | 0.1218     | 0.0002     | 0.1114     | 0.0782     | 0.0000   |

Sample size: 30 observations

**ESTIMATED COEFFICIENTS: STANDARD DEVIATION**  
**DEPENDENT VARIABLE:  $Y_2$**

**Table 13**

|                    | $b_{11}^2$ | $b_{12}^2$ | $b_{21}^2$ | $b_{22}^2$ | constant |
|--------------------|------------|------------|------------|------------|----------|
| OLS                | 0.2359     | 0.2225     | 0.1932     | 0.1794     | 0.0772   |
| Johansen           | 0.2358     | 0.2226     | 0.1975     | 0.1913     | 0.0647   |
| Standard prior     | 0.0940     | 0.0949     | 0.0012     | 0.0006     | 0.0000   |
| Optimal prior      | 0.1575     | 0.1348     | 0.1191     | 0.0004     | 0.0000   |
| Cointegrated prior | 0.1443     | 0.1242     | 0.1217     | 0.0002     | 0.0000   |

Sample size: 30 observations

## REFERENCES

- AHN, S.K. and REINSEL, G.C. (1990): "Estimation for partially nonstationary multivariate autoregressive models". *Journal of the American Statistical Association*, 85, 813-823.
- BALLABRIGA, F. (1988): "Transmission mechanisms in an open economy: Spain 1969-1984". *Ph.D.Thesis. University of Minnesota*.
- BANERJEE, A., DOLADO, J.J., HENDRY, D.F. and SMITH G.W. (1986): "Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence". *Oxford Bulletin of Economics and Statistics*, 48, 253-277.
- BANERJEE, A., DOLADO, J. GALBRAITH, J.W. AND HENDRY, D.F. (1993): "*Cointegration, error-correction, and the econometric analysis of non-stationary data*". Oxford University Press. Oxford.
- BOX, G.E.P. and TIAO, G.C. (1977): "A canonical analysis of multiple time series". *Biometrika*, 64, 355-365.
- CAMPBELL, J.Y. and PERRON, P. (1991): "Pitfalls and opportunities: what macroeconomists should know about unit roots". Blanchard, O.J. and Fischer, S. (eds). *NBER Macroeconomics Annual*. The MIT Press, Cambridge, Massachusetts, 141-201
- CANOVA, F. (1988): "Essays in seasonality in time-series". *Ph.D.thesis. University of Minnesota*.
- CANOVA, F. (1992): "An alternative approach to modeling and forecasting seasonal time series". *Journal of Business and Economic Statistics*, 10, 97-108.
- CANOVA, F. (1993): "Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model". *Journal of Economic Dynamics and Control*, 17, 233-261.
- CLEMENTS, M.P. and MIZON, G.E. (1991): "Empirical analysis of macroeconomic time series": VAR and structural models. *European Economic Review*, 35, 887-917.
- CHRISTIANO, L. and EICHENBAUM, M. (1990): "Unit roots in real GNP: Do we know and do we care?". *Carnegie-Rochester Conference Series on Public Policy*, 32, 7-61.

- DOAN, T. LITTERMAN, R.F. and SIMS, C. (1984): "Forecasting and conditional projections using realistic prior distributions". *Econometric Reviews*, 3, 1-100.
- DOLADO, J.J. (1992): "A note on weak exogeneity in VAR cointegrated models". *Economics Letters*. 38, 139-143.
- EITRHEIM, O. (1992): "Inference in small cointegrated systems: some Monte Carlo results". Discussion Paper 92-31. *University of California, San Diego*.
- ENGLE, R.F. and GRANGER, C.W.J. (1987): "Cointegration and error correction: representation, estimation and testing". *Econometrica*, 55, 251-76.
- GONZALO, J. (1991): "Estimation of long run equilibrium relationships and common long memory components in cointegrated systems". *Ph.D.thesis. University of California, San Diego*.
- GRANGER, C.W.J. (1981): "Some properties of time series data and their use in econometric model specification". *Journal of Econometrics*, 16, 121-130.
- GRANGER, C.W.J. and WEISS A.A. (1983): "Time series analysis of error correction models". *Studies in Econometrics, Time Series and Multivariate Statistics*. Karlin, S., Amemiya, T. and Goddman, L.A. (eds). New York, Academic Press, 255-278.
- JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors". *Journal of Economics Dynamics and Control*, 12, 231-254.
- JOHANSEN, S. (1991a): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". *Econometrica*, 59, 1551-1580.
- JOHANSEN, S. (1991b): "The power function of the likelihood ratio test for cointegration". *Econometric Decision Models: New Methods of Modeling and Applications*. Gruber J. (ed) Springer Verlag, 323-335.
- JOHANSEN, S. (1992a): "An I(2) cointegration analysis of the purchasing power parity between Australia and the United States". Hargreaves, C. (ed.). *Macroeconomic modelling of the long run*. Edward Elgar, Aldershot, 229-248.
- JOHANSEN, S. (1992b): "Cointegration in partial systems and the efficiency of single-equation analysis". *Journal of Econometrics*, 52, 389-402.
- JOHANSEN, S. and JUSELIUS, K. (1992): "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK". *Journal of Econometrics*, 53, 211-44.

- JUSELIUS, K. (1992): "On the duality between long-run relations and common trends in the I(1) versus I(2) models. An application to aggregate money holdings". Forthcoming in *Econometric Reviews*.
- LITTERMAN, R. (1980): "Techniques for forecasting with vector autoregressions". *Ph.D.thesis. University of Minnesota*.
- LITTERMAN, R. (1984): "Specifying vector autoregressions for macroeconomic forecasting". *Federal Reserve of Minneapolis. Staff Report 92*.
- LITTERMAN, R. (1986): "Forecasting with Bayesian vector autoregressions - five years of experience". *Journal of Business and Economic Statistics* 4, 25-38.
- LÜTKEPOHL, H. (1991): *Introduction to Multiple Time Series Analysis*. Springer Verlag. New York.
- PARK, J.Y. and PHILLIPS, P.C.B. (1989): "Statistical inference in regressions with integrated processes: Part 2". *Econometric Theory*, 5, 95-131.
- PEÑA, D. and BOX, G.E.P. (1987): "Identifying a simplifying structure in time series". *Journal of the American Statistical Association*, 82, 836-843.
- PHILLIPS, P.C.B (1991a): "To criticize the critics: an objective Bayesian analysis of stochastic trends". *Journal of Applied Econometrics*, 6, 333-334.
- PHILLIPS, P.C.B. (1991b): "Bayesian routes and unit roots: de rebus prioribus semper est disputandum". *Journal of Applied Econometrics*, 6, 435-473.
- PHILLIPS, P.C.B. (1991c): "Optimal inference in cointegrated systems". *Econometrica*, 59, 283-306.
- QUAH, D. (1992): "The relative importance of permanent and transitory components: identification and some theoretical bounds". *Econometrica*, 60, 107-118.
- RAYNAULD, J. and SIMONATO, J.C. (1993): "Seasonal BVAR models". *Journal of Econometrics*, 55, 203-229.
- REIMERS, H.E. (1992): "Comparisons of tests for multivariate cointegration". *Statistical Papers*, 33, 335-359.
- REINSEL, G.C. and AHN, S.K. (1992): "Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test and forecasting". *Journal of Time Series Analysis*, 13, 353-75.
- SIMS, C. (1986): "Bayesmith: a program for multivariate Bayesian interpolation". Discussion Paper n° 234. *Center of Economic Research. University of Minnesota*.



- SIMS, C. (1988): "Bayesian skepticism on unit root econometrics". *Journal of Economic Dynamics and Control*, 12, 463-474.
- SIMS, C. (1989): "A nine variable probabilistic model of the US economy". *Institute for Empirical Macroeconomics, Federal Reserve of Minneapolis*. Discussion Paper, 14.
- SIMS, C. (1991a): "VAR macroeconometrics: an update". *Yale University*. Manuscript.
- SIMS, C. (1991b): Comment on "To criticize the critics" by Peter C.B. Phillips. *Journal of Applied Econometrics*, 6, 423-34.
- SIMS, C. (1991c): Comment on "Empirical analysis of macroeconomic time series: VAR and structural models" by Michael P. Clements and Grayham E. Mizon. *European Economic Review*, 35, 922-932.
- SIMS, C., STOCK, J. and WATSON, M. (1990): "Inference in linear time series models with some unit roots". *Econometrica*, 58, 113-144.
- SIMS, C. and UHLIG, H. (1991): "Understanding unit rooters: a helicopter tour". *Econometrica*, 59, 1591-1599.
- THEIL, H. (1963): "On the use of incomplete prior information in regression analysis". *Journal of the American Statistical Association*, 58, 401-414.
- THEIL, H. and GOLDBERGER, A. S. (1961): "On pure and mixed statistical estimation in economics". *International Economic Review*, 2, 65-78.
- VELU, R.P., WICHERN, D.W. and REINSEL, G.C. (1986): "Reduced rank models for multiple time series". *Biometrika*, 73, 105-111.

