Valuing Structure, Model Uncertainty and Model Averaging in Vector Autoregressive Processes

Econometric Institute Report EI 2004-23

Rodney W. Strachan[1]
Herman K. van Dijk[2]
[1]Department of Economics, Keele University,
Keele, Staffs ST5 5BG, U.K.
email: r.w.strachan@econ.keele.ac.uk

[2]Econometrics Institute,
Erasmus University Rotterdam,
Rotterdam, The Netherlands
email: hkvandijk@few.eur.nl

May 18, 2004

ABSTRACT

Economic policy decisions are often informed by empirical analysis based on accurate econometric modeling. However, a decision-maker is usually only interested in good estimates of outcomes, while an analyst must also be interested in estimating the model. Accurate inference on structural features of a model improves policy analysis as it improves estimation, inference and forecast efficiency. In this paper a Bayesian inferential procedure is presented which allows for unconditional inference on structural features of vector autoregressive (VAR) processes. We employ measures on manifolds in order to elicit uniform priors on subspaces defined by particular structural features of VARs. The features considered are cointegration, exogeneity, deterministic processes and overidentification. Posterior probabilities of these features are used in a model averaging approach for forecasting and impulse response analysis. The methods are applied to three empirical economic issues: stability of Australian money demand; relative weights of permanent and transitory shocks in a US real business cycle model; and possible evidence on an inflationary oil price shock and a liquidity trap in a UK macroeconomic model. The results obtained illustrate the feasibility of the proposed methods.

Key Words: Posterior probability; Grassman manifold; Orthogonal group; Cointegration; Exogeneity; Model averaging.

JEL Codes: C11, C32, C52

1 Introduction.

An important function of empirical economic analysis is to provide accurate information for decision making. This information is generally provided in the form of estimates of objects of interest such as forecasts of endogenous variables like money demand, effects of oil price shocks on domestic inflation, or as relative weights of transitory and permanent components in a real business cycle model. One may infer from this brief list that, in many cases, the decision maker is not directly interested in the underlying model used to produce such estimates. It is, however, in the analyst's interest to detail how the results she provides rely upon the model. That is, the analyst, when providing estimates of the objects of interest, must point out "The results are conditional upon the validity of the assumptions used in the model ..." Such restrictions upon the interpretation of the results do not necessarily aid the decision-maker in her task.

It is generally accepted that, in order to improve policy analysis, it is important to have accurate inference on the support for the alternative models considered or to have such inference on the structural features of an encompassing model. As such, much effort is expended in investigating the empirical support for various economically and statistically plausible features. Examples of features of models that are of interest to analysts - but not necessarily decision-makers - include numbers of long run relationships among variables, forms of these long run relationships, persistent and predictable long run behaviour of variables, short term behaviour, and the dimension of the system in variables or in parameters required for the problem of interest. Each of these features implies zero restrictions on particular parameters in a general model. If these features are supported by the data - and so are credible in the sense of Sims (1980) - and if they hold outside the sample, then imposing them can improve forecasts and inference, and hence policy suggestions. The support in the data is, however, in many cases not clear or dogmatically for or against the restriction, and researchers usually do not have strong prior belief in particular restrictions. Yet it is common to *condition* upon such features, effectively assigning a weight of one to the model implied by the restrictions being true and zero to all other plausible models. Even if the support is strongly for or against a particular restriction, with only slight support for the alternative unrestricted model, imposing the restriction ignores information from that less likely model which, if appropriately weighted, could improve forecasts.

There is therefore a conflict between the analyst's need to obtain the best model and the decision-maker's need for the least restrictive interpretation of the information provided by the analyst. As an alternative to conditioning on structural features, it is possible to improve policy analysis by presenting unconditional or averaged information. Gains in forecasting accuracy by simple averaging have been pioneered by Bates and Granger (1969) and discussed recently by Diebold and Lopez (1996), Newbold and Harvey (2001) and Terui and van Dijk (2002). Some explanation for this phenomenon in particular cases was provided by Hendry and Clements (2002). Alternatively, the averaging weights can be determined to reflect the support for the model from which each estimate derives. This requires accurate reflection upon the uncertainty associated with the structural features defining the model.

In this paper we present a Bayesian approach for conducting unconditional inference on structural features of the vector autoregressive model. Specifically, we focus on three contributions. First, a general operational procedure is presented for specifying diffuse prior information on structural features of interest which implies well-defined posteriors and existence of moments. For a more intuitive, specific application we refer to Strachan and Inder (2004). Given the diffuse prior, the information in the likelihood function is supposed to dominate. As a result one can evaluate the relative weights or probabilities of such structural features as cointegration, overidentification, deterministic processes, and exogeneity. In order to obtain these results we make use manifolds and orthogonal groups and their measures. Then we can elicit uniform prior measures on relevant subspaces of the parameter space. From these measures we develop prior distributions for elements of these subspaces as the parameter of interest. Thus we choose prior specification on models directly rather than on parameters that are subsequently restricted.

Second, using this methodology we show in this paper how to obtain posterior inference and forecasts from *model averages* in which the economically and econometrically important structural features may have weights other than zero or one.

Third, we apply the proposed methodology to three empirical economic issues. We start with the issue of stability of the money demand equation using Australian post war data. Next, we investigate the relative weights of permanent and transitory components in a real business cycle model of the USA due to King, Plosser, Stock and Watson (1991). Third, we analyze the posterior evidence of a UK inflationary oil price shock and the evidence on a UK liquidity trap using the macro-econometric model of Garratt, Lee, Pesaran and Shin (2000).

The structure of the paper is as follows. In the Section 2 we introduce the general model of interest in this paper - the vector autoregressive model, the general structural features of interest, and the restrictions they imply. In Section 3 we present the priors, the likelihood and a general expression for the posterior. The tools for inference in this paper, the Bayes factor and posterior probabilities, are introduced and expressions are derived for specific features of interest like impulse responses. Our approach is a significant divergence from much of the earlier work. This section therefore provides a discussion of the advantages of this approach in the context of model averaging. We demonstrate the approach in Section 4 with three applications using a model of Australian money demand, A US real business cycle model and a UK macroeconomic model. In Section 5 we summarize conclusions and discuss possibilities for further research.

2 The Vector Autoregressive Model.

Ever since the influential work by Sims (1980), modeling economically important issues - for instance possible effects of oil price shocks on domestic inflation - has lead to the use of vector autoregressive processes. These processes can incorporate a wide range of short and long run dynamic, equilibrium and deterministic behaviours. Further, it has been observed in empirical studies, that many economic variables of interest are not stationary, yet economic theory, or empirical evidence, suggests stable long run relationships to exist among these variables. The statistical theory of cointegration (Granger, 1983, and Engle and Granger, 1987), in which a set of nonstationary variables combine linearly to form stationary relationships, and the attendant Granger's representation theorem provide a useful specification to incorporate this economic behaviour into the error correction model and allows the separation of long run and short run behaviour. We work with the vector autoregressive model in the error correction form to simplify expressions of restrictions. For more details on a - likelihood - analysis of VAR models with cointegration restrictions we refer to Johansen (1995).

The error correction model (ECM) of the $1 \times n$ vector time series process y_t, $t = 1, \ldots, T$, conditioning on the l observations $t = -l + 1, \ldots, 0$, is

$$\Delta y_t = y_{t-1}\beta^{+}\alpha + d_t\theta + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \tag{1}$$

$$
\begin{aligned}
&= y_{t-1}\beta^{+}\alpha + d_t\theta_1\alpha + d_t\theta_0 + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \\
&= z_{1,t}\beta\alpha + z_{2,t}\Phi + \varepsilon_t \qquad (2)
\end{aligned}
$$

where $\Delta y_t = y_t - y_{t-1}$, $z_{1,t} = (d_t, y_{t-1})$, $z_{2,t} = (d_t, \Delta y_{t-1}, \ldots, \Delta y_{t-l})$, $d_t = (1, t)$, $\Phi = (\theta_0', \Gamma_1', \ldots, \Gamma_l')'$ and $\beta = \left(\theta_1', \beta^{+\prime}\right)'$. The matrices β^{+} and α' are $n \times r$ and assumed to have rank r, and if $r = n$ then $\beta^{+} = I_n$. In (2) we have used the decomposition $\theta = \theta_1\alpha + \theta_0$ which we will explain further in a later subsection.

The following subsections define the restrictions of interest, combinations of which define different model features of interest which we may compare or weight using posterior probabilities.

As we consider a wide range of models in this paper, we will use a consistent notation to index each model to identify the cointegrating rank, the identifying restrictions, the form of exogeneity, and the deterministic processes in the model. We will denote the cointegrating rank of a model by r, where $r = 0, 1, \ldots, n$. The particular identifying restrictions placed upon β will be denoted by o, where $o = 0, 1, \ldots, \mathfrak{I}$ and $o = 0$ will be understood to refer to the just identified model. Partitioning y_t as $y_t = (y_{1,t}\ y_{2,t})$ where $y_{1,t}$ is a $1 \times n_1$ vector, $n_1 \geq r$, exogeneity of $y_{2,t}$ will be considered with respect to subsets of the parameters in the equation for $y_{1,t}$, where will use ϕ_1 and ϕ_2 to denote these subsets. The particular form of exogeneity restrictions in the model will be denoted by e, where $e = 0, \ldots, 4$ and these are defined below. Finally, the particular form of deterministic processes will be denoted by the couplet (i, j) where $i, j = 0, 1, 2$. The i refers to the processes in Δy_t and j refers to the processes in $y_t\beta^{+}$. Five commonly considered combinations of processes are detailed in the subsection on deterministic processes below. We also allow for a range of lags of differences, however as these have little economic importance for the studies we look at, and for space considerations, we do not discuss these further.

The vector identifying a particular model will therefore be $\omega = (r, o, e, i, j)$. For example, the least restricted model will be $(n, 0, 0, 0, 0)$, while the most restricted model will be $(0, o, 4, 2, 2)$. The models will be identified as M_{ω}. When we are considering only a particular feature such as exogeneity, we will indicate this by referring to the model as $M_{(.,.,e,.)}$, and if we are conditioning upon a particular feature, such as rank, $M_{(e|r)}$. Where we have averaged across or marginalised with respect to the other features, we will indicate this by $M_{(r)}$, and the marginal likelihood for a model will be m_{ω}.

Finally, we introduce the following terms to simplify the expressions in the posteriors. Let $\widetilde{z}_t = (z_{1,t}\beta \quad z_{2,t})$, and the $(r + k_i) \times n$ matrix $B = [\alpha' \ \Phi']'$. We may now write the model, given in (1) as

$$\Delta y_t = \widetilde{z}_t B + \varepsilon_t. \tag{3}$$

2.1 Structural features

Within model (1), a number of structural features are commonly of interest to economists and or econometricians. Here we detail five of these and the restrictions they imply for (1). To demonstrate we use a simple example: money demand. The variables, all of which appear in logarithmic form, are defined as $y_t = (m_t \quad inc_t)$, where m_t is the log measure of real money and inc_t is the measure of real income. A decision maker may be interested in the one step ahead forecast of m_t or the overall response path of m_t to a shock in inc_{t-h} for $h = 0, 1, 2, \ldots$.The analyst has to estimate the parameters determining the long and short run behaviour of m_t and produce forecasts of m_t, where the forecasts may also be over the long run, or both the long run and short run.

2.1.1 Cointegration

For cointegration analysis of (1), of interest is the coefficient matrix β^+ and α which are of rank $r \leq n$. Of particular interest then, is r which implies there are $(n - r)$ common stochastic trends in y_t, and r is the number of $I(0)$ combinations of the element of y_t extant. In the case $r < n$ and assuming for now $\theta = 0$, β^+ is the matrix of cointegration coefficients, $y_t\beta^+$ are the stationary relations towards which the elements of y_t are attracted, and α is the matrix of factor loading coefficients or adjustment coefficients determining the rate of adjustment of y_t towards $y_t\beta^+$.

In the money demand example, $r \in [0, 1, 2]$. It is common to regard the money demand relation as the cointegrating relation between the integrated variables in $y_t \sim I(1)$, and supply is exogenous (see for example Johansen, 1995 and Funke, Hall and Beeby, 1997). That is, $\zeta_t = \beta_1 m_t + \beta_2 inc_t = y_t\beta^+ \sim I(0)$, $E(\zeta_t) = d_t\theta_1$ and possibly $\theta_1 \neq 0$. Therefore, for the analysis to make sense, we require that cointegration should hold (and so $r = 1$). In this case for (1) one has $\beta^+ = (\beta_1, \beta_2)'$ and $\alpha = (\alpha_1, \alpha_2)$.

2.1.2 Exogeneity

As is usually accepted in econometric modelling, there are benefits from parsimony. Thus, an important issue is the dimension of the system to be estimated in terms of the number of equations. Recall the partition $y_t = (y_{1,t}\ y_{2,t})$. If the set of variables in $y_{2,t}$ can be treated as exogenous for inferential purposes, a partial system may be estimated in which no equations are estimated for these variables. This is essentially ignoring information that contributes nothing to the inference. As an example, it is not uncommon to assume that to estimate the income elasticity of money, a researcher would be interested in whether an equation for income need be estimated, or could this analysis be done with a single equation. In certain cases, exogeneity has been given an economic interpretation such as in Garratt, Lee, Pesaran, and Shin (2002) who treat oil prices as a 'long-run forcing' variable.

Under the condition of cointegration, the representation of the model in (1) will be useful for the analysis of exogeneity. Partition $\alpha = (\alpha_1 \quad \alpha_2)$ conformably with the dimensions of $y_{1,t}$ and $y_{2,t}$. In this paper we consider weak exogeneity of $y_{2,t}$ with respect to the parameters influencing long run behaviour of $y_{1,t}$, $\phi_1 = \left(vec\,(\beta)', vec\,(\alpha_1)'\right)'$. As shown in Urbain (1992) and Johansen (1992) *inter alia*, $y_{2,t}$ will be weakly exogenous with respect to ϕ_1 if $\alpha_2 = 0$. To preserve the rank of α requires that $n_1 \geq r$, which implies we cannot have more than $n - r$ variables weakly exogenous with respect to ϕ_1. An important model in the literature which relies upon this assumption is the triangular model (Phillips, 1991) used by Phillips (1994) in which $n_1 = r$.

For a given cointegrating rank r, denote by $M_{(e|r)}$ the various models of exogeneity. The model with no exogeneity restrictions imposed is $e = 0$ and the model with weak exogeneity of $y_{2,t}$ with respect to ϕ_1 is $e = 1$. Other forms of exogeneity include: strong exogeneity of $y_{2,t}$ with respect to the parameters influencing long run behaviour of $y_{1,t}$, ϕ_1 $(e = 2)$; weak exogeneity of $y_{2,t}$ with respect to the parameters influencing long and short run behaviour of $y_{1,t}$ $\left(\phi_2 = \left(\phi_1', vec\,(\Gamma_{11})', vec\,(\Gamma_{21})'\right)'\right)$ $(e = 3)$; and strong exogeneity of $y_{2,t}$ with respect to the parameters influencing long and short run behaviour of $y_{1,t}$ $(e = 4)$. These imply further restrictions upon the parameters in (1) such as Granger noncausality. We do not explore them here as the first case is sufficient to demonstrate the approach.

If we are interested in whether we may estimate the money demand equation ζ_t (and so estimate β^+) from a single equation for m_t, then this would require that the variable inc_t be weakly exogenous with respect to β^+ $(e = 1)$.

α_1 is the adjustment coefficient in the equation for Δm_t and α_2 is the same in the equation for inc_t such that these parameters determine the response in y_t to a nonzero value of ζ_{t-1}. Weak exogeneity of inc_t with respect to β^+ therefore implies $\alpha_2 = 0$.

2.1.3 Overidentifying restrictions on the cointegrating vectors

As discussed in Garratt *et al.* (2002), when modelling economic systems, economic theory tends be more useful when it focuses upon the form of long run, or equilibrium, relationships between variables and leaves the short run relations unrestricted (see Sims 1980 for discussion about the dangers of imposing incredible restrictions on short run dynamics). For money demand, the stability of the (log of the) inverse velocity of money, $\nu_t = m_t - inc_t$ is an important issue for econometric analysis. We note that stability is here interpreted in the sense that velocity is $I(0)$ but may have deterministic trends - we discuss this latter possibility in the following subsection.

In both the classical and Bayesian approaches, to test the appropriateness of such restrictions and to estimate the restricted model, requires a specification of the model subject to these restrictions. In the classical maximum likelihood approach, Johansen (1995) has provided methods for estimation with, and testing of, these restrictions. The three restrictions commonly investigated are presented in Johansen (1995, Chapter 5) as the following hypotheses.

$(o = 0)$ No restrictions upon β.

$(o = 1)$ $\beta = H\psi$

where the dimensions of the respective matrices are: H $n \times s$, ψ $s \times r$, $r \leq s$.

$(o = 2)$ $\beta = (b\ \varphi) = (b\ b_\perp \psi)$

where the dimensions of the respective matrices are: b $n \times s$, $b_\perp$ $n \times (n-s)$, ψ $(n-s) \times (r-s)$, $s \leq r$.

$(o = 3)$ $\beta = (H_1\psi_1, H_2\psi_2, \ldots, H_l\psi_l)$

where the dimensions of the respective matrices are: H_i $n \times s_i$, ψ_i is $s_i \times r_i$, $r_i \leq s_i$, $l \leq r$, $\sum_i r_i = r$.

The restriction in $o = 0$ imposes no restriction on the space of β, in $o = 1$ the cointegrating space is completely determined (given r). The third restriction, $o = 2$, restricts the cointegrating space to pass through a known vector or set of s vectors, b, and the remaining $r - s$ vectors, $b_\perp\psi$, are unknown except that they are orthogonal to b, such that the space of β is not completely known. The final hypothesis, $o = 3$, generalizes the first two.

2.1.4 Deterministic terms

Economists are commonly interested in the presence or absence of deterministic processes in y_t or $y_t\beta^+$. For both statistical and economic reasons, the persistent and predictable, or deterministic, component economic behaviour is important. Of interest are questions such as whether linear or quadratic drifts are present in y_t and whether nonzero constant terms and deterministic trends are present in $y_t\beta^+$. For example, the velocity of money in many countries has not remained stable over the long run. For extended periods it has displayed what appears to be a clear trend. If we were to assume the velocity was an equilibrium or long run relation of interest, it would be important to allow for some trend in this relation. It is well known, however, that simplistic treatment of the deterministic terms by testing whether θ or some elements of θ are zero leads to the strange and unsatisfactory situation that very different trending behaviour is implied in the levels of the process for different values of r. For an example involving a Bayesian analysis of purchasing power parity we refer to Schotman and Van Dijk (1991) . Therefore we decompose θ into $\theta = \theta_1\alpha + \theta_0$ where $\theta_1 = \theta\alpha' (\alpha\alpha')^{-1}$ and $\theta_0 = \theta\alpha'_\perp (\alpha_\perp\alpha'_\perp)^{-1} \alpha_\perp$ such that $\theta_1 = (\mu_1, \delta_1)'$ represents the deterministic processes associated with $y_t\beta^+$ and $\theta_0 = (\mu_0, \delta_0)'$ represents those for Δy_t (see Johansen, 1995 Section 5.7 for further discussion). Assuming $d_t = (1, t)$ (and $l = 0$ for simplicity) then

$$E(\Delta y_t) = d_t\theta_0 = \mu_0 + t\delta_0$$

and

$$E(y_t\beta^+) = d_t\theta_1 = \mu_1 + t\delta_1.$$

Although a wider range of models are clearly available, the five most commonly considered may be stated as follows, where in $M_{(i,j|r)}$, the i denotes the model for Δy_t and the j denotes the model for $y_t\beta^+$ at given rank r :

$$\begin{aligned}
M_{(0,0|r)} &: E(\Delta y_t) = \mu_0 + \delta_0 t \text{ and } E(y_t\beta^+) = \mu_1 + \delta_1 t \\
M_{(1,0|r)} &: E(\Delta y_t) = \mu_0 \text{ and } E(y_t\beta^+) = \mu_1 + \delta_1 t \\
M_{(1,1|r)} &: E(\Delta y_t) = \mu_0 \text{ and } E(y_t\beta^+) = \mu_1 \\
M_{(2,1|r)} &: E(\Delta y_t) = 0 \text{ and } E(y_t\beta^+) = \mu_1 \\
M_{(2,2|r)} &: E(\Delta y_t) = 0 \text{ and } E(y_t\beta^+) = 0
\end{aligned}$$

3 Priors and posteriors.

In this section the forms of the priors and resultant posterior are presented. We restrict ourselves to flat priors where possible, although consideration is given to informative priors when discussing the parameters of interest. For the model in (3), assume the rows of the $T \times n$ matrix $\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_T')'$ are $\varepsilon_t \sim iidN(0, \Sigma)$. The likelihood can then be written as

$$L\left(y|\Sigma, B, \beta, \omega, \widetilde{Z}\right) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\varepsilon'\varepsilon\right)\right\}. \quad (4)$$

3.1 The prior for (Σ, B, ω).

The priors for the elements of $\omega = (r, o, e, i, j)$ are not independent, as certain combinations are either impossible, meaningless (such as, for example, $r = 0$ with $o = 2$) or observationally equivalent to another combination (such as, for example, $r = 0$ with $o = 2$ or $r = n$ with $i = 1$ or 2). However, we specify all combinations of the indices in ω be equally likely to avoid biasing the evidence in favour of other classes of models. For example, at $r = 0$ the models with $(i, j) = (0, 1)$ and $(i, j) = (1, 1)$ are observationally equivalent. If we were to treat these two as one model such that they receive half the prior probability of other models, systematic employment of this principle would bias the prior weight in favour of models with $0 < r < n$, and thus bias the evidence in favour of some economic theories for which we wish to determine the support. If a user wishes to treat these models differently, they made do so by reweighting the output from our analysis. This implies we use $p(r) = (n+1)^{-1}$ and $p(i, j) = 0.2$. As we consider weak exogeneity, the prior density for the states of exogeneity $e \in [1, 2]$ is $p(e|r) = 1/2$ and for the states of overidentification of β we consider, $o \in [1, 4]$, $p(o) = 1/2$. The standard diffuse prior for Σ, $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$ is used.

As B changes dimensions across the different models of ω and each element of the matrix B has the real line as its support, the Bayes factors for different models will not be well defined if an improper prior on B, such as $p(B|\beta, \omega) \propto 1$ were used. For discussion on this point see (among many others) Lindley (1957), Bartlett (1957), Jeffreys (1961) and more recently O'Hagan (1995), Strachan and van Dijk (2003) and Strachan and van Dijk (2004). For this reason a weakly informative proper prior for B must be used. We take the prior for B conditional upon (Σ, β, ω) as normal with zero mean and covariance $\Sigma \otimes \left(\widetilde{\beta}' \underline{H} \widetilde{\beta}\right)^{-1}$ where $\underline{H} = \tau I_{(r+k_i)}$ and

$$\widetilde{\beta} = \begin{bmatrix} \beta & 0 \\ 0 & I_{k_i} \end{bmatrix}$$

such that $\widetilde{\beta}' \underline{H} \widetilde{\beta} = \tau I_{(r+k_i)}$. We choose the value of $\tau = 0.5$ as this provides a reasonably flat prior. Evidence on the influence of this choice can be found in Strachan and Inder (2004).

3.2 Eliciting a prior on β.

There is a large body of work in Bayesian cointegration analysis in which methods are proposed to deal with a range of problems. A feature of much of this work is that linear identifying restrictions were adopted to enable estimation of β. We start with a brief discussion of this work as it relates to the aim of this paper - model averaging. We then present a general analysis of our alternative approach. For specific applications and a more intuitive explanation of this approach in different contexts, we refer to Strachan and Inder (2004) and Strachan and van Dijk (2003).

Linear restrictions and the cointegrating space: It is well known that as β and α appear as a product in (2), r^2 restrictions need to be imposed on the elements of β and α to just identify these elements. These restrictions are commonly imposed upon β by assuming $c\beta$ is invertible for known $(r \times n)$ matrix c and the restricted β to be estimated is $\overline{\beta} = \beta (c\beta)^{-1}$. The free elements are collected in $\overline{\beta}_2 = c_\perp \overline{\beta}$ where $c_\perp c' = 0$. A common choice in theoretical work is $c = [I_r \ 0]$ such that $\overline{\beta} = \left[I_r \ \overline{\beta}_2'\right]'$. A prior is then specified for $\overline{\beta}_2$ which is then estimated and often its value is interpreted.[1]

Assuming that c is known, Kleibergen and van Dijk (1994, 1998) and Bauwens and Lubrano (1996) detail remaining pathologies and features which complicate analysis associated with the posterior for $\overline{\beta}_2$ with a flat prior. Kleibergen and van Dijk (1994) demonstrate how a variable addition specification - which would provide a natural way of performing inference on r by nesting the reduced rank model within a full rank model - results in an

[1] There exist practical problems with incorrectly selecting c. The implications for classical analysis of this issue are discussed in Boswijk (1996) and Luukkonen, Ripatti and Saikkonen (1999) and in Bayesian analysis by Strachan (2003). In each of these papers examples are provided which demonstrate the importance of correctly determining c.

improper posterior distribution at reduced ranks, thus precluding inference. For the non-nested reduced rank model, as in (2), Kleibergen and van Dijk (1994) outline the additional issue of local nonidentification which manifests itself in the likelihood and results in asymptotes in the marginal posterior distributions, nonexistence of moments of $\overline{\beta}_2$, and precludes the use of MCMC due to reducibility of the Markov chain. As a solution they propose using the Jeffreys prior as the behaviour of this prior in problem areas of the support offsets the problematic behaviour of the likelihood. Kleibergen and van Dijk (1998), Kleibergen and Paap (2002), and Paap and Van Dijk (2003) use a singular value decomposition to nest the rank $r < n$ model within the rank n model. Importantly, they include in the posterior the Jacobian for the transformation from the full rank model to the parameters of the reduced rank model into the posterior. In this specification, the Jacobian behaves in a similar way to the Jeffreys prior in the problem areas of the support, however this approach allows freer expression of prior beliefs than the Jeffreys prior. Use of the Jeffreys prior or the singular value decomposition avoid the issue of local nonidentification, result in proper posteriors and allow use of MCMC, however the posterior again has no moments of $\overline{\beta}_2$.

Bauwens and Lubrano (1996) begin with the reduced rank model and provide a study of the posterior distribution of $\overline{\beta}_2$. Using the results for the 1-1 $poly-t$ density of Drèze (1978), they show the posterior has no moments due to a deficiency of degrees of freedom. Similar results have been shown for the simultaneous equations model (Drèze 1978, Kleibergen and van Dijk 1998). Nonexistence of moments is not commonly a concern for estimation as modal estimates exist as alternative estimates of location. However, as the kernel of the 1-1 $poly-t$ is a ratio of the kernels of two $Student-t$ densities, the posterior may be bimodal - with the modes sometimes well apart from each other - making it difficult to both locate the global mode and bringing into question the interpretation of the mode as a measure of location.

Exogeneity is a commonly employed restriction and is important in two of our applications. For our applications in which we combine restrictions to define new models, we have the additional problem that the posterior for $\overline{\beta}_2$ is improper when exogeneity is imposed. A proof of this result is provided in Appendix I. An improper posterior is a significant issue as it implies we know *a priori* we cannot obtain inference on any estimate of an object of interest if exogeneity is imposed. Nonexistence of moments is a significant issue as this implies we know *a priori* any estimate of an object of interest, $g\left(\overline{\beta}_2\right)$ - obtained by averaging across the set of models - will not exist (or be infinite) if $g\left(\overline{\beta}_2\right)$ is a convex or linear function of $\overline{\beta}_2$.

Further, it is clear from the discussion on the prior for B that a flat prior on $\overline{\beta}_2$ cannot be employed to obtain posterior probabilities for ω, since the dimensions of $\overline{\beta}_2$ depend upon ω. As argued in the introduction, an advantage of the Bayesian approach is the ability to explicitly incorporate prior beliefs into the analysis. A flat improper prior is generally intended to reflect ignorance about the parameter of interest, therefore the above issues with the posterior at least, may be resolved by relinquishing this option and making use of an informative prior on $\overline{\beta}_2$. For example, a student-t prior may be used, or inequality restrictions - such as a marginal propensity to consume between zero and one - are often useful. Priors such as the Jeffreys prior have been proposed which may resolve some of the above problems, however this prior does not allow for evaluating proper posterior probabilities and model averaging. Therefore, to preserve the options of both informative and uninformative priors, to preserve the function of the prior as a representation of prior beliefs, to simplify the application and estimation, and as we do not see $\overline{\beta}_2$ as the parameter of interest, we diverge at this point from much of the earlier literature in both specifying our parameter of interest and eliciting an uninformative prior on that parameter.

The parameter of interest: It is necessary that we explain the above comment regarding 'the parameter of interest' and the implications of focussing upon and using $\overline{\beta}_2$. If we denote the space spanned by a matrix A by $sp\,(A)$, we can say that in cointegration analysis it is not the values of the elements of β that are the primary object of interest, rather it is the space spanned by β, $\mathfrak{p} = sp\,(\beta)$, and this space is in fact all we are able to uniquely estimate. The parameter $\mathfrak{p}$ is an r-dimensional hyperplane in R^n containing the origin and as such is an element of the Grassman manifold[2] $G_{r,n-r}$ (James, 1954), $\mathfrak{p} \in G_{r,n-r}$. Before we derive the priors for $\mathfrak{p}$ we briefly comment on the relationship between priors for $\overline{\beta}_2$ and $\mathfrak{p}$. First we must introduce some notation for matrix spaces and measures on these spaces. For an introduction to these concepts see Muirehead (1982) and for a more intuitive discussion see Strachan and Inder (2004).

The $r \times r$ orthogonal matrix C is an element of the orthogonal group of $r \times r$ orthogonal matrices denoted by $O\,(r) = \{C\,(r \times r) : C'C = I_r\}$, that is

[2]The authors would like to thank Soren Johansen for making this point to one of the author's while visiting the EUI in Florence in 1998. Villani (2000, 2004) also makes use of a prior on $\mathfrak{p}$

$C \in O(r)$. The $n \times r$ semi-orthogonal matrix V is an element of the Stiefel manifold denoted by $V_{r,n} = \{V(n \times r) : V'V = I_r\}$, that is $V \in V_{r,n}$. As the vectors of any V are linearly independent (since they are orthogonal) the columns of V define a plane, $\mathfrak{p}$, which is an element of the $(n-r)r$ dimensional Grassman manifold, that is $\mathfrak{p} = sp(V) \in G_{r,n-r}$. That is, all of the vectors in V will lie in only one $r-$ dimensional plane, $\mathfrak{p}$. The cointegrating space for an n dimensional system with cointegrating rank r is an example of an element of $G_{r,n-r}$. Finally, let the j^{th} largest eigenvalue of the matrix A be denoted $\lambda_j(A)$.

As discussed in James (1954), the invariant measures on the orthogonal group, the Stiefel manifold and the Grassman manifold are defined in exterior product differential forms (for measures on the orthogonal group and the Stiefel manifold, see also Muirhead 1982, Ch. 2). For brevity we denote these measures as follows. For a $(n \times n)$ orthogonal matrix $[b_1, b_2, \ldots, b_n] \in O(n)$ where b_i is a unit n-vector such that $\beta = [b_1, b_2, \ldots, b_r] \in V_{r,n}$, $r < n$, the measure on the orthogonal group $O(n)$ is denoted $dv_n^n \equiv \Lambda_{i=1}^n \Lambda_{j=i+1}^n b_j' db_i$, the measure on the Stiefel manifold $V_{r,n}$ is denoted $dv_r^n \equiv \Lambda_{i=1}^r \Lambda_{j=i+1}^n b_j' db_i$, and the the measure on the Grassman manifold $G_{r,n-r}$ is denoted $dg_r^n \equiv \Lambda_{i=1}^r \Lambda_{j=n-r+1}^n b_j' db_i$. These measures are invariant (to left and right orthogonal translations).The underscore denotes the normalised measure such that $\int_{G_{r,n-r}} \underline{dg_r^n} = 1$.

We begin with a discussion of the relationship between the cointegrating vectors identified using the linear restrictions, $\overline{\beta}_2$, and the cointegrating space as the object of interest, $\mathfrak{p}$.

Theorem 1 *The Jacobian for the transformation from* $\mathfrak{p} \in G_{r,n-r}$ *to* $vec\left(\overline{\beta}_2\right) \in R^{(n-r)r}$ *is defined by*

$$\underline{dg_r^n} = \pi^{-(n-r)r} \Pi_{j=1}^r \frac{\Gamma\left[(n+1-j)/2\right]}{\Gamma\left[(r+1-j)/2\right]} \left| I_r + \overline{\beta}_2' \overline{\beta}_2 \right|^{-n/2} \left(d\overline{\beta}_2\right) \tag{5}$$

where $\Gamma(q) = \int_0^\infty u^{q-1} e^{-u} du$ *for* $q > 0$.

Proof. See Appendix II.

Thus while a uniform distribution on $G_{r,n-r}$ implies a uniform distribution on $V_{r,n}$, this uniform distribution on $G_{r,n-r}$ implies a Cauchy distribution for $\overline{\beta}_2$. This last result was also derived by Phillips (1989).[3]

[3]This transformation of the measure is relevant in both Bayesian and classical appli-

Generally, estimating the cointegrating space using linear identifying restrictions will result in Cauchy tail behaviour unless there are other terms - such as prior information - offsetting the effect of this transformation. As one example of this effect of prior information, Bauwens and Lubrano (1996) show that overidentifying restrictions - which therefore reduce the number of free parameters to be estimated and, importantly, restrict the range of $\mathfrak{p}$ within $G_{r,n-r}$ - will result in a posterior with as many moments as overidentifying restrictions.

A common justification for the linear restrictions is that an economist will usually have some idea about which variables will enter the cointegrating relations and so she chooses c to select the rows of coefficients most likely to be nonzero - more generally linearly independent from eachother - and then normalise on these coefficients. This is a necessary assumption to ensure $(c\beta)^{-1}$ in $\overline{\beta}_2 = c_{\perp}\beta\,(c\beta)^{-1}$ exists. Using these linear restrictions, however, has the unexpected and undesirable result that the Jacobian for $\overline{\beta}_2 \rightarrow \mathfrak{p}$ places more weight in the direction where the coefficients thought most likely to be different from zero are, in fact, zero (or linearly dependent). In fact, normalisation of β by choice of c with a flat prior on $\overline{\beta}_2$ implies infinite prior odds against this normalisation (see Theorems 2 and 3 in Appendix II for a proof of this statement and simple demonstration). In other words, assuming $(c\beta)^{-1}$ exists and then using this assumption makes this assumption *a priori* impossible.

A uniform prior on the cointegrating space: We wish to avoid the problems outlined above deriving from the use of linear restrictions with normalisation to identify the elements of β and the subsequent treatment of $\overline{\beta}_2$ as the parameter of interest. Our recommendation is, if the economist wishes to incorporate prior beliefs about the cointegrating relations, these

cations. As discussed in Phillips (1994), the form in (13) which introduces Cauchy tails into the distribution for $\overline{\beta}_2$ explains why applying linear restrictions to the maximum likelihood estimator of Johansen, $\widehat{\beta} = \begin{bmatrix}\widehat{\beta}_1' & \widehat{\beta}_2'\end{bmatrix}'$ results in an estimator, $\widehat{\overline{\beta}} = \widehat{\beta}_2\widehat{\beta}_1^{-1}$, which is occasionally unreliable. The finite sample distribution for $\widehat{\overline{\beta}}$ has Cauchy tails and this Cauchy behaviour is a direct result of imposing the linear restrictions. This form also provides an alternative explanation for the rather similar but Bayesian results of Bauwens and Lubrano (1996). They show posterior Cauchy tail behaviour of the Bayesian estimator of $\overline{\beta} = \beta_2\beta_1^{-1}$ where no (additional) prior information on the cointegrating space is employed, although they use a 1-1 poly-t argument to find this result. Similar results can be found for the simultaneous equations model in Kleibergen and van Dijk (1998) and Drèze (1976).

should be expressed in the prior distribution for the cointegrating space.

As we have claimed the cointegrating space to be the parameter of interest, rather than $\overline{\beta}_2$, we propose working directly with $\mathfrak{p} = sp(\beta)$ avoiding the linear restrictions and normalisation. Initially we present a distribution and identifying restrictions for β from the form of the uniform distribution for $\mathfrak{p}$ over $G_{r,n-r}$ using the results of James (1954) (see also Strachan and Inder, 2004). The identifying restrictions on β follow naturally from this approach. This prior has the form

$$p(\beta) = \frac{1}{\int_{G_{r,n-r}} dg_r^n} \tag{6}$$

where β is the r-frame with fixed orientation in $\mathfrak{p}$. In the proof of Theorem 1 in Appendix II, the measure on $G_{r,n-r}$ used in the above expression is derived from its relationship with the spaces $V_{r,n}$ and $O(r)$.

To avoid using linear restrictions with a normalisation to identify β it is necessary to find an alternative set of restrictions that do not require knowledge of c and which avoid the issues associated with the posterior for $\overline{\beta}_2$. Fortunately the definition (11) and the discussion in the proof of Theorem 1 provide a natural solution to this question. That is use $\beta \in V_{r,n_i}$ which implies $r(r+1)/2$ restrictions. The dimension of the Grassman manifold is only $(n-r)r$ while the dimension of the Stiefel manifold $V_{r,n}$ is $nr - r(r+1)/2$, which exceeds that of $G_{r,n-r}$ by $r(r-1)/2$. In (11), these remaining restrictions come from the orientation of β in $\mathfrak{p}$ by $C \in O(r)$. The prior, the posterior (as is made clear later) and the differential form for β are all invariant to translations of the form $\beta \to \beta H$, $H \in O(r)$. Therefore it is possible to work directly with β as an element of the Stiefel manifold and adjust the integrals with respect to β by $\left(\int_{O(r)} dv_r^r\right)^{-1}$. Note that these identifying restrictions do not distort the weight on the space of the parameter of interest, $\mathfrak{p}$, and it is never necessary to actually specify the orientation of β in $\mathfrak{p}$.

Thus, contrary to the situation when using linear identifying restrictions, we are able to employ innocuous identifying restrictions, place a prior directly on the parameter of interest and, as we show below, we achieve a better behaved posterior about which we know much more. Before we discuss the posterior, however, we note that in Appendix III extends this approach to informative distributions on the cointegrating space.

3.3 The posteriors.

Using the priors specified above and the likelihood in (4), the general form of the posterior is then

$$\begin{aligned} p\left(B,\Sigma,\beta,r,i|y\right) &\propto p\left(\beta\right)\ \left|\Sigma\right|^{-(T+n+k_i+r+1)/2} \\ &\quad\times \exp\left\{-\frac{1}{2}tr\Sigma^{-1}\left[TS+\left(B-\widetilde{B}\right)'V\left(B-\widetilde{B}\right)\right]\right\} \\ &\quad\times\left(2\pi\right)^{-n(k_i+r)/2}\tau^{-n(k_i+r)/2} \\ &= k\left(B,\Sigma,\beta,\omega|y\right). \end{aligned} \tag{7}$$

We refer the reader to Appendix IV for definition of the terms in the above and following expressions.

For $B\in R^{(k_i+r)n}$ and Σ positive definite (denoted $\Sigma>0$), to estimate the relevant Bayes factors, $B_{jl}=\frac{m_j}{m_l}$, for the models of interest, estimates of the marginal likelihoods, e.g.

$$m_j=\sum_{\omega}\int_{R^{(k_i+r)n}}\int_{\Sigma>0}\int_{G_{r,n-r}}k_{\phi}\left(B,\Sigma,\beta,\omega|y\right)\left(dg_r^n\right)\left(d\Sigma\right)\left(dB\right), \tag{8}$$

are required. To perform the integration in (8) of $\phi=(\Sigma,B,\beta)$, we first analytically integrate (7) with respect to (Σ,B) as these parameters have conditional posteriors of standard form. This integration gives us the following.

Theorem 2 *The marginal posterior for* (β,ω) *is*

$$p\left(\beta,\omega|y\right)\propto\mathfrak{g}_{\omega}\left|S_{00}\right|^{-T/2}\left|M_{22}\right|^{-n/2}\left|\beta'D_0\beta\right|^{-T/2}\left|\beta'D_1\beta\right|^{(T-n)/2}p\left(\beta\right) \tag{9}$$

where in this case $\mathfrak{g}_{\omega}=T^{-nr/2}\pi^{-(n_i-r)r/2}\tau^{-n(k_i+r)/2}$.

Proof. See, for example, Zellner (1971) or Bauwens and van Dijk (1990).■

Remark: From the expression (9) that we see that not only is dg_r^n invariant to $\beta\rightarrow\beta C$ for $C\in O\left(r\right)$, but so is $k\left(\beta\right)$ and thus the posterior.

Next we need to integrate (9) with respect to β to obtain the posterior for ω. Here we find one of the advantages of our approach over previous approaches in that for all model specifications we consider, the posterior will be proper and all finite moments of β exist (see Appendix II for proof). The importance of this statement becomes evident when we consider that economic objects of interest to decision-makers are often linear or convex functions of the cointegrating vectors. As discussed in a previous section, with linear identifying restrictions expectations of such objects are not defined unless overidentifying restrictions are imposed or an informative prior is used. Further, this result holds even when exogeneity is imposed - again in contrast to when linear identifying restrictions are used.

To obtain the posterior distribution of $\omega = (r, o, e, i)$, $p(\omega|y)$, it is necessary to integrate (9) with respect to β and so obtain an expression for

$$p(\omega|y) = \int p(\beta, \omega|y) \, dg_r^n. \tag{10}$$

The marginal density of β conditional on ω implied by (9) is not of standard form. Although one may exist, we do not currently know of a simple, general analytical solution for $\mathfrak{c}_\omega = \int_{V_{r,n}} k_\beta(\beta) \, dg_r^n$ and so we estimate $\mathfrak{c}_\omega$.

Two possible approaches to estimating $\mathfrak{c}_\omega$ are either to use Markov Chain Monte Carlo (MCMC) methods or to use deterministic methods to approximate the integral. Kleibergen and van Dijk (1998) develop a MCMC scheme in the simultaneous equations model and Kleibergen and Paap (2002) extend this to the cointegrating error correction model. Bauwens and Lubrano (1996) demonstrate an alternative approach. In each of these applications a method is presented to evaluate integrals using MCMC when β has been identified using linear restrictions rather than those used in this paper. Strachan (2003) demonstrates the MCMC approach when β has been identified using restrictions related to those of the ML estimator of Johansen (1992). An approach commonly used in classical work to approximate integrals over $V_{r,n}$, is to use the Laplace approximation (see Strachan and Inder, 2004) which is computationally much faster than MCMC.

The Laplace approximation is a second order asymptotic approximation to the marginal likelihood. There is an alternative, simpler, first order asymptotic approximation to the marginal likelihood which assumes dominance by the likelihood. That is, we may treat the Bayesian information criteria of Schwarz (1978) (BIC) as an asymptotic approximation to $-T/2$ times the log marginal likelihood, $\mathfrak{c}_\omega$, for each model. Thus we are able to obtain estimates of the posterior probabilities of the models. In the Applications section we estimate the integrals both by MCMC approximation and the BIC approaches.

As we wish to obtain estimates of economic objects of interest averaged across models we need to be able to obtain draws of β from the posterior.

The next subsection outlines an approach to obtaining MCMC draws from the posterior with the uniform prior used in this paper.

Obtaining MCMC draws from the posterior with an uninformative prior on the cointegrating space: The mode of the marginal posterior for β, $\widetilde{\beta}$,is relatively straight forward to obtain (see Strachan and Inder, 2004). This gives us a method of developing a candidate density for the posterior with mass in the same location as the posterior. The technical details on the development and form of this distribution are left to Appendix III. The steps in this approach are:

1) Specify a distribution for τ which includes specifying its standard deviation, σ.

2) Let $H = \widetilde{\beta}$ and construct the orthogonal complement $H_\perp$ such that $H'H_\perp = 0$.

3) Take a draw of τ and construct $P_\tau = HH' + H_\perp H_\perp' \tau$.

4) Draw Z from the multivariate standard normal.

5) Construct $X = P_\tau Z$.

6) Construct β^* from the orthogonal-triangular (QR) decomposition $X = \beta^* \kappa$.

β^* is then a draw from the candidate density for $\mathfrak{p}$ with location $\mathfrak{p}^H = sp\left(\widetilde{\beta}\right)$ and we iterate over steps 3) to 6) to obtain draws from the candidate distribution.

Acceptance for a Metropolis Hastings scheme or weighting in an Importance sampling method will be determined by a function the ratio of the posterior to the candidate.

4 Empirical Applications

In this section we provide empirical results for three economic issues. The first is relatively simple and involves the issue of stability of Australian money demand. In the second we evaluate the relative weights of permanent and transitory shocks in a US real business cycle model. Here we make use of the classic study of King, Plosser, Stock and Watson (1991). In the third case we evaluate posterior evidence on an inflationary oil price shock and a liquidity trap in a UK macroeconomic model developed by Garratt, Lee, Pesaran en Shin (2002).

4.1 Stability of Australian money demand.

We consider a simple study of the stability of Australian money demand. The variables, all of which appear in logarithmic form, are defined as

$$y_t = (m_t \quad p_t \quad inc_t),$$

where m_t is the measure of M1, p_t is the price level, such that $m_t - p_t$ measures real money, and inc_t is real gross national income. The data are quarterly observations from September 1976 to December 2002 and were sourced from the web site of The Australian Bureau of Statistics, specifically tables D03, G09 and G02.

As is commonly done (see for example Johansen, 1995 and Funke, Hall and Beeby, 1997), we regard the money demand relation as the cointegrating relation among the variables

$$z_t = \beta_1 m_t + \beta_2 p_t + \beta_3 inc_t + \mu_1 + \delta_1 t = y_t \beta$$

and are interested in the stability of this relation and whether this relation is the velocity of money. This implies a secondary question of interest is whether the velocity of money is stable. From the money demand relation above we can see that stability of the (negative log) velocity of money, $\nu = m_t - p_t - inc_t$, is implied by the identifying restrictions $\beta_1 = -\beta_2 = -\beta_3$. Figure 1 shows a plot of ν over the sample period and we see immediately that there is a clear break in the behaviour of ν around 1990. There is some uncertainty about the date of this break so we will simply denote it as $\underline{\tau}$. This graphical evidence could lead us to believe that ν is not a cointegrating relation and therefore is not stable (in the sense of being $I(0)$). An alternative explanation is that there was a change in the equilibrium relationship among these variables and that the velocity is still $I(0)$ but not necessarily stationary. This could occur if there were a change in the mean of z_t, $E(z_t) = \mu_1$, for example, or a trend were introduced into z_t, $E(z_t) = \mu_1 + \delta_1 t$, after the date $\underline{\tau}$. Either of these could reflect the financial deregulation and advances in finance technology that began in the 1980s and either consolidated in the 1990s (mean shift) or a continuing expansion in the real money supply (trend). It is also possible that the cointegrating relation is not ν but that there is still a break in that relation.

Thus we allow for a break in the cointegrating relation such that $\beta = \beta_{\underline{\tau}}$ for $t \leq \underline{\tau}$ and $\beta = \beta_{\overline{\tau}}$ for $t > \underline{\tau}$. We also allow for the possbility that ν is $I(0)$

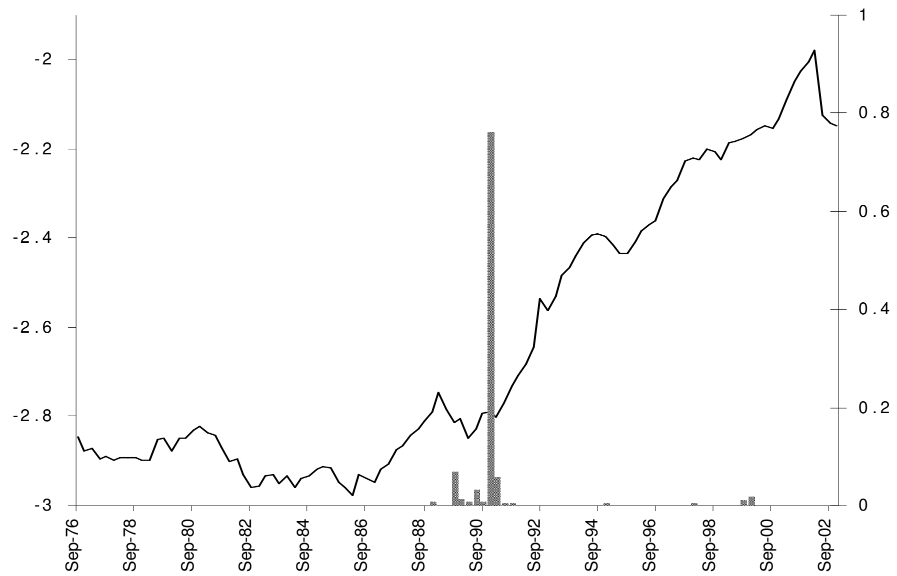

Figure 1: Plot of the velocity of money (ν) as the solid line (against the left hand axis), and the probability of a break in the money demand relation at each date (vertical bars against the right hand axis).

which implies a restriction on the cointegrating space. As we would also like to capture any possible change in the deterministic behaviour of the money demand relation, we allow the deterministic process to differ pre and post τ. One final structural feature we allow is that income is weakly exogenous with respect to the long run relations for which we find mixed support.

With no break in the cointegrating relation, weak exogeneity of income has a posterior probability of zero, while with a break there is a 22.5% posterior probability that income is weakly exogenous. Not allowing for a break in the behaviour of the money demand relation, the probability that ν is $I(0)$ is zero and there is a 53% probability that there is a linear deterministic drift in Δy_t and a nonzero mean in $y_t\beta^+$ and a 44% probability that there is a trend in $y_t\beta^+$. The probability of the break occurring at each date is plotted against ν as the vertical bars in Figure 1 and against each of the variables in the system in Figure 2. We see from Figure 1 that the probabilities of a break seem to lag the actual break in the behaviour of the velocity of money. However, as the unconditional probability that ν is $I(0)$ is reasonably low at

22.16%, it may not be appropriate to consider the stability of this relation when our concern is the stability of the money demand relation.

We can, however, say something about the deterministic behaviour of the money demand relation before and after the break. That is, there is a 42.6% posterior probability that there was a trend in the relation before and after any break date, and a 21.9% posterior probability that there was no trend prior to the break and there was a trend after the break. If we consider the evidence for a break at each date against the plots in Figure 2, we see that the probabilities coincide well with changes in the behaviours of prices and income. At or around the most probably break date, the rate of inflation slowed and the rate of income growth first became negative and then accelerated to a faster rate than prior to the break. Noteable events at this time were the recession at the begining of the 1990s and the completion of a period of financial deregulation over the 1980s which is evidenced in the behaviour of the velocity of money.

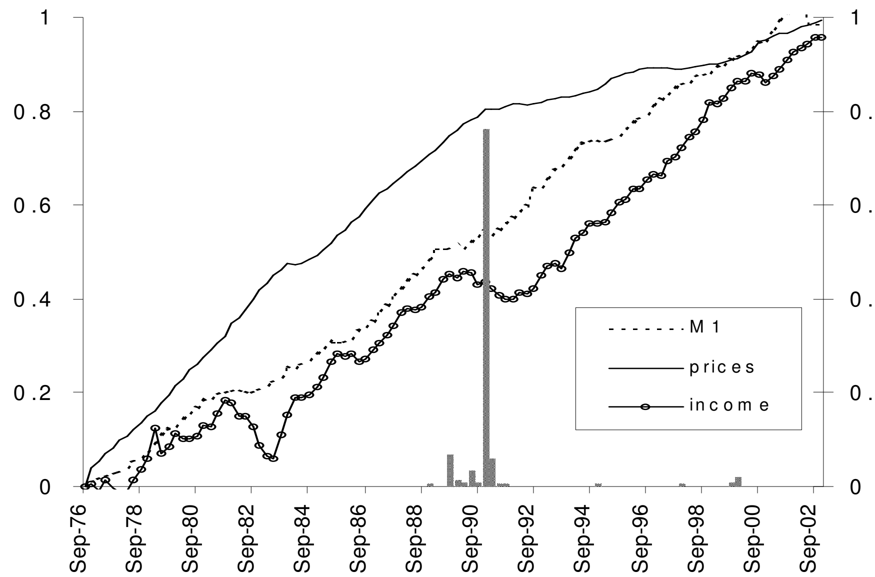

Figure 2: Plot of the probability of a break in the money demand relation as vertical bars (right hand axis) and each of M1, prices and income (left hand axis).

This is not a trivial modelling exercise. The results were averaged over

80 models that assumed no structural break, and for each date at which we allowed a break we considered 288 models. As we allow for a break at any date between March 1978 ($t = 2$) and September 2002 ($t = T - 1$), we considered in total some 28,016 models. With a burn-in of 5000 draws and 10000 draws from the posterior for each model, the computation of the probabilities on a desk top PC took 90 minutes.

4.2 Relative weights of permanent and transitory components in a US real business cycle model.

In this study we investigate the Bayesian evidence for the presence of a common stochastic trend in real US consumption (c_t), investment (i_t) and income (inc_t) as implied by the Real Business Cycle model (RBC) studied in King, Plosser, Stock and Watson (1991). We extend this study to a decomposition of the variance of the three variables in the system $y_t = (c_t, i_t, inc_t)$ into that part due to permanent shocks and that part due to transitory shocks. We use the approach of Cubadda and Hecq (2001) for achieving this decomposition in the time domain. The data are quarterly covering the period from the first quarter 1951 to the final quarter of 1999 and come from the study by Paap and van Dijk (2003).

King, *et al.* investigate the support for a feature of the process that is implied by the RBC: the number of stochastic trends, $n - r$. We therefore allow the rank, r, to vary over all possible values, $r \in [0, 1, \ldots, n]$. The other important feature of the RBC is that the Great Ratios of consumption on income and investment on income should be stationary. We therefore allow the log differences $c_t - inc_t$ and $i_t - inc_t$ to either form the cointegrating relations (if $r = 2$) or the variables will enter the cointegrating relations via these relations (if $r = 1$). This implies we allow the cointegrating vectors to be overidentified as $\beta = H\varphi$ where the $2 \times r$ matrix φ is semiorthogonal such that $\varphi'\varphi = 1$ and

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}.$$

Finally we also allow for the range of five combinations of deterministic processes suggested in Subsection 2.1.4. Thus the set of 320 models may be summarised as $r \in [0, 1, 2, 3]$, $(i, j) \in [(0, 0), \ldots, (2, 2)]$, $l \in [1, \ldots, 4]$ and $o \in [0, 1]$.

Beginning with the support for the alternative models in the model set, the modal model with posterior probability of 41.4%, has one stochastic trend ($r = 2$) and the great ratios form the cointegrating relations, $\beta = H$. The posterior probabilities of the models are given in Table 1. These results show that without the overidentifying restrictions, the weight of support is upon there being three stochastic trends in y_t, with some support for there being no stochastic trends. With the overidentifying restrictions imposed, we find that the support shifts to a single stochastic trend. This gives an indication of the dangers of a sequential testing procedure for these restrictions. While the marginal probability of the overidentification restrictions is only 44.9%, 49.5% of the posterior mass is on models in which these restrictions are either impossible ($r = n$) or meaningless ($r = 0$). These results give a clear example of where there is a high degree of uncertainty as to what is the appropriate or 'best' model, and averaging offers an attractive approach to accounting for this uncertainty in subsequent inference.

Table 1: Posterior probabilities of structural features for real business cycle model. Note that the cells for observationally equivalent models have been merged and their probabilities added together.

Just Identified Models ($o = 0$)

r	$(i,j) = (0,0)$	$(i,j) = (1,0)$	$(i,j) = (1,1)$	$(i,j) = (2,1)$	$(i,j) = (2,2)$
0	0.000	0.000		0.338	
1	0.000	0.001	0.023	0.028	0.000
2	0.000	0.000	0.002	0.000	0.000
3	0.000		0.156		0.001
Over Identified Models ($o = 1$)					
1	0.000	0.001	0.007	0.012	0.001
2	0.000	0.000	0.414	0.008	0.006

Decomposing the impulse response into the transitory and permanent shocks we gain an impression of the importance of the effects for the variability of the consumption, investment and income. Table 2 shows the variance decomposition into the two components in three separate cases where h denotes the number of quarters after the shock. The first part of the table, part A, gives the results averaged over all of the models considered. In this case the variance is shared almost equally between the two sources of shocks and there is little change over the horizons considered. Part B shows the results for the best (highest posterior probability) model. Here we see that for all three variables the transitory shock accounts for most of the variability and, for consumption and income, this effect increases slightly from a horizon of one to 36. A similar pattern is found when we only consider models with cointegration but no trend in the error correction term and no linear deterministic drift. These three cases demonstrate the implications of conditioning upon individual models or even subsets of models. This is not to suggest that this conditioning is necessarily inappropriate, but it does have implications for the conclusions we obtain.

Table 2: Estimated variance decompositions into permanent and transitory components in the time domain.

A: Averaged over all models						
	Permanent			Transitory		
h	c_t	i_t	inc_t	c_t	i_t	inc_t
1	0.543	0.496	0.498	0.454	0.501	0.499
4	0.510	0.496	0.495	0.487	0.501	0.502
36	0.499	0.496	0.492	0.498	0.502	0.505

B: From the best model $(ij = (1,1), r = 2, o = 1)$						
	Permanent			Transitory		
h	c_t	i_t	inc_t	c_t	i_t	inc_t
1	0.374	0.249	0.259	0.626	0.751	0.741
4	0.294	0.249	0.253	0.706	0.751	0.748
36	0.272	0.249	0.248	0.728	0.751	0.752

C: Averaged over all models with $(i > 0, j > 0) \cap (0 < r < n)$						
	Permanent			Transitory		
h	c_t	i_t	inc_t	c_t	i_t	inc_t
1	0.407	0.313	0.317	0.593	0.687	0.683
4	0.342	0.313	0.311	0.658	0.687	0.689
36	0.319	0.312	0.306	0.681	0.688	0.694

4.3 Evidence on inflationary oil price shock and liquidity trap in the UK.

Garratt, Lee, Pesaran, and Shin (2003) provide an extensive model of the UK economy which focuses upon the long run relations, but incorporates useful short run restrictions to improve modelling. In their paper, Garratt *et al.* highlight two differences in their approach from other large models. First it is developed for a small open economy, and second it takes a new and practical approach to incorporating long run relations while leaving short run relations largely unrestricted. The variables in the econometric model are

$$y_t = (r_t, w_t, \Delta p_t, p_t - p_t^*, e_t, h_t - w_t, r_t^*, w_t^*, p_t^o),$$

where, in logarithms, p_t^o is the price of oil, w_t is UK real per capita GDP and w_t^* is the foreign (OECD) real per capita GDP, p_t is the UK producer price index, p_t^* is foreign (OECD) producer prices, e_t is the nominal Sterling effective exchange rate, h_t UK real per capita M0 money stock, $r_t = 0.25 \ln(1 + R_t/100)$ where R_t is a function of 90 day interest rates and r_t^* a similar function of the US, Germany, Japan and France 90 day rates.

The long run relations which form the cointegrating relations, subject to all restrictions finally imposed as a result of the analysis by Garratt, *et al.*, are

$$\begin{aligned} p_t - p_t^* - e_t &= u_{1,t} \\ r_t - r_t^* &= u_{2,t} \\ w_t - w_t^* &= u_{3,t} \\ r_t - \Delta p_t &= u_{4,t} \text{ and} \\ \beta_{32}(h_t - w_t) &= \mu_{11,\delta} t + \beta_{22} r_t + u_{5,t} \end{aligned}$$

where the $u_{i,t}$ are $I(0)$ with unrestricted means. Assuming the rank $r = 5$, these results suggest a cointegrating space spanning the space of the matrix $\beta = (H_1 \quad \beta_2^c)$ where

$$H_1' = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\begin{aligned} \beta_2^c &= H_2 \varphi \\ H_2' &= \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ \varphi' &= \begin{pmatrix} \mu_{11,\delta} & \beta_{22} & \beta_{32} \end{pmatrix}. \end{aligned}$$

There are three parameters[4] to be estimated in β. In their paper, Garratt *et al.* make oil prices strictly exogenous with respect to the rest of the system[5]. The parameterisation they use implies weak exogeneity of oil prices with respect to α and β. The restriction that there is no quadratic trend in y_t implies $\mu_{11,\delta}\alpha = 0$. Further, the exclusion of a trend from all long run relations except the money-income relation, $u_{5,t}$, implies the restriction upon the first row of β is $\mu_{1,\delta} = \left(0, 0, 0, 0, \mu_{11,\delta}\right)$.

The combinations of restrictions implied by the above model can be denoted in the notation of Section 2 as M_ω with $\omega = (5, 2, 1, 1, 0)$, that is, the cointegrating rank is 5, we employ the overidentifying restrictions on β of type 2, oil prices are weakly exogenous with respect to α and β, and there is no quadratic drift in y_t but there may be a trend in $y_t\beta$ $(ij = (1, 0))$

The range of models we include in our model set are defined by $r \in [0, 1, \ldots, 9]$, $e \in [0, 1]$, $o \in [0, 2]$, and $(i, j) \in [(0, 0), \ldots, (2, 2)]$, for a total of 200 models. As a number of the models implied by combinations of these restrictions are either impossible or observationally equivalent, we need only estimate 87 models. Analysing their macroeconomic model within (1), Garratt, *et al.* find support for $r = 5$ using Johansen's trace test. They also find support for the overidentifying restrictions and trend restrictions using a log-likelihood ratio test, where they used bootstrap estimates of critical values. They do not appear to test support for the weak exogeneity of oil prices. Below we present the posterior probabilities of the various models (zeros or near zeros are suppressed or omitted) where $e = 1$ implies weak exogeneity of oil prices.

[4] Note that we do not use linear identifying restrictions (or normalisation) for the vector β_2^c, in which coefficients must be estimated. Instead, as discussed below, we identify φ by nonlinear restrictions of the form $\varphi'\varphi = 1$. We do this to simplify estimation, and to avoid the potential problem that the posterior may have no moments and possibly be improper, particularly when we impose exogeneity.

[5] The concept of strict exogeneity has been criticised (Engle, Hendry and Richard 1983 and Hendry 1995) for introducing ambiguity of interpretation. The concepts of weak, strong and super exogeneity do, however, have clear interpretations and implications. Therefore, it is fortunate that in making oil prices strictly exogenous, Garratt *et al.* in fact make them weakly exogenous with respect to β^+ and α. The weak exogeneity of oil prices implies $\alpha_2 = 0$.

	$\widehat{p}(r,i,e=0\|y)$				
r	$i,j=0,0$	$i,j=1,0$	$i,j=1,1$	$i,j=2,1$	$i,j=2,2$
0					
1		0.0101	0.0008	0.0324	0.0128
2		0.0004	0.0001	0.0040	0.0190

	$\widehat{p}(r,i,e=1\|y)$				
r	$i,j=0,0$	$i,j=1,0$	$i,j=1,1$	$i,j=2,1$	$i,j=2,2$
0					
1		0.0213	0.0085	0.2507	0.1262
2		0.0001	0.0002	0.0756	0.4376
3					0.0001

Marginal Probabilities

$\widehat{p}(i\|y)$	$i,j=0,0$	$i,j=1,0$	$i,j=1,1$	$i,j=2,1$	$i,j=2,2$
	0.0000	0.0320	0.0096	0.3626	0.5958

$\widehat{p}(r\|y)$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4$
		0.4628	0.5371	0.0001	

The posterior probabilities for the rank suggest support for a rank of one or two with $P(r=1|y)=0.4628$ and $P(r=2|y)=0.5371$. We also find marginal probabilities of no deterministic processes $(i,j=2,2)$ of 0.5958 and of an intercept in the cointegrating relations $(i,j=2,1)$ of 0.3626. The posterior probability that the oil prices are weakly exogenous is 0.9203 providing strong support for this restriction. The combined restrictions of overidentification, exogeneity, four stochastic trends and a linear trend in the long run money-income relation had a joint probability of effectively zero within this model set.

With the overidentifying restrictions, the only coefficients to be estimated in the long run relations, ignoring the intercepts, are in the money market equilibrium condition given by

$$h_t - w_t = \mu_{21,\tau} t + \beta_{22} r_t + u_{4,t}.$$

Estimating the coefficients in this relation subject to the restrictions proposed by Garratt, *et al.*, we obtain

$$h_t - w_t = -0.0070t - 43.2148 r_t + u_{4,t}$$

which compares with the classical estimate of Garratt *et al.* of

$$h_t - w_t = -0.0073t - 56.0975 r_t + u_{4,t}.$$

Both results suggest a downward trend in the money-income ratio which may be attributed to technological innovations in the finance sector (Garratt, *et al. 2003*).

Although there is a clear modal model, $M_{(r,o,e,i,j)} = M_{(2,1,2,2,2)}$, there is just as clearly some support for nearby models such as $M_{(1,1,2,2,1)}$ and $M_{(1,1,2,2,2)}$. We incorporate the information value of these models for decision making through averaging the economic object of interest. As an example of an averaged output which can be used as an input for decision making, Figure 3 presents the higher posterior density regions (hpds) for the impulse response function over 60 months for a response in relative UK prices, $p_t - p_t^*$, to a shock in oil prices, p_t^o. This output is averaged across all models and was produced from 100,000 draws from the full posterior. The intervals plot the boundaries of the 20%, 40%, 60% and 80% hpds. The UK during the period of the sample was a net oil exporter and we see the effect of this reflected in the figure as the distribution of the response path indicates initially that the rest of the world experiences a larger response to an oil price shock than the UK, after which the UK appears to catch up slightly. However, the greater impact on world prices relative to UK prices seems to persist as after 60 months the path is centred around a slightly negative mode just above negative 1%. This is not a surprising result given the likely exchange rate adjustment in the pound.

It should be pointed out that these intervals are not comparable with the usual classical confidence intervals as, in addition to variable uncertainty, they incorporate parameter uncertainty and model uncertainty. With this extra uncertainty it is sensible then that the intervals containing a given mass will be wider and the mass in any particular region does not have the same interpretation. Trimming the model set of unreasonable models would likely produce smaller intervals. However, the results we present are more informative on the question 'What will happen to relative prices in the UK if there is an oil price shock?' as they do not require the addendum: '... if this model and these parameter values are correct?'.

Figure 4 plots the hpds for the impulse response function over 60 months for a response in UK inflation, Δp_t, to a shock in oil prices, p_t^o, again produced from 100,000 draws from the posterior. The median response after 60 months shows a moderate increase in the level of inflation of around 2.5% and so the median impulse response is about where we would expect it and the 20% and 40% hpds are reasonable.

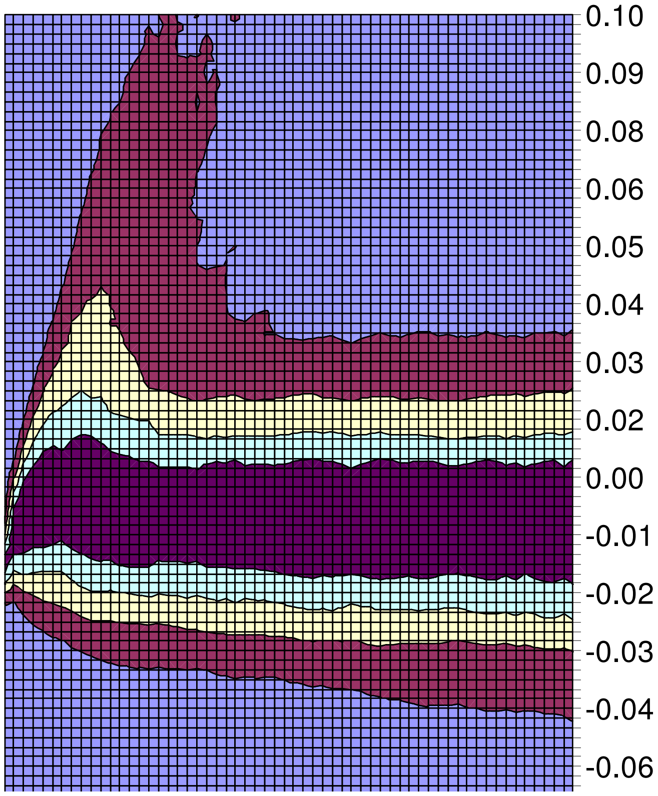

Figure 3: Higher posterior density regions for the impulse response of relative UK prices $(p_t - p_t^*)$ to a shock in oil prices. The x-axis spans zero to sixty months.

An interesting feature of both figures are the long tails at low lags. This tail behaviour is due entirely to the set of 40 models (out of 97 models) in which oil prices are not constrained to be weakly exogenous. Although these models are given a small (but not negligible) posterior probability (around 8%), their implied response paths are so extreme that they have a noticeable influence upon the marginal distribution of the response.

It is to demonstrate this rather strange behaviour that we have reported the results using the BIC approximation to the posterior probabilities. The same plots of the hpds for the impulse response paths when we used the Laplace approximation or the MCMC estimation do not demonstrate such an extreme diversion in the tail and look similar to what we obtain if we use BIC but exclude the models in which oil prices are not exogenous ($e = 0$). The

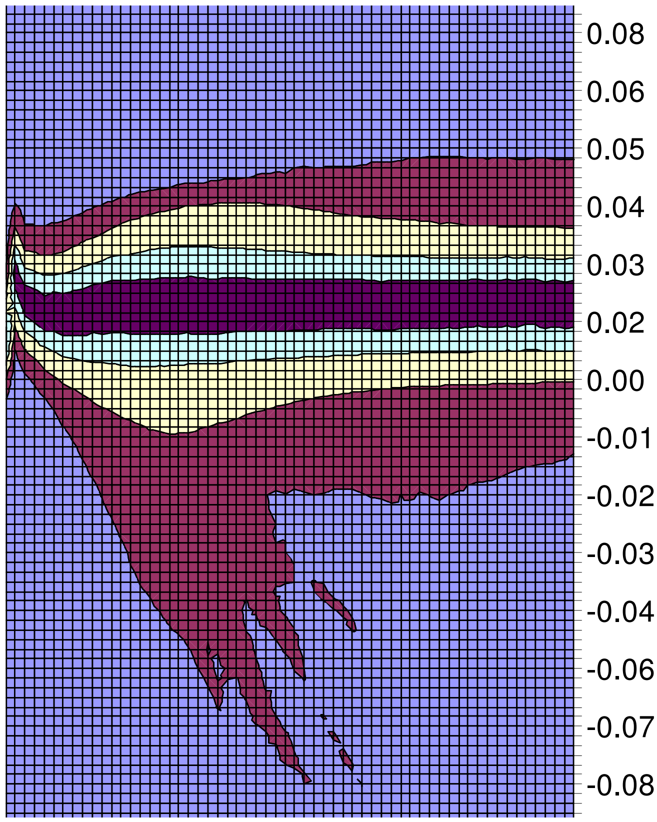

Figure 4: Higher posterior density regions for the impulse response of relative UK inflation (Δp_t) to a shock in oil prices. The x-axis spans zero to sixty months.

reason for this is that the Laplace and MCMC methods tend to concentrate the mass of the density for the models on fewer models and attribute no mass to the models with $e = 0$. The behaviour in Figure 4 demonstrates the risks of conditioning on particular models, but also the risks - also inherent in our approach - of not using a sufficiently well considered model set.

The final output from this analysis is the forecasts of the probability that the UK would encounter both negative prices changes and negative growth from the year 2000 on, after the end of the sample (the near future with respect to the sample). The occurance of these two events would lessen the effectiveness of interest rate cuts (as they cannot go negative) thus removing an important policy tool from the central bank. Although more technical definitions exist, for conciseness we will term an occurance of both negative inflation ($\Delta p_{t+h} < 0$) and negative growth ($\Delta w_{t+h} < 0$) in one quarter as an occurance of the the liquidity trap. To obtain the predicted probability of this joint event, we must forecast of the joint predictive density for income and inflation and estimate the mass of the density in the region where both are negative. Figure 5 plots the observed growth (dashed line) and inflation (solid line) over the final eight quarters of the sample. While growth was positive in all quarters (but zero in December 1998), there were three quarters in which there was negative inflation. Figure 6 plots the predicted marginal probabilities of negative growth (dashed line) and negative inflation (solid line) and we see the recent behaviour shown in Figure 5 reflected in these probabilities in the immediate future. The behaviour of the data and the signals from predictions that each will go negative give a mixed signal as to the chance of encountering the liquidity trap. Figure 7 shows the probability of encountering the liquidity trap in the near future as well as the product of the marginal probabilities of the two events of negative growth and negative price changes. The difference between these two plots gives a clear indication of the importance of being able to forecast the joint predictive distribution and not just the marginals. The two lines would of course be identical if the two events were independent. One interesting observation is that although the likelihood of the liquidity trap occuring is very low initially, due to the strong growth in the UK at the end of 1999, the probability increases two years later then falls again. The increase in the probability coincides with a period when central banks were beginning to express some slight concern for the possibility of such an event.

5 Conclusion.

In this paper we have presented a Bayesian approach to obtain unconditional inference on structural features of the vector autoregressive model by means of evaluating posterior probabilities of alternative model specifications using a diffuse prior on the structural features of interest. The output produced this way allows forecasts and policy recommendations to be made that are not conditional on a particular model. Thus this model averaging approach provides an alternative to the more commonly used model selection approach. Specifically we provide techniques for estimating marginal likelihoods for models of cointegration, deterministic processes, exogeneity,

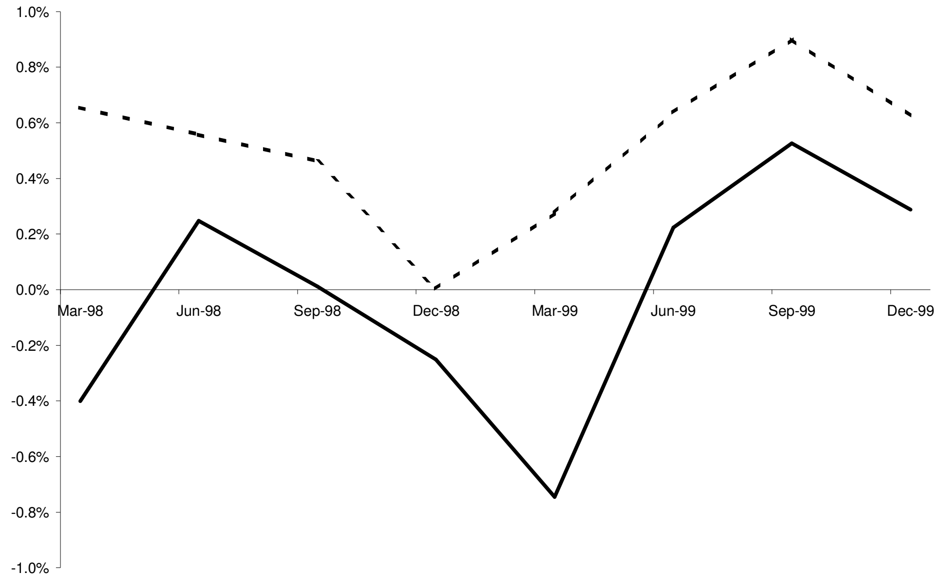

Figure 5: Plot of UK growth (dashed line) and inflation (solid line) over the quarters March 1998 to December 1999.

and overidentifying restrictions upon the cointegrating space. The estimates are derived using a mixture of analytical integration and MCMC or asymptotic approximations to integrals. Three applications of the methodology are provided: stability of Australian money demand, permanent and transitory shocks in the US business cycle and UK evidence on an inflationary oil price shock and a liquidity trap.

We end with mentioning two topics for further research. First, there exists the issue of the robustness of the results with respect to prior and model specification. Very natural extensions of our approach are to include prior inequality conditions in the parameter space of structural VARs and consider forms of nonlinearity and time variation in the model itself. For instance, in using a SVAR for business cycle analysis one may use prior information on the length and amplitude of the period of oscillation. An example of a possible nonlinear time varying structure that may prove useful is presented in Paap and van Dijk (2003). Systematic use of inequality conditions and nonlinearity implies a more intense use of MCMC algorithms. Second, one may use the results of our approach in explicit decision problems in international and

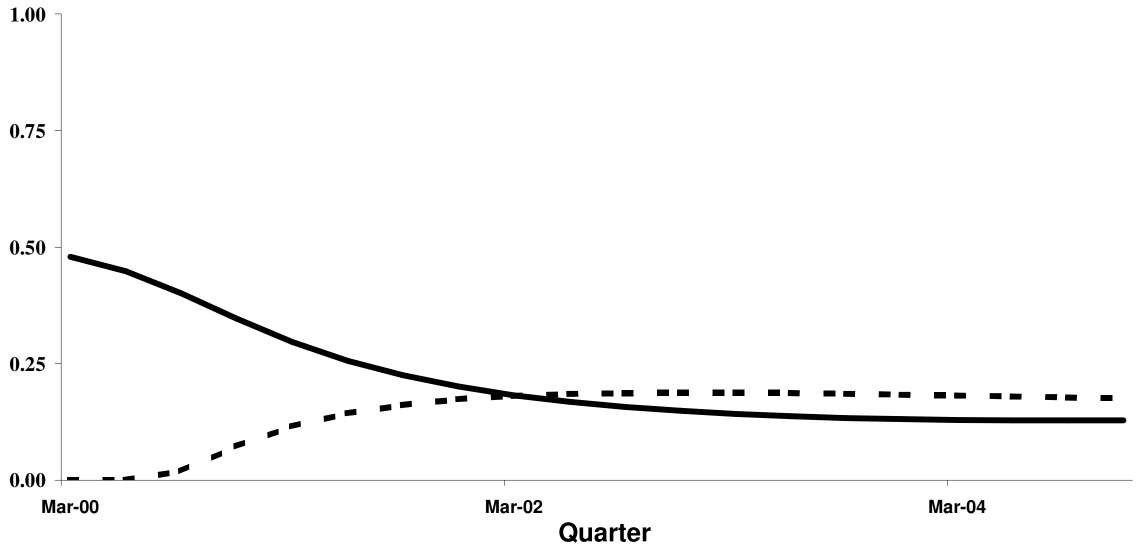

Figure 6: Plot of predicted marginal probabilities of negative growth (dashed line) and negative inflation (solid line).

financial markets like hedging currency risk or evaluation of option prices.

6 Acknowledgements

A very preliminary version of this paper has been presented at the ESEM-2002 meeting in Venice, the EC-2 meeting in Bologna and the Tokyo meeting in 2004 on financial time series. We would like to thank David Hendry, Soren Johansen, Helmut Lutkepohl, and Christopher Sims for helpful discussion on the topic of this paper. We also wish to acknowledge the assistance of funding from the University of Liverpool grant number 4128.

7 References.

Bauwens, L. and M. Lubrano (1996): "Identification Restrictions and Posterior Densities in Cointegrated Gaussian VAR Systems," in: T.B. Fomby, ed., *Advances in Econometrics, Vol. 11B, Bayesian methods applied to time*

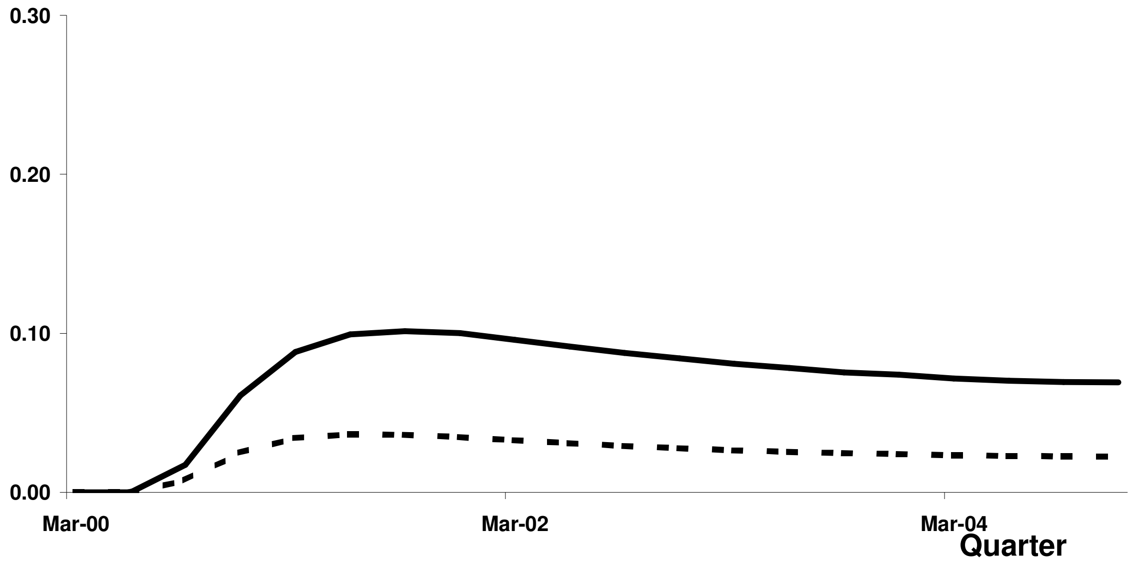

Figure 7: Plot of the predicted probability of the liquidity trap (solid line) and the product of the marginal probabilities of negative growth and negative inflation (dashed line).

series data. JAI Press 3-28.

Bauwens, L. and H.K. van Dijk (1990): "Bayesian limited information analysis revisited," in: J. Gabszewicz, J.F. Richard and L. Wolsey, eds., *Economic decision-making: Games, Econometrics, and Decision-making, Contributions in honour of Jacques Drèze*, North-Holland, Amsterdam, 385-424.

Billingsley, P. (1979): *Probability and Measure.* New York: Wiley.

Boswijk, H.P. (1996): "Testing Identifiability of Cointegrating Vectors," *Journal of Business and Economic Statistics* **14**, 153-160.

Bartlett, M, S. (1957): "A Comment on D. V. Lindley's Statistical Paradox," *Biometrica* **44**, 533–534.

Bates, J. M., & C. W. J. Granger (1969): "The Combination of Forecasts," *Operational Research Quarterly* **20**, 451–468.

Cubadda, G. and A. Hecq (2001): "On non-contemporaneous short-run comovements," *Economics Letters* **73**, 389–397.

Chikuse, Y. (1998): "Density Estimation on the Stiefel Manifold," *Journal of Multivariate Analysis* **66**, 188-206.

Diebold, F.X. and Lopez, J. (1996): "Forecast Evaluation and Combination," in G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics.* Amsterdam: North-Holland, 241-268.

Drèze, J.H. (1978): "Bayesian Regression Analysis using Poly-t Densities," *Journal of Econometrics* **6**, 329-354.

Engle, R.F. and C.W.J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica* **55**, 251-276.

Funk, M., S.G. Hall and M. Beeby (1997): "The Relationship Between Money and Prices: An Economic Appraisal Based on Cointegration and Causality," in: C. Heij, J.M. Schumacher, B. Hanzon and C. Praagman, eds., *System Dynamics in Economic and Financial Models*, New York: Wiley.

Garratt, A, K. Lee, M. H. Pesaran, and Y. Shin (2003): "A Long Run Structural Macroeconomic Model of the UK," forthcoming in *Economic Journal.*

Granger, C.W.J. (1983): "Cointegrated Variables and Error Correction Models," unpublished USCD Discussion Paper 83-13.

Hendry, D. F. & M. P. Clements (2002): "Pooling of Forecasts," *Econometrics Journal* **5**, 1-26.

James, A.T. (1954): "Normal Multivariate Analysis and the Orthogonal Group," *Annals of Mathematical Statistics* **25**, 40-75.

Jeffreys, H. (1961): *Theory of Probability.* London: Oxford University Press.

Johansen, S. (1992): "Cointegration in Partial Systems and the Efficiency of Single Equation Analysis," *Econometrica* **52**, 389-402.

Johansen, S. (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson (1991). "Stochastic trends and economic fluctuations." *The American Economic Review*, 81, 819-840.

Kleibergen, F. and R. Paap (2002): "Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration," *Journal of Econometrics* **111**, 223-249.

Kleibergen, F. and H.K. van Dijk (1994): "On the Shape of the Likelihood/Posterior in Cointegration Models," *Econometric Theory* **10**, 514-551.

Kleibergen, F. and H.K. van Dijk (1998): "Bayesian Simultaneous Equations Analysis Using Reduced Rank Structures," *Econometric Theory* **14**, 701-743.

Lindley, D. V. (1957): "A Statistical Paradox," *Biometrica* **44**, 187-192.

Luukkonen, R., A. Ripatti and P. Saikkonen (1999): "Testing for a Valid Normalisation of Cointegrating Vectors in Vector Autoregressive Processes," *Journal of Business and Economic Statistics* **17**, 195-204.

Muirhead, R.J. (1982): *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Newbold P. & D. Harvey (2001): "Tests for Multiple Forecast Encompassing," *Journal of Applied Econometrics* **15,** 471–482.

O'Hagan, A. (1995): "Fractional Bayes Factors for Model Comparison," *Journal of the Royal Statistical Society, Series B (Methodological)* **57, 1**, 99-138.

Paap, R. and H. K. van Dijk (2003): "Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to US Consumption and Income," *Journal of Business and Economic Statistics* **21**, 547-563.

Phillips, P. C. B. (1989): "Spherical matrix distributions and cauchy quotients," *Statistics and Probability Letters* **8**, 51–53.

Phillips, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica* **59**, 283-306.

Phillips, P. C. B. (1994): "Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models," *Econometrica* **62**, 73-93.

Schotman, P.C. and H.K. van Dijk (1991): "A Bayesian analysis of the unit root in real exchange rates," *Journal of Econometrics* **49**, 195-238.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics* **6**:2, 461-464.

Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica* **48**, 1-48.

Strachan, R. (2003): "Valid Bayesian Estimation of the Cointegrating Error Correction Model," *Journal of Business and Economic Statistics* **21**, 185-195.

Strachan, R. W. and B. Inder (2004): "Bayesian Analysis of The Error Correction Model," forthcoming in *Journal of Econometrics*, forthcoming

Strachan, R and H. K. van Dijk (2003): "Bayesian Model Selection for a Sharp Null and a Diffuse Alternative with Econometric Applications," *Oxford Bulletin of Economics and Statistics* **65**, 681-688.

Strachan, R and H. K. van Dijk (2004): "Improper priors with well defined Bayes factors," Econometric Institute Report EI 2004-18, Erasmus University.

Terui, N. and H. K. van Dijk (2002): "Combined forecasts from linear and nonlinear time series models," *International Journal of Forecasting* **18**(3), 421-438.

Urbain, J. P. (1992): "On Weak Exogeneity in Error Correction Models," *Oxford Bulletin of Economics and Statistics* **52**, 187-202.

Villani, M. (2000). *Aspects of Bayesian Cointegration.* PhD thesis, University of Stockholm.

Villani, M. (2004). Bayesian reference analysis of cointegration, forthcoming in *Econometric Theory.*

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

8 Appendix I

8.1 Posterior distribution of β_2 given exogeneity.

In this section we show the marginal likelihoods are not well defined for β_2 when weak exogeneity is imposed. The following results apply for a wide class of priors. To consider weak exogeneity with respect to β, we partition the matrix α as $\alpha = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix}$ such that the exogeneity restriction is implied by $\alpha_2 = 0$ and derive the marginal distribution of (α_2, β). Next we set $\alpha_2 = 0$ in $p(\alpha_2, \beta|y)$. If $\int p(\beta|\alpha_2 = 0, y)(\beta' d\beta) = \infty$, then the posterior does not integrate to a finite constant and Bayes factors are not defined. Thus by demonstrating that the above integral is not finite when linear restrictions are imposed on β, such that $\beta = [\beta_2' \; I_r]'$ and $\beta_2 \in R^{(n-r)r}$, we show the marginal likelihoods are not finite.

The marginal, joint posterior distribution for (α, β) given $r = n_2$, is

$$p(\alpha, \beta | r, y) \propto \left| TS + (\alpha - \widehat{\alpha})' \beta' S_{11} \beta (\alpha - \widehat{\alpha}) \right|^{-(\nu+r)/2}.$$

such that

$$\begin{aligned} p(\alpha, \beta | r, y) &\propto \left| TS + (\alpha - \widehat{\alpha})' \beta' S_{11} \beta (\alpha - \widehat{\alpha}) \right|^{-(\nu+r)/2} \\ &= \left| (\beta' S_{11} \beta)^{-1} + T^{-1} (\alpha - \widehat{\alpha}) S^{-1} (\alpha - \widehat{\alpha})' \right|^{-(\nu+r)/2} \\ &\quad \times |TS|^{-(\nu+r)/2} |\beta' S_{11} \beta|^{-(\nu+r)/2}. \end{aligned}$$

Let σ_{22} denote the last n_2 rows and columns of TS and partition TS as

$$TS = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & \sigma_{22} \end{bmatrix}.$$

Next, denote the $n_2 \times n_2$ matrix made up of the last n_2 rows and columns of S_{00} by $S_{00,22}$, and note that $\sigma_{22} = S_{00,22} - \widehat{\alpha}_2' \beta' S_{11} \beta \widehat{\alpha}_2$. Next, we integrate with respect to α_1. The conditional distribution of $\alpha | \beta$ is

$$\begin{aligned} p(\alpha | \beta, y) &\propto \left| (\beta' S_{11} \beta)^{-1} + T^{-1} (\alpha - \widehat{\alpha}) S^{-1} (\alpha - \widehat{\alpha})' \right|^{-(\nu+r)/2} \\ &= \left| (\beta' S_{11} \beta)^{-1} + g_1(\alpha_2) + g_2(\alpha_1) \right|^{-(\nu+r)/2} \\ &\quad \text{where} \\ g_1(\alpha_2) &= (\alpha_2 - \widehat{\alpha}_2) \sigma_{22}^{-1} (\alpha_2 - \widehat{\alpha}_2)' \\ g_2(a_1) &= \left(\alpha_1 - \widehat{\Delta}\right) \left(S_{11} - S_{12} \sigma_{22}^{-1} S_{21}\right)^{-1} \left(\alpha_1 - \widehat{\Delta}\right)'. \end{aligned}$$

Integrating with respect to α_1 gives us the marginal distribution of (α_2, β) as

$$\begin{aligned} p(\alpha_2, \beta | r, y) &\propto \left| \sigma_{22} + (\alpha_2 - \widehat{\alpha}_2)' \beta' S_{11} \beta (\alpha_2 - \widehat{\alpha}_2) \right|^{-(\nu-n+r+n_2)/2} \\ &\quad \times |\beta' S_{11} \beta|^{-(n-n_2)2} |\sigma_{22}|^{(\nu-n+n_2)/2} |S|^{-\nu/2}. \end{aligned}$$

Since $S_{00,22} = \sigma_{22} + \widehat{\alpha}_2' \beta' S_{11} \beta \widehat{\alpha}_2 = \sigma_{22} + S_{01,2} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10,2}$, then evaluating this expression at $\alpha_2 = 0$ and rearranging we have

$$p(\beta | \alpha_2 = 0, r, y) \propto |\beta' D \beta|^{-\nu/2} |\beta' D_{0,2} \beta|^{(\nu-n+n_2)/2}$$

where $D_{0,2} = S_{11} - S_{10,2}S_{00,22}^{-1}S_{01,2}$, $S_{10,2} = S_{01,2}'$ is the last n_2 rows of S_{10}. If we partition D and $D_{0,2}$ conformably as

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & d \end{bmatrix} \text{ and } D_{0,2} = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \delta \end{bmatrix}$$

use the linear restrictions $\beta = \begin{bmatrix} \beta_2' & I_r \end{bmatrix}'$, then let $d_s = d - D_{21}D_{11}^{-1}D_{12}$, $\delta_s = \delta - \Delta_{21}\Delta_{11}^{-1}\Delta_{12}$, $\widehat{\beta}_2 = D_{11}^{-1}D_{12}$ and $\widetilde{\beta}_2 = \Delta_{11}^{-1}\Delta_{12}$,

$$\begin{aligned} p(\beta_2|\alpha_2 = 0, r, y) \;\; \propto \;\; & \left| d_s + \left(\beta_2 - \widehat{\beta}_2\right)' D_{11} \left(\beta_2 - \widehat{\beta}_2\right)\right|^{-l_0} \\ & \times \left| \delta_s + \left(\beta_2 - \widetilde{\beta}_2\right)' \Delta_{11} \left(\beta_2 - \widetilde{\beta}_2\right)\right|^{l_1}. \end{aligned}$$

Thus we have the 1-1 poly-t form for the posterior of $\beta|\alpha_2 = 0$. As the posterior is integrable only if $\gamma = 2(l_0 - l_1) - (n - r) > 0$. In this case, then, since $n_2 = r$

$$\gamma = 2(l_0 - l_1) - (n - r) = \nu - \nu + n - n_2 - n + r = 0$$

and the posterior is clearly not integrable. Note that is is possible to take $n_2 > r$ provided $n_1 > n - r$. In this case $\gamma = r - n_2 < 0$, again producing an improper posterior.

Taking strong exogeneity with respect to β will result in n_2 being replaced by $k_2 = n_2 + ln$ giving

$$\begin{aligned} 2(l_0 - l_1) - (n - r) &= \nu - \nu + n - k_2 - n + r \\ &= -n_2 - ln + r < 0 \end{aligned}$$

and the posterior is not proper in any situation.

9 Appendix II

9.1 Proof of Theorem 1:

In deriving the invariant measure on the Grassman manifold, James (1954) presents a relationship between an element of the Stiefel manifold, $V \in V_{r,n}$, and element of the Grassman manifold, $\mathfrak{p} = sp(\beta) \in G_{r,n-r}$ where the r-frame $\beta \in V_{r,n}$ and an element of the orthogonal group, $C \in O(r)$. β has a particular (fixed) orientation in $\mathfrak{p}$ such that it has only $(n-r)\,r$ free elements. Thus as $\mathfrak{p}$ is permitted to vary over all of $G_{r,n-r}$, β is *not* free to vary over all of $V_{r,n}$. For $\mathfrak{p} = sp\,(V)$, V is determined uniquely given $\mathfrak{p}$ and orientation of V in $\mathfrak{p}$ by $C \in O\,(r)$, such that $V = \beta C$. Note that as $\mathfrak{p}$ is permitted to vary over all of $G_{r,n-r}$, V *is* free to vary over all of $V_{r,n}$. The resulting relationship between the measures is

$$\begin{aligned} dv_r^n &= dg_r^n dv_r^r \\ \text{or } \underline{dv_r^n} &= \underline{dg_r^n}\ \underline{dv_r^r}. \end{aligned} \tag{11}$$

James[6] obtains the volume of $G_{r,n-r}$ as

$$\begin{aligned} \int_{G_{r,n-r}} dg_r^n &= \frac{\int_{V_{r,n}} dv_r^n}{\int_{O(r)} dv_r^r} \\ &= \pi^{(n-r)r}\Pi_{j=1}^r \frac{\Gamma\left[(r+1-j)/2\right]}{\Gamma\left[(n+1-j)/2\right]}. \end{aligned} \tag{12}$$

Since the polynomial term accompanying the exterior product of the differential forms is equivalent to the Jacobian for the transformation (Muirhead 1982, Theorem 2.1.1), we can see from the expression (11) that the Jacobian for the transformation V to (β, C) is one.

Next consider the transformation from $V \in V_{r,n}$, to $\overline{\beta}_2 \in R^{(n-r)r}$ and $C \in O\,(r)$ presented by Phillips (1989 and 1994, Lemma 5.2 and see also Chikuse, 1998) and reproduced here:

$$V = \left[c' + c'_\perp \overline{\beta}_2\right]\left[I_r + \overline{\beta}'_2\overline{\beta}_2\right]^{-1/2} C.$$

The differential form for this transformation is

$$\underline{dv_r^n} = \pi^{-(n-r)r}\Pi_{j=1}^r \frac{\Gamma\left[(n+1-j)/2\right]}{\Gamma\left[(r+1-j)/2\right]}\left|I_r + \overline{\beta}'_2\overline{\beta}_2\right|^{-n/2} d\overline{\beta}_2 \left(\underline{dv_r^r}\right) \tag{13}$$

(Phillips, 1989, 1994).

Equating (11) and (13) gives the result. Another, slightly more general proof for the same result is presented in Chikuse (1998).■

[6] We note that the sums, Σ, in (5.23) of James (1954) should be products, Π.

9.2 Proof that linear identifying restrictions with a flat prior give zero weight to the chosen linear restrictions:

The Jacobian defined by (5) implies that a flat prior on $\mathfrak{p}$ is informative with respect to β_2 and vice versa. This leads us to consider the implications of a flat prior on β_2 for the prior on $\mathfrak{p}$.

Theorem 3 *The Jacobian for the transformation from $\overline{\beta}_2 \in R^{(n-r)r}$ to $\mathfrak{p} \in G_{r,n-r}$ is defined by*

$$\begin{aligned}\left(d\overline{\beta}_2\right) &= \pi^{(n-r)r}\Pi_{j=1}^{r}\frac{\Gamma\left[(r+1-j)/2\right]}{\Gamma\left[(n+1-j)/2\right]}\left|I_r + (c\beta)'^{-1}\beta' c_\perp' c_\perp \beta\,(c\beta)^{-1}\right|^{n/2}(dg_r^n) \\ &= J\ dg_r^n. \end{aligned} \tag{14}$$

Proof. Invert (13) and replace $\overline{\beta}_2$ by $c_\perp\beta\,(c\beta)^{-1}$.■

Theorem 4 *Given r, use of the normalisation $\overline{\beta}_2 = c_\perp\beta\,(c\beta)^{-1}$ results in a transformation of measures for the transformation $\overline{\beta}_2 \in R^{(n-r)r} \to \mathfrak{p} \in G_{r,n-r}$ that places infinite mass in the region of null space of c relative to the complement of this region.*

Proof. Let $\rho_{c_\perp}$ be the plane defined by the null space of c. Define a ball, $\mathfrak{B}$, of fixed diameter, d, around $\rho_{c_\perp}$ and let $N_0 = \mathfrak{B} \cap G_{r,n-r}$ and $N = G_{r,n-r} - N_0$. Since for $d > 0$, $\int_N J dg_r^n$ is finite whereas $\int_{N_0} J dg_r^n = \infty$, we have

$$\frac{\int_{N_0} J dg_r^n}{\int_N J dg_r^n} = \infty.$$

■

Discussion: To demonstrate this result, consider a $n-$dimensional system for $y = (x', z')'$ where x is a r vector. To implement linear restrictions a normalisation must begin by first choosing c. Suppose it is believed that if a cointegrating relationship exists then it will most likely involve the elements of x in linearly independent relations. That is in $y\beta = x\beta_1 + z\beta_2 \backsim I(0)$, $\det(\beta_1)$ is believed far from zero making it safe to normalise on β_1, and so choose $c = [I_r\ 0]$ and estimate $\overline{\beta}_2 = c_\perp\beta\,(c\beta)^{-1}$.

$>$From (14) we see as $\mathfrak{p} = sp(\beta) \to sp(c)$, $c_\perp\beta \to 0_{(n-r)\times r}$ and $c\beta \to O(r)$ and $J \to 1$. However, as vectors in β approach the null space of c, that is $\det(c\beta) \to 0$, then $(c\beta)^{-1} \to \infty$, and thus $J \to \infty$. As a result the prior will more heavily weight regions where $\det(c\beta) = \det(\beta_1) \approx 0$, contrary to the intention of the economist. As a trivial example, consider our money demand study with $r = 1$ and $\zeta_t = \beta_1 m_t + \beta_2 inc_t$. If we believe money is most likely to enter the cointegrating relation, we would choose $c = (1, 0)$ as we believe $\beta_1 \neq 0$. Yet the Jacobian places infinite weight in the region $\beta_1 = 0$ excluding m_t from the cointegrating relation.

9.3 Proofs that the posterior will be proper and all finite moments of β exist.

Since $\mathfrak{g}_\omega$ is finite for the class of priors considered, that the Bayes factor is finite requires the integral with respect to β to be finite. The following are some general results with respect to this integral.

Theorem 5 *The marginal posterior density for β conditional upon ω has the same form for each model considered:*

$$\begin{aligned} p(\beta|\omega, y) &\propto |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} \\ &= k_\beta(\beta) \end{aligned} \tag{15}$$

where $k_\beta(\beta) = |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2}$.

Theorem 6 *The marginal posterior density for β conditional upon (r, i) in (15) is proper and all finite moments exist.*

Proof. Denote by b_{ij} any element of β. The proof follows from the result that the integral

$$M_\beta = \int_{V_{r,n}} |b_{ij}|^m k_\beta(\beta) \, dv_r^n$$

for $m = 0, 1, 2, \ldots$ is bounded above almost everywhere by the finite integral $M \int_{-1}^{1} |b_{ij}|^m db_{ij}$. As the elements of β, b_{ij}, have compact support, it is only necessary for this proof to show that $k_\beta(\beta) \, dv_r^n$ is bounded above almost everywhere by some finite constant function over $V_{r,n}$ (note the adjustment to the integral over $G_{r,n-r}$ simply requires division by the finite volume of $O(r)$, thus we only need consider the integral over $V_{r,n}$). As demonstrated in the proof to Theorem 1 in Section (4.2), dg_r^n is integrable and therefore bounded above almost everywhere by some finite constant, M_1.

The eigenvalues $\lambda_j(D_l)$ for $l = 0, 1$, will be positive and finite with probability one. By the Poincaré separation theorem, since $\beta \in V_{r,n}$, then

$$\Pi_{j=1}^{r} \lambda_{n-r+j}(D_l) \leq |\beta' D_1 \beta| \leq \Pi_{j=1}^{r} \lambda_j(D_l)$$

and so $k_\beta(\beta)$ is bounded above (and below) by some positive finite constant, M_2. Thus $k_\beta(\beta)\, dg_r^n$ has a finite upper bound, $M = M_1 M_2$. With the compact support for b_{ij}, these conditions are sufficient to ensure the posterior for β will be proper and all finite moments exist (see Billingsley 1979, pp. 174 and 180).■

10 Appendix III

10.1 An informative distribution on the cointegrating space

In this Appendix we present a method of developing an informative distribution on the cointegrating space. Uses of such a distribution include as an informative prior or as a candidate density in a Markov Chain sampling scheme. The development of the distribution makes use of a specification of the location of the distribution and a specification of the dispersion. The estimator for the mode of the posterior presented in Strachan and Inder (2004) provides a means of locating the distribution near the mass of the posterior, as for an informative prior the researcher specifies the location space in the form of a matrix.

When considering the cointegrating space $\mathfrak{p}$, we will denote our desired location or the likely value as $\mathfrak{p}^H = sp(H\kappa)$ (as in the Garrett *et al.* case) where $H \in V_{s,n}$ is a known $n \times s$ $(s \geq r)$ matrix, $H_\perp \in V_{n-s,n}$ its orthogonal complement and κ is an $s \times r$ full rank r matrix. To obtain H in $V_{s,n}$, first specify the general matrix H^g with the desired coefficient values (that is H^g may be the modal estimate of β in the posterior, or some matrix specified by the researcher). One might consider as an example the matrix H presented in Section 4.2. Next map this to $V_{r,n}$ by the transformation $H = H^g \left(H^{g\prime} H^g\right)^{-1/2}$.

At the extreme, a dogmatic distribution for $\mathfrak{p}$ could be specified by letting $\beta = H\kappa V$, $V \in O(r)$. Next define $\kappa V = V_\kappa \in V_{r,s}$ and specify the distribution in (6) for V_κ. This resulting distribution assigns unit probability mass to $\mathfrak{p} = \mathfrak{p}^H$.

Next we specify an informative, nondogmatic, distribution for $\mathfrak{p}$ centered at $\mathfrak{p} = \mathfrak{p}^H$ but with positive mass elsewhere in $G_{r,n-r}$.

Let the random scalar τ have $E(\tau) = 0$ and $E(\tau^2) = \sigma^2$. The value of σ will control the tightness of the density around $\mathfrak{p}^H$. Next construct

$$\begin{aligned} P_\tau &= HH' + H_\perp H_\perp' \tau \\ &= [H\ H_\perp] \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r}\tau \end{bmatrix} \begin{bmatrix} H' \\ H_\perp' \end{bmatrix} \end{aligned}$$

and let the elements of the $n \times r$ matrix Z be independently distributed as standard normal, $N(0,1)$. The matrix $X = P_\tau Z$ can be decomposed as $X = \beta^* \kappa$ where $\beta^* \in V_{r,n}$ and κ is an $r \times r$ upper triangular matrix. For $\tau \neq 0$ and $|\tau| < \infty$, the space of β^*, $\mathfrak{p} = sp(\beta^*)$, is a direct weighted sum of the spaces $\mathfrak{p}^H$ and $\mathfrak{p}^{H_\perp}$ with the weight determined by τ.

At $\tau = 0$ and $\tau = \pm\infty$, $\mathfrak{p}$ is respectively $\mathfrak{p}^H$ and $\mathfrak{p}^{H_\perp}$. It is for this reason that we chose $E(\tau) = 0$ such that with respect to τ, the space will on average be $\mathfrak{p}^H$. One choice for τ is $N(0,1)$ and the form of the resultant density for β^* and the hyperparameter τ is

$$p(\tau, \beta) = \tau^{-(n-r)r} \exp\left\{-\frac{\tau^2}{2}\right\} \left|\beta' P_{\tau^2}^{-1} \beta\right|^{-n/2} \mathfrak{c}_r \qquad (16)$$

where $\mathfrak{c}_r = 2^{-r-1/2} \pi^{r(r-1)/4-(n+1)r/2} \Pi_{j=1}^r \Gamma\left[(n+1-j)/2\right]$.

As an alternative, if the researcher would prefer to assign more weight in the direction of $\mathfrak{p}^{H_\perp}$ but preserve $\dim(\mathfrak{p}) = r$ with probability one, she may choose $P_\tau = HH'(1-\tau^2)^{1/2} + H_\perp H_\perp' \tau$ with $\tau \in [-1,1]$. Again the choice of $E(\tau) = 0$ would make sense and $E(\tau^2) = \sigma^2$ controls the tightness of the density around $\mathfrak{p}^H$. A possible choice of a distribution for $\eta = \tau + 1$ may be Beta over $\eta \in [0,2]$ which allows some mass to be distributed around $\mathfrak{p}^{H_\perp}$ by appropriate choice of parameter values

11 Appendix IV

11.1 Terms in the posterior:

We define the terms in (7) by $S = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1} \beta' S_{10}$, $\widetilde{B} = \begin{bmatrix} \widetilde{\alpha}' & \widetilde{\Phi}' \end{bmatrix}'$, $\widetilde{\alpha} = (\beta' S_{11} \beta)^{-1} \beta' S_{10}$, $\widetilde{\Phi} = S_{22}^{-1} S_{20}$, and $V = \widetilde{\beta}' \left(\Sigma_{t=1}^T z_t' z_t + \underline{H}\right) \widetilde{\beta}$ where

$z_t = (z_{1,t}\ z_{2,t})$. The values for the S_{ij} are defined as

$$
\begin{array}{rcll}
TM_{ij} & = & h_{ij} + \Sigma_{t=1}^{T} z'_{i,t} z_{j,t} & \text{for } i \text{ and } j = 1, 2, \\
h_{ij} & = & 0 \text{ if } i \neq j \text{ and } h_{ii} = 0.01I, & \\
TM_{20} & = & \Sigma_{t=1}^{T} z'_{2,t} \Delta y_t, & TM_{10} = \Sigma_{t=1}^{T} z'_{1,t} \Delta y_t, \\
TM_{00} & = & \Sigma_{t=1}^{T} \Delta y'_t \Delta y_t \quad \text{and so} & \\
S_{ij} & = & M_{ij} - M_{i2} M_{22}^{-1} M_{2j} & \text{for } ij = 0, 1, 2, \\
\text{except } i & = & j = 2 \text{ where} & \\
S_{22} & = & M_{22} - M_{21} M_{11}^{-1} M_{12} \text{ and} & \\
S_{20} & = & M_{20} - M_{21} M_{11}^{-1} M_{10}. &
\end{array}
$$

For later use we also define $D_0 = D_1 - D_2$, $D_1 = S_{11}$ and $D_2 = S_{01} S_{11}^{-1} S_{10}$.