University of Leicester

DEPARTMENT OF ECONOMICS

MODEL UNCERTAINTY AND BAYESIAN MODEL AVERAGING IN VECTOR AUTOREGRESSIVE PROCESSES

Rodney W. Strachan, University of Leicester, UK

Herman K. van Dijk, Erasmus University, The Netherlands

Working Paper No. 06/5
February 2006

Model Uncertainty and Bayesian Model Averaging in Vector Autoregressive Processes

Rodney W. Strachan[1*]
Herman K. van Dijk[2]
[1]Department of Economics, University of Leicester,
Leicester LE1 7RH UK.
email: rws7@le.ac.uk

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

February 6, 2006

*Corresponding author.

ABSTRACT

Economic forecasts and policy decisions are often informed by empirical analysis based on econometric models. However, inference based upon a single model, when several viable models exist, limits its usefulness. Taking account of model uncertainty, a Bayesian model averaging procedure is presented which allows for unconditional inference within the class of vector autoregressive (VAR) processes. Several features of VAR process are investigated. Measures on manifolds are employed in order to elicit uniform priors on subspaces defined by particular structural features of VARs. The features considered are the number and form of the equilibrium economic relations and deterministic processes. Posterior probabilities of these features are used in a model averaging approach for forecasting and impulse response analysis. The methods are applied to investigate stability of the "Great Ratios" in U.S. consumption, investment and income, and the presence and effects of permanent shocks in these series. The results obtained indicate the feasibility of the proposed method.

Key Words: Posterior probability; Grassman manifold; Orthogonal group; Cointegration; Model averaging; Stochastic trend; Impulse response; Vector autoregressive model.

JEL Codes: C11, C32, C52

1 Introduction.

In this paper we take account of model uncertainty and introduce a method of using Bayesian model averaging in the class of vector autoregressive processes. We demonstrate the operational implications of our approach by investigating the stability of the "Great Ratios" in U.S. consumption, investment and income, and analysing the presence and effects of permanent shocks for the duration of the business cycle in these series.

The idea underlying Bayesian model averaging is relatively straightforward. Model specific estimates are weighted by the corresponding posterior model probability and then averaged over the set of models considered. Although many statistical arguments have been made in the literature to support model averaging (e.g., Leamer, 1978, Hodges, 1987, Draper, 1995, Min and Zellner, 1993 and Raftery, Madigan and Hoeting, 1997), recent applications suggest its relevance for macroeconometrics (Fernández, Ley and Steel, 2001 and Sala-i-Martin, Doppelhoffer and Miller, 2004). Here we mention three reasons for this relevance.

The first reason is relevance for forecasting and policy analysis. An important function of empirical economic analysis is to provide accurate information for decision making. For example, there is evidence that permanent - possibly productivity - shocks account for most fluctuations in consumption (King, Plosser, Stock and Watson, 1991, and Lettau and Ludvigson, 2004) and information may be required on the form of the response in consumption to such a permanent shock. Centoni and Cubadda (2003), however, focus upon business cycle fluctuations and find permanent shocks are not very important. While the decision maker is not directly interested in the underlying model used to estimate the response, it is, however, the econometrician's responsibility to detail the model upon which these estimates rely. If there is any uncertainty about the veracity of the model, the expected loss (from choosing a policy action) from that single model cannot equal the expected loss that accurately accounts for model uncertainty.

A second reason for considering model averaging is methodological. There are well known issues relating to the complexity of the model set and the sequences used to select a model. The standard approach to providing inference is to select a single model and present empirical results based upon this model. The usual strategy of model selection using sequential testing procedures, however, introduces problems of model uncertainty. In the context of sequential hypothesis testing, the pre-test problem is well understood (see, for example, Poirier, 1995, pp. 519-523) and has received considerable attention in the statistical and econometric literature. We do not intend (nor are able) to survey the literature here, but mention that just within the unit root and cointegration testing there have been several studies such as Elliott and Stock (1994), Elliott (1998), Phillips (1996), Chang and Phillips (1995) and Chang (2000) (see for useful discussion, Maddala and Kim, 1998, pp. 139-140 and 229-231).

The problem is self evident. Whether we accept or do not accept an hypothesis, the veracity of the adopted hypothesis is uncertain. Subsequent tests condition upon that uncertain outcome and have their own uncertain outcomes. This process can lead to significant size distortions and inappropriate reported standard errors. Generally, the resulting standard errors will not fully reflect the uncertainty associated with the estimates. The longer the sequence of tests the more the problem compounds, and the sequence can become very long if, for example, we consider: lag length; the type of deterministic processes present; the number of cointegrating relations; overidentifying restrictions on the cointegrating space; and even whether certain variables are in some sense (weakly or strongly) exogenous for the inference in question. Despite the extensive concern shown in the literature for the pretest problem, however, a generally applicable strategy for dealing with this issue does not appear to be available. It would seem the usual (implicit) approach is to "... entirely ignore the problems caused by pretesting, not because they are unimportant, but because, in practice, they are generally intractable" (Davidson and MacKinnon, 1993, pp. 97-98).

An additional, related, problem due to the complexity of the model, is the conflicting inferences that may arise depending upon which sequence of tests is employed. For example, using the Johansen trace test and data on consumption, investment and income from Paap and van Dijk (2003), we find that the chosen cointegrating rank depends upon the chosen deterministic term[1] and the rank may be zero or one. This suggests it is important to determine the correct deterministic process before investigating the cointegrating rank. However, the range of deterministic process that can occur differs if cointegration occurs or not. To take this example further, let us assume a rank of one for these variables and we are now interested in 1)

[1] As the deterministic processes enter the error correction term, testing for presence of a trend in a VAR in levels, when cointegration is present, does not identify the deterministic process.

whether the error correction term, z_t, has a trend and 2) if the Great Ratios of consumption to income and investment to income enter z_t.[2] Depending upon whether we test stability of the Great Ratios first or test the presence of various deterministic terms first, we find either we have no trend in z_t and that the Great Ratios do not enter z_t, or that the Great Ratios do enter z_t and z_t has a linear trend.

A third reason for considering Bayesian model averaging is a pragmatic one. The support in the data is in many cases not clear or dogmatically for or against a restriction, and researchers often do not have strong prior belief in particular restrictions. The strategy of testing hypotheses on restrictions and conditioning upon the outcome, effectively assigns a weight of one to the model implied by the restriction 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 inference.

Thus, there is 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 of the uncertainty associated with the structural features defining the model.

We present a Bayesian approach for conducting unconditional inference from 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. Given the 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 the number of stable equilibrium relationships among economic variables,

[2]This implies a particular overidentifying restriction on the cointegrating space holds.

the forms of those equilibrium relationships, the dynamic responses to disequilibria, and the type of deterministic processes that may be present. In order to obtain these results we make of 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.

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 briefly demonstrate the proposed methodology with an investigation of the stability of the 'Great Ratios' as discussed in King, Plosser, Stock and Watson (1991) (hereafter KPSW), and the relative weights of permanent and transitory components in US consumption, investment and income, and, finally, the credibility of alternative paths of responses to a possible productivity shock.

There exist several Bayesian analyses of VAR processes in the literature. A complete survey is outside the scope of our paper, although we mention the following approaches. Using so-called "Minnesota" priors, which are of a random walk nature, Doan, Litterman and Sims (1984) investigate Bayesian forecasting and impulse response analysis using unrestricted VARs. Sims and Zha (1999) investigate confidence bands of impulse responses using unrestricted VARs. Other papers using unrestricted VARs include Koop (1991 and 1994) and Canova and Matteo (2004). Structural features in VAR models, like cointegration, are investigated by Kleibergen and Van Dijk (1994), Strachan (2003), Villani (2005) using diffuse type of priors. We extend the analysis of these two approaches by considering priors on structural features and by investigating the implied forecasts and impulse responses using Bayesian model averaging.

The structure of the paper is as follows. In the Section 2 we introduce the models of interest in this paper - the vector autoregressive models, the general structural features of interest, and the restrictions they imply. In Section 3 we present the priors, the likelihood and useful expressions for the posterior. The tools for inference in this paper, posterior probabilities, are introduced and general expressions are derived for estimators of 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 an investigation of the responses of consumption, investment and income to a permanent shock allowing for stability of consumption to income and investment to income ratios. In Section 5 we summarize conclusions and discuss possibilities for further research.

2 A Set of Vector Autoregressive Models.

Since the influential work by Sims (1980), the vector autoregressive model has enjoyed much success in macroeconometrics. These models 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).

When a VAR process cointegrates, the model may be written in the vector error correction model (VECM) form. The VECM 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

$$\begin{aligned} \Delta y_t &= \left(d_{1,t}\theta_1 + y_{t-1}\beta^{+}\right)\alpha + d_{2,t}\theta_2 + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \quad (1) \\ &= z_{1,t}\beta\alpha + z_{2,t}\Phi + \varepsilon_t \quad (2) \end{aligned}$$

where $\Delta y_t = y_t - y_{t-1}$, $z_{1,t} = (d_{1,t}, y_{t-1})$, $z_{2,t} = (d_{2,t}, \Delta y_{t-1}, \ldots, \Delta y_{t-l})$, $\Phi = (\theta_2', \Gamma_1', \ldots, \Gamma_l')'$ and $\beta = \left(\theta_1', \beta^{+\prime}\right)'$. The matrices Γ_i are $n \times n$ and β^{+} and α' are $n \times r$ and assumed to have rank r, and if $r = n$ then $\beta^{+} = I_n$. We define the deterministic terms $d_{i,t}\theta_i$ formally below.

Here we define the restrictions of interest, combinations of which define different model features of interest which we may compare or weight using posterior probabilities. The restrictions refer to the number of equilibrium relations, to the form of these relations, the lag length and to particular types of deterministic processes.

We denote the number of stable equilibrium relationships or, more precisely, the cointegrating rank by r, where $r = 0, 1, \ldots, n$. For cointegration analysis of (1), the parameters of interest are the coefficient matrices 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 simplicity $\theta_1 = 0$, β^+ is the matrix of cointegration coefficients, $y_t\beta^+ = 0$ 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 $y_t\beta^+ = 0$.

A second feature of interest is the particular identifying restrictions placed upon β. These will be denoted by o, where $o = 0, 1, \ldots, \mathfrak{J}$ and $o = 0$ will be understood to refer to the just identified model. A range of restrictions commonly investigated are presented in Johansen (1995, Chapter 5). We restrict ourselvest to two cases: no restriction on β $(o = 0)$; and $\beta = H\psi$ $(o = 1)$ where ψ is an $s \times r$ matrix such that the cointegrating space is either completely determined (if $r = s$) or is restricted to be within the space spanned by H.

The deterministic processes in the level, y_t, and the equilibrium relations, $y_t\beta^+$, are given respectively by the terms $d_{1,t}\theta_1$ and $d_{2,t}\theta_2$ in (1). The contents and dimensions of the $d_{i,t}$ and the θ_i depend upon the particular deterministic process that occur in $y_t\beta^+$ and Δy_t (and therefore y_t). In the discussion that follows, μ_1 and δ_1 are $1 \times r$ vectors, while μ_2 and δ_2 are $1 \times n$ vectors. These processes can be linear trends, non-zero means or zero mean for $y_t\beta^+$, and no drift, linear drift and quadratic drift in y_t. For example, if $\theta_2 = (\mu_2' \quad \delta_2')'$ then $d_{2,t} = (1, t)$ and this implies y_t will have a quadratic drift. If $\theta_2 = \mu_2$ then $d_{2,t} = (1)$ and this implies y_t will have a linear drift. We consider the five commonly used combinations in the table below (see, for example, Johansen, 1995):

d	$d_{1,t}\theta_1$	$d_{1,t}$	$y_t\beta^+$	$d_{2,t}\theta_2$	$d_{2,t}$	y_t
1	$\mu_1+\delta_1 t$	$(1,t)$	linear trend	$\mu_2+\delta_2 t$	$(1,t)$	quadratic drift
2	$\mu_1+\delta_1 t$	$(1,t)$	linear trend	μ_2	(1)	linear drift
3	μ_1	(1)	non-zero mean	μ_2	(1)	linear drift
4	μ_1	(1)	non-zero mean	0	$\{\}$	no drift
5	0	$\{\}$	zero mean	0	$\{\}$	no drift

Each model will be identified by M_ω where $\omega=(r,o,d)$ and $\omega\in\Omega$ where Ω is the set of all ω we consider. For example, the least restricted model will be $M_{(n,0,1)}$, while the most restricted model will be $M_{(0,1,5)}$. As an example of models we consider, KPSW begin their investigation with results using two VAR models with six lags: the first having only a constant, $M_{(n,0,3)}$, and the second having a constant and a trend, $M_{(n,0,1)}$. From these models they find evidence that suggests support for two equilibrium relations of known form and a linear drift which within our model set is $M_{(2,1,3)}$.[3]

Thus, with $n=3$ in our application, we deal with a case of $4\times 2\times 5=40$ models[4]. 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 except to note here that this increases the number of models to 40 times the number of lags we consider.

3 Priors and Posteriors.

In this section the forms of the priors and resultant posterior are presented. We begin with discussion of the distribution of the prior probabilities over the model space which contains some models that are impossible and others that are observationally equivalent. Next we consider the priors for the parameters (Σ,B). Conditional upon β the model in (2) is linear. This fact makes it relatively straightforward to elicit priors on these parameters. We next give careful consideration to the prior for β before presenting the method of posterior analysis.

[3] Interestingly, they later report results on responses to permanent (productivity) shocks using $M_{(2,1,3)}$ but with eight lags of differences.

[4] This reduces to 26 models when we account for impossible models and observationally equivalent models. See Subection 3.1 below for further discussion on this point.

3.1 The Prior for (Σ, B, M_ω).

In this paper we wish to treat all models as *a priori* equally likely, however this is not a straightforward issue[5]. The priors for the individual elements of $\omega = (r, o, d)$ 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 the models with $r = n$ and $d = 1$ or 2). The natural prior probability to assign to impossible models is zero[6]. However, the researcher must carefully consider how she wishes to treat observationally equivalent models.

It would seem sensible to regard observationally equivalent models as one model and then assign equal prior probabilities to all models. For example, at $r = 0$ the models with $d = 2$ and $d = 3$ are observationally equivalent. If we were to treat these two as one model, they would receive half the prior probability of other models with rank $0 < r < n$. Systematic employment of this principle, however, would bias the prior weight in favour of models with $0 < r < n$. This could shift the posterior weight of evidence in favour of some economic theories for which we wish to determine the support.

Alternatively we could specify all possible combinations of the indices in ω be equally likely to avoid biasing the evidence in favour of other classes of models. However, any bias towards some models can be viewed as simply a result of Bayes Theorem. This is the view we take and we implement the first approach (treating observationally equivalent models as one model) in the following way. We first assign probabilities to various values of the model features such as different cointegrating ranks, $p(r)$, or deterministic processes, $p(d)$. We then set the prior weighting for each model as $k(M_\omega) = p(r) p(o) p(d)$. Next, set $k(M_\omega) = 0$ for impossible combinations and for each set of combinations of ω that imply observationally equivalent models, we set $k(M_\omega) = 0$ for all but one of the combinations. Finally we compute the prior model probabilities as $p(M_\omega) = k(\omega) / \Sigma_\omega k(\omega)$ where in the denominator we have summed $k(M_\omega)$ over all ω.

To demonstrate these prior probabilities we use the application in this paper. As we have $n = 3$, $r \in [0, 1, 2, 3]$ so we use $p(r) = (n+1)^{-1} = 0.25$

[5] The authors are grateful to Geert Dhaene and an anonymous referee for useful comments on this issue.

[6] Although the actual prior probability we assign to impossible models - provided it is less than one - is irrelevant as the marginal likelihood for these models will be zero, such that the posterior probability will be zero by design.

and with $d \in [1, 2, 3, 4, 5]$ we set $p(d) = 0.2$. In our application we consider two states of overidentification of β. In the first state β is unrestricted $(o = 0)$ and in the second we have $\beta = H\psi$ $(o = 1)$ and so we set $p(o) = 1/2$ for $o \in [0, 1]$.

For each model implied by a particular value of ω, we need to specify a prior for the parameters in the model. We use the standard diffuse prior for Σ, $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$.

As B changes dimensions across the different versions of ω implied by different models 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 $p(B|\beta, M_\omega) \propto 1$ were used. For the original discussion on this point see Bartlett (1957) and more recently O'Hagan (1995), Strachan and van Dijk (2003) and Strachan and van Dijk (2005). For this reason a weakly informative proper prior for B must be used. We take the prior for $b = vec(B)$ conditional upon $(\Sigma, \beta, M_\omega)$ as Normal with zero mean and covariance $\underline{V} = \Sigma \otimes \eta^{-1} I_{(r+k_i)}$.[7] We choose the value of $\eta = 10$ as this provides a mild degree of shrinkage towards zero which has been shown to improve estimation (See Ni and Sun, 2003). Further evidence on the influence of this choice can be found in Strachan and Inder (2004).

3.2 Eliciting a Prior on β.

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. Much of the work to date in Bayesian cointegration analysis has used linear identifying restrictions. That is, 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 $c_\perp c' = 0$. For example, if $c = [I_r \ 0]$ then $\overline{\beta} = \left[I_r \ \overline{\beta}_2' \right]'$. A prior is then specified for $\overline{\beta}_2$ which is then estimated and often its value is interpreted.[8]

[7] If an informative prior is used on for the cointegrating space then we recommend the prior for B in Koop, León-González and Strachan (2005), of which the prior presented here is a specific case.

[8] 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.

Assuming that c is known, Kleibergen and van Dijk (1994 & 1998) demonstrate how a flat prior on $\overline{\beta}_2$ can result in, at best, nonexistence of moments of $\overline{\beta}_2$, and, at worst, an improper posterior distribution thus precluding inference. They also outline how local nonidentification 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, and a related solution is proposed in Kleibergen and Paap (2002) and Paap and Van Dijk (2003). Using these approaches avoids the issue of local nonidentification, results in proper posteriors and allows use of MCMC, however the posterior again has no moments of $\overline{\beta}_2$.

Bauwens and Lubrano (1996) 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. Nonexistence of moments is not commonly a concern for estimation as modal estimates exist as alternative measures 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.

For the model averaging we require posterior probabilities, however, as is well known, a flat prior on $\overline{\beta}_2$ cannot be employed to obtain posterior probabilities for M_ω since the dimensions of $\overline{\beta}_2$ depend upon ω. It would appear, then, that we need to be informative to obtain inference.

Denoting the space spanned by β by $\mathfrak{p} = sp(\beta)$, we can say it is $\mathfrak{p}$, and not β, that is the primary object of interest 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[9] $G_{r,n-r}$ (James, 1954), $\mathfrak{p} \in G_{r,n-r}$. A requirement to employ linear restrictions is that we know enough about the cointegrating space to be able to choose c such that $c\beta$ is nonsingular such that $\overline{\beta}_2 = c_\perp \beta (c\beta)^{-1}$ exists. Making use of this assumption to impose these linear restrictions, however, has the unexpected and undesirable result that it makes this assumption *a priori* impossible (see the Appendix, Theorem 3).

[9] 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 (2005) also makes use of a prior on $\mathfrak{p}$.

To employ uninformative priors, to simplify the application and estimation, and as we do not see $\overline{\beta}_2$ (but rather $\mathfrak{p}$) 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.

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. We save the technical discussion for the Appendix, but to implement this approach, we specify β to be semiorthogonal, i.e., $\beta'\beta = I_r$, and specify a Uniform distribution for β (for details see Strachan and Inder (2004) and Strachan and van Dijk (2003)).

A Uniform prior for $\mathfrak{p}$ over $G_{r,n-r}$ is implied by a Uniform prior for β over $V_{r,n}$. This prior has the form

$$p(\beta|M_\omega) = \frac{1}{\mathfrak{c}_\beta} \tag{3}$$

where $\mathfrak{c}_\beta = \int_{G_{r,n-r}} d\beta$[10] and β is the r-frame with fixed orientation in $\mathfrak{p}$. The measure on $G_{r,n-r}$ used in the above expression is derived from its relationship with the spaces $V_{r,n}$ and $O(r)$ in the proof of Theorem 2 in the Appendix. This proof also provides an expression for $\mathfrak{c}_\beta$.

For the cases in which we impose identifying restrictions discussed in Section 2 of the form $\beta = H\psi$ $(o = 2)$, we impose $\psi \in V_{r,s}$ and impose the Uniform prior on $V_{r,s}$. This implies that we are unformative about the orientation of the vectors β in $sp(\beta)$. For computational and mathematical simplicity we also convert H to be semiorthogonal by the transformation $H \to H(H'H)^{-1/2}$. This transformation is innocuous since the space of H, which is the important parameter, is unchanged by this transformation.

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. We note at this point that Strachan and Inder (2004) extend this approach to informative distributions on the cointegrating space.

The full prior distribution for the parameters in a given model is then given by $p(\Sigma, B, \beta|M_\omega) = p(B|\Sigma, \beta, M_\omega)\, p(\Sigma|M_\omega)\, p(\beta|M_\omega)$.

[10] We acknowledge that this notation is not technically correct. If we were to denote the measure for the Grassman manifold as dg_r^n, then we should really write $\mathfrak{c}_\beta = \int_{G_{r,n-r}} dg_r^n$. However, for notational clarity we use the notation $d\beta$.

3.3 Posterior Analysis.

An expression for posterior distribution of the parameters for any model given the data, $p(\Sigma, B, \beta|M_\omega, y)$, is obtained by combining the prior, $p(\Sigma, B, \beta|M_\omega)$, with the likelihood for the data $L(y|\Sigma, B, \beta, M_\omega)$. That is,

$$p(\Sigma, B, \beta|M_\omega, y) \propto p(\Sigma, B, \beta|M_\omega) L(y|\Sigma, B, \beta, M_\omega) = k(B, \Sigma, \beta, M_\omega|y). \tag{4}$$

As we will be using a Gibbs sampling scheme we need to present the conditional posterior for each parameter. To simplify the presentation of the posteriors, we use the following two facts. First, conditional upon β the model is linear. Second, the matrices α and β always occur in a product form as $\beta\alpha$ such that we can introduce any full rank square $r \times r$ matrix (unidentified) D such that $\beta\alpha = \beta D D^{-1}\alpha = \beta^*\alpha^*$. Note that the matrices α^* and α have the same support. However, β is semiorthogonal with the Stiefel manifold as its support while β^* has as its support the nr dimensional real space. The approach we use follows that of Koop, León-González and Strachan (2005). Development of the sampling scheme and further details may be found in that paper.

To further simplify the expressions of the posteriors we introduce the following notation. For the model in (2), assume the rows of the $T \times n$ matrix $E = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_T')'$ are $\varepsilon_t \sim iidN(0, \Sigma)$ and define the $T \times n$ matrix $Z_0 = (\Delta y_1', \Delta y_2', \ldots, \Delta y_T')'$ and the $T \times (r + k_i)$ matrix $Z = (Z_1\beta \quad Z_2)$ where $Z_1 = \left(z_{1,1}', z_{1,2}', \ldots, z_{1,T}'\right)'$ and $Z_2 = \left(z_{2,1}', z_{2,2}', \ldots, z_{2,T}'\right)'$. Finally, let B be the $(r + k_i) \times n$ matrix $B = [\alpha' \ \Phi']'$. We may now write the model, given in equation (1) as

$$Z_0 = Z_1\beta\alpha + Z_2\Phi = ZB + E.$$

Vectorising this expression we have

$$z_0 = zb + e \tag{5}$$

where $z_0 = vec(Z_0)$, $z = (I_n \otimes Z)$, $b = vec(B)$ and $e = vec(E)$. The form of the likelihood is then

$$L(y|\Sigma, B, \beta, M_\omega) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr\Sigma^{-1}E'E\right\}.$$

Combining the likelihood with the prior $p(\Sigma)$, we can see the covariance matrix Σ has a posterior distribution conditional upon (B, β) that is inverted Wishart with scale matrix $E'E$ and degrees of freedom T.

We can now use standard algebraic operations (see e.g., Zellner, 1971) to show

$$\begin{aligned} tr\Sigma^{-1}E'E &= e'\left(\Sigma^{-1}\otimes I_T\right)e \\ &= s^2+\left(b-\widehat{b}\right)'V^{-1}\left(b-\widehat{b}\right) \end{aligned}$$

where $s^2 = z_0' M_V z_0$, $M_V = \Sigma^{-1}\otimes\left(I_T - Z\left(Z'Z\right)^{-1}Z'\right)$, $\widehat{b} = \left(I_n\otimes\left(Z'Z\right)^{-1}Z\right)z_0$ and $V = \Sigma\otimes\left(Z'Z\right)^{-1}$. The likelihood can then be written as

$$L\left(y|\Sigma, B, \beta, M_\omega\right) \propto |\Sigma|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}\left[s^2+\left(b-\widehat{b}\right)'V^{-1}\left(b-\widehat{b}\right)\right]\right\}. \quad (6)$$

Thus we can see that conditional upon β, if we were to use a flat prior for B, the conditional posterior distribution would be Normal with mean $\widehat{b}$ and covariance matrix V. Combining this form with the informative prior given in the previous subsection, we obtain the conditional posterior with mean $\overline{b} = \overline{V}V^{-1}\widehat{b} = \left(I_n\otimes\left(Z'Z+\eta I\right)^{-1}Z'\right)z_0$ and covariance matrix $\overline{V} = \left[V^{-1}+\underline{V}^{-1}\right]^{-1} = \Sigma\otimes\left(Z'Z+\eta I\right)^{-1}$.

As β is semiorthogonal, it is clear that the posterior distribution will be nonstandard regardless of the form we choose for the prior. Therefore, to obtain a useful expression for the posterior for obtaining draws of β, we make use of the transformation to β^* and α^*. We give β^* a Normal prior with zero mean and covariance matrix $n^{-1}I_{nr}$. We can easily transform back to the parameters of interest via $\beta = \beta^* D^{-1}$ and $\alpha = \alpha^* D$. The prior for β^* implicitly specifies a proper prior for D and that the marginal prior for $\beta = \beta^* D^{-1}$ is Uniform as specified in the above subsection.

Next we vectorise $Z_1\beta^*\alpha^*$ to obtain $vec\left(Z_1\beta^*\alpha^*\right) = xb_{\beta^*}$ where $b_{\beta^*} = vec\left(\beta^*\right)$ and $z_1 = \left(\alpha^{*\prime}\otimes Z_1\right)$. Thus we can rewrite the expression in (5) as

$$\begin{aligned} vec\left(Z_0 - Z_2\Phi\right) &= vec\left(Z_1\beta\alpha\right) + vec\left(E\right) \text{ or} \\ \widetilde{z}_0 &= z_1 b_{\beta^*} + e \end{aligned}$$

where $\widetilde{z}_0 = vec\left(Z_0 - Z_2\Phi\right)$. Thus the likelihood can be written as

$$L\left(y|\Sigma, B, \beta, M_\omega\right) \propto |\Sigma|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}tr\Sigma^{-1}\left[s_\beta^2+\left(b_{\beta^*}-\widehat{b}_{\beta^*}\right)'V_{\beta^*}^{-1}\left(b_{\beta^*}-\widehat{b}_{\beta^*}\right)\right]\right\}$$

where $s_{\beta^*}^2 = \widetilde{z}_0' M_{V_{\beta^*}}\widetilde{z}_0$, $M_{V_{\beta^*}} = \left(\Sigma^{-1}\otimes I_T\right) - \left(\Sigma^{-1}\left(\alpha^*\Sigma^{-1}\alpha^{*\prime}\right)^{-1}\Sigma^{-1}\otimes\left(Z_1'Z_1\right)^{-1}\right)$, $\widehat{b}_{\beta^*} = V_{\beta^*}\left(\alpha^{*\prime}\Sigma^{-1}\otimes Z_1\right)\widetilde{z}_0$, and $V_{\beta^*} = \left(\alpha^*\Sigma^{-1}\alpha^{*\prime}\right)^{-1}\otimes\left(Z_1'Z_1\right)^{-1}$. In this case if we were to use a flat prior for β we see the posterior distribution of b^*_β, conditional upon (Σ, B), would be Normal with mean $\widehat{b}_{\beta^*}$ and covariance matrix V_{β^*}. Combining this form with the informative prior given above, we obtain the conditional posterior with mean $\overline{b}_{\beta^*} = \overline{V}_{\beta^*} V_{\beta^*}^{-1} \widehat{b}_{\beta^*}$ and covariance matrix $\overline{V}_{\beta^*} = \left[V_{\beta^*}^{-1} + \underline{V}_{\beta^*}^{-1} \right]^{-1} = \left[(\alpha^* \Sigma^{-1} \alpha^{*\prime} \otimes Z_1' Z_1) + n I_{nr} \right]^{-1}$.

An important component of Bayesian inference is the posterior probability of each model, $p(M_\omega|y)$. These can be derived from the marginal likelihoods for each model via the expression

$$p(M_i|y) = \frac{m_i p(M_i)}{\sum_{\omega \in \Omega} m_\omega p(M_\omega)}$$

where the summation in the denominator is over all elements of Ω. The marginal likelihood for a model will be m_ω where

$$m_i = \int_{R^{(k_i+r)n}} \int_{\Sigma>0} \int_{G_{r,n-r}} k(B, \Sigma, \beta, M_\omega|y)\,(d\beta)\,(d\Sigma)\,(dB), \tag{7}$$

where $B \in R^{(k_i+r)n}$, Σ is positive definite (denoted $\Sigma > 0$). To integrate (7) with respect to (B, Σ, β) we first analytically integrate (4) with respect to (B, Σ) as these parameters have conditional posteriors of standard form. This integration gives us the following.

Theorem 1 *The marginal posterior for (β, M_ω) is*

$$p(\beta, M_\omega|y) \propto \mathfrak{g}_\omega \left|\beta' D_0 \beta\right|^{-T/2} \left|\beta' D_1 \beta\right|^{(T-n)/2} \tag{8}$$

where in this case

$$\mathfrak{g}_\omega = |S_{00}|^{-T/2} |M_{22}|^{-n/2} T^{-n(k_i+r)/2} \eta^{n(k_i+r)/2} \times \mathfrak{c}_\beta^{-1}.$$

The expressions for D_0 and D_1 are

$$\begin{aligned} D_1 &= (Z_1' Z_1 + \eta I_r) - Z_1' Z_2 (Z_2' Z_2 + \eta I_{k_i})^{-1} Z_2' Z_1 \text{ and} \\ D_0 &= D_1 - S_{01} S_{11}^{-1} S_{10} \end{aligned}$$

where

$$\begin{aligned} S_{10} &= Z_1' Z_0 - Z_1' Z_2 (Z_2' Z_2 + \eta I_{k_i})^{-1} Z_2' Z_0, \\ S_{00} &= Z_0' Z_0 - Z_0' Z_2 (Z_2' Z_2 + \eta I_{k_i})^{-1} Z_2' Z_0, \end{aligned}$$

Proof. See, for example, Zellner (1971) or Bauwens and van Dijk (1990)[11].■

Next we need to integrate (8) with respect to β to obtain the posterior for M_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 the Appendix 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 we wish to report expectations of these objects, we require the existence of moments of β.

To obtain the posterior distribution of $\omega = (r, o, d)$, $p(M_\omega|y)$, it is necessary to integrate (8) with respect to β and so obtain an expression for

$$p(M_\omega|y) = \int p(\beta, M_\omega|y)\, d\beta. \tag{9}$$

The marginal density of β conditional on ω implied by (3) and (8) is

$$p(\beta|M_\omega, y) = |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} / \mathfrak{c}_\omega \tag{10}$$

and is not of standard form. Although one may exist, we do not currently know of a simple, general analytical solution for

$$\mathfrak{c}_\omega = \int_{G_{r,n-r}} |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} d\beta \tag{11}$$

and so we estimate $\mathfrak{c}_\omega$ and obtain our estimate of m_i from $m_i = \mathfrak{c}_\omega \mathfrak{g}_\omega$.

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

[11] Remark: From the expression (8) that we see that not only is $d\beta$ invariant to $\beta \to \beta C$ for $C \in O(r)$, but so is the kernel of the marginal density for β given M_ω, $k(\beta|M_\omega, y)$, and thus the complete posterior for β given M_ω.

using restrictions related to those of the ML estimator of Johansen (1992). An approach commonly used in classical work to approximate integrals over $V_{r,n}$, and therefore $G_{r,n-r}$, is to use the Laplace approximation (see Strachan and Inder, 2004) which is computationally much faster than MCMC but the accuracy depends upon the sample size. An additional contribution of this paper is the development of an MCMC method to estimate the integral (11).

3.4 Bayesian Model Averaging with MCMC.

In this section we outline how we implement Bayesian model averaging to provide unconditional inference. We then present the steps in the sampling scheme and, finally, we discuss how we obtain estimates of the marginal likelihoods.

Suppose we have an economic object of interest ϕ which is a function of the parameters for a given model $(B, \Sigma, \beta|M_\omega)$, $\phi = \phi\,(B, \Sigma, \beta|M_\omega)$. We wish to report the unconditional (upon any particular model) expectation of this object. That is, we wish to report an estimate of

$$E\,(\phi|y) = \sum_{\omega\in\Omega} E\,(\phi|y, M_\omega)\,p\,(M_\omega|y)$$

where $E\,(\phi|y, M_\omega)$ is the expectation of ϕ from model ω. To obtain this estimate, denote the i^{th} draw of the parameters from the posterior distribution for model M_ω as $\left(B^{(i)}, \Sigma^{(i)}, \beta^{(i)}\right)$ and so the i^{th} draw of ϕ as $\phi^{(i)} = \phi\left(B^{(i)}, \Sigma^{(i)}, \beta^{(i)}|M_\omega\right)$. Next suppose we have $i = 1, \ldots, J$ draws of the parameters from the posterior distribution for each model. To approximate $E\,(\phi|y)$, we first obtain estimates of $E\,(\phi|y, M_\omega)$ from each model by

$$\widehat{E}\,(\phi|y, M_\omega) = \frac{1}{M}\Sigma_{i=1}^{M}\phi^{(i)} \text{ for each } \omega.$$

These estimates are then averaged as

$$\widehat{E}\,(\phi|y) = \sum_{j=1}^{J} \widehat{E}\,(\phi|y, M_\omega)\,\widehat{p}\,(M_\omega|y)$$

in which $\widehat{p}\,(M_\omega|y)$ is an estimate of $p\,(M_\omega|y)$. Therefore we require draws of the parameters $\left(B^{(i)}, \Sigma^{(i)}, \beta^{(i)}\right)$ for each model and estimates of the posterior model probabilities, $\widehat{p}\,(M_\omega|y)$. We use the following scheme at each step i to obtain draws of (B, Σ, β).

1. Initialize $(b, \Sigma, b_{\beta^*}) = \left(b^{(0)}, \Sigma^{(0)}, b_{\beta^*}^{(0)}\right)$.

2. Draw $\Sigma | b, b_{\beta^*}$ from $IW\left(E'E, T\right)$

3. Draw $b | \Sigma, b_{\beta^*}$ from $N\left(\bar{b}, \overline{V}\right)$

4. Draw $b_{\beta^*} | \Sigma, b$ from $N\left(\bar{b}_{\beta^*}, \overline{V}_{\beta^*}\right)$.

5. Repeat steps 2 to 4 for a suitable number of replications.

For the computation of the posterior probabilities, we need only draws of β to approximate the integral in (11). If the model set becomes large then computation times for the above strategy may become rather large. A sensible strategy then would be to include the model in the sampling scheme. This could be achieved using a method such as the reversible jump methodology of Greene (1995).

To estimate the marginal likelihood, we must estimate the term

$$\mathfrak{c}_\omega = \int \left|\beta' D_0 \beta\right|^{-T/2} \left|\beta' D_1 \beta\right|^{(T-n)/2} d\beta = \int k\left(\beta\right) d\beta$$

where $k\left(\beta\right) = \left|\beta' D_0 \beta\right|^{-T/2} \left|\beta' D_1 \beta\right|^{(T-n)/2}$. We approximate this integral using the method proposed by Gelfand and Dey (1994) which uses the relation

$$\frac{1}{\mathfrak{c}_\omega} = \int \frac{q\left(\beta\right)}{k\left(\beta\right)} \frac{k\left(\beta\right)}{\mathfrak{c}_\omega} d\beta.$$

in which $q\left(\beta\right)$ is a proper known density. As we have we have a sequence of draws $\beta^{(i)}$, $i = 1, ..., J$, from the posterior distribution for β, we can estimate $\mathfrak{c}_\omega$ by

$$\widehat{\mathfrak{c}_\omega} = J \left(\Sigma_{i=1}^{J} \frac{q\left(\beta^{(i)}\right)}{k\left(\beta^{(i)}\right)} \right)^{-1}.$$

As our choice for $q\left(\beta\right)$, we use a Matrix Angular Central Gaussian distribution of Chikuse (1990). We find that if we locate $q\left(\beta\right)$ reasonably close to the mode of the posterior $\widetilde{\beta}$, the method works well.

To develop $q(\beta)$, we begin by constructing the matrix $P = \widetilde{\beta}\widetilde{\beta}' + \tau\widetilde{\beta}_{\perp}\widetilde{\beta}_{\perp}'$ where $\widetilde{\beta}_{\perp}'\widetilde{\beta} = 0$ and $\widetilde{\beta}_{\perp}'\widetilde{\beta}_{\perp} = I_{n-r}$, and τ is small (we use $\tau = 0.1$)[12]. Then we take

$$q(\beta) = \left|\beta' P^{-1}\beta\right|^{-n/2} \frac{\tau^{-(n-r)r/2}}{\mathfrak{c}_{\beta}} \tag{12}$$

where $\mathfrak{c}_{\beta}$ is defined in (3). Note that we could use a Uniform distribution such as the prior for $q(\beta)$ by setting $\tau = 1$, but we find the estimates of $\widehat{\mathfrak{c}_{\omega}}$ are less stable in this case.

4 Empirical Application

In this section we provide empirical evidence on the role of permanent shocks in U.S. consumption (c_t), investment (i_t) and income (inc_t) as studied by KPSW. The KPSW study proposes these variables are subject to a single common permanent productivity shock and that the consumption/income and investment/income ratios are stable. They also report evidence that the bulk of the fluctuations in these variables is due to the permanent shock. Using an extended data set up to and including July 2005[13], we report evidence upon the number of common permanent shocks, the support for the stability of the consumption/income and investment/income ratios as implied by the KPSW model, and the proportion of variability in the three variables in the system $y_t = (c_t, i_t, inc_t)$ over the business cycle that is due to permanent shocks. Finally, we report full densities of impulse responses to permanent shocks to demonstrate the importance of model uncertainty.

4.1 Evidence on Permanent Shocks and the 'Great Ratios'.

KPSW translate the above features of the system of variables into restrictions upon a VECM and investigate the support for these restrictions. These model restrictions are that there is one common stochastic trend and $c_t - inc_t$ and

[12] In earlier work we used τ drawn from $N\left(0, \sigma^2\right)$ with small σ. This placed the location of $q(\beta)$ on the posterior mode. However, this added a step to the sampling scheme that did not markedly improve estimation.

[13] The data are quarterly covering the period from the first quarter 1951 to the second quarter of 2005, on Personal Consumption Expenditures, Gross Private Domestic Investment, and GDP (Source: Bureau of Economic Analysis).

$i_t - inc_t$ will both be stationary $I(0)$ processes. We therefore allow the rank, r, to vary over all possible values, $r \in [0, 1, \dots, n]$ and for 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 $r = 1$). Finally we also allow for the range of five combinations of deterministic processes suggested in Section 2. An additional feature of the model of KPSW is that if $c_t - inc_t$ and $i_t - inc_t$ are stationary, we would not expect them to contain trends. Thus we would expect the evidence to suggest $d < 2$. The set of 120 models may be summarised as $r \in [0, 1, 2, 3]$, $o \in [0, 1]$, $d \in [1, 2, 3, 4, 5]$ and $l \in [5, 6, 7, 8]$.[14]

Beginning with the support for the alternative models in the model set, the modal model with posterior probability of 78%, has six lags of differences, one stochastic trend ($r = 2$), the great ratios do not form the cointegrating relations, $o = 0$, and the equilibrium relations and the levels contain deterministic trends ($d = 2$). The posterior probabilities of the models (averaged over lags) are given in Table 1. These results show that both with and without the overidentifying restrictions, the weight of support is upon there being one common stochastic trend in y_t ($p(r = 2|y) = 91\%$), with some support for a second stochastic trend ($p(r = 1|y) = 8.2\%$). This result gives substantial support to the first feature suggested by the model proposed in KPSW, that these variables share a single permanent shock. The second feature, that $c_t - inc_t$ and $i_t - inc_t$ are cointegrating relations, however, has a posterior probability of only 9.1%. These two conclusions agree with the findings of Centoni and Cubadda (2003) (hereafter CC) who use an extended data set to April 2001. Finally, we also find strong evidence that the equilibrium relations are $I(0)$ with linear deterministic trends as $p(d = 2|y) = 87.9\%$.[15]

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

[14] Simply multiplying up the cardinality of each set of (r, o, d, l) would produce 160 models. However, several models are impossible and so excluded, or observationally equivalent to another and so we count these as one model. See Section 3.1 for discussion on this point.

[15] The results $p(o = 1|y) = 0.91$ and $p(d = 2|y) = 0.789$ most likely reflect the (probably temporary) fall in the savings ratio and the rise in the investment ratio towards the end of the 1990s which is very evident in the data. Thus these conclusions are possibly sample dependent. The issue of structural breaks are not considered in this demonstration, but we note it as a possible direction for a more serious investigation of this issue.

Just Identified Models ($o = 0$)

r	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
0	0.001	0.001		0.001	
1	0.016	0.012	0.002	0.009	0.014
2	0.005	0.808	0.013	0.009	0.013
3	0.001		0.001		0.001
Over Identified Models ($o = 1$)					
1	0.013	0.005	0.004	0.003	0.004
2	0.001	0.055	0.001	0.003	0.001

4.2 Effects on Permanent Shocks.

Next we consider the importance of the permanent shocks in the business cycle. Decomposing the variance into the components due to transitory and permanent shocks, we gain an impression of the relative importance of these effects for the variability of the consumption, investment and income. KPSW derive an identification scheme for this decomposition based upon a particular economic theory. In our data there is uncertainty associated with this theory. Therefore, we use the approach of CC which produces a decomposition without the need for a particular identification scheme.

KPSW estimate the proportion of variance of due to permanent shocks in the time domain for the model $M_{(2,1,3)}$ with 8 lags of differences. For , i_t and inc_t they report proportions varying from 0.88 (c_t), 0.12 (i_t) and 0.45 (inc_t) at one quarter after the shock to 0.89 (c_t), 0.47 (i_t) and 0.81 (inc_t) respectively at 24 quarters after the shock. Our interest is in the proportion of business cycle fluctuations due to permanent shocks and so follow CC who consider the variance decomposition within the frequency domain.

With their slightly shorter sample, CC found proportions of variability over an 8-32 quarter period of 0.574 for c_t, 0.139 for i_t and 0.181 for inc_t. Table 2 reports the proportions of fluctuations over 8 to 32 quarters that are due to permanent shocks for the three variables using our updated data set and extended model set. We see from these results that the KPSW model assigns a larger proportion of the variability in consumption and investment to the permanent (productivity) shock than the other models. The remaining models generally agree with each other, at least in the relative sizes if not the exact values. Thus, using our Bayesian model averaging approach we find support for the conclusion of CC that the single permanent shock is not the main determinant of business cycle fluctuations.

Table 2: Estimated variance decompositions into permanent components in the frequency domain.

Esimation method	c_t	i_t	inc_t
Averaged over all models	0.168	0.540	0.212
CC model $M_{(2,0,3)}$	0.187	0.376	0.262
KPSW model $M_{(2,1,3)}$	0.251	0.360	0.300
Best model $M_{(2,0,2)}$	0.165	0.554	0.202

We conclude by reporting for each variable the impulse response path from a permanent shock. We assume there is only one permanent shock (and so condition upon $r = 2$), but average over the other model features. The impulses for c_t, i_t and inc_t are shown in Figures 1, 2 and 3 respectively. The upper panel in each figure shows the full density over all 60 periods. The bands represent the boundaries of 20%, 40%, 60% and 80% highest posterior density regions (HPDs). These are contours of the density that define the smallest possible regions containing the stated mass. To aid with the interpretation of these figures we have included in the lower panels the profiles of the density of the impulses at three points in time after the shock. These are at $h = 10$, $h = 30$ and $h = 60$ periods after the shock.

We see that the 20% and 40% HPDs are very sensitive to changes in the shape of the density and so reflect small movements in the bulk of the mass. The 60% and 80% HPDs are less sensitive and tend to show the general direction of the response. The lower panels show the changes in the shape of the densities that cause the movements in these HPDs. In each case there are two or more paths that influence the densities at different intervals.

For consumption, the peak near zero slowly loses prominence to the more disperse, higher mass as the period increases. For investment, there are three paths at early stages, with the higher path dominating. Over time, however the central path becomes the only important path. Looking at the response path of income, we see there are two separate paths that compete immediately after the shock, and they both remain important as we move out along the time horizon. The higher path, however, moves slightly down toward zero and becomes more significant.

The form of these densities are important for giving a full account of the uncertainty associated with the responses. In each case the secondary (or more) paths derive from models with low posterior probabilities. However, neglecting these models and only using the best model (effectively assuming model certainty) would produce very different estimates of, say, expected loss from a particular action.

5 Conclusion.

In this paper we have presented a Bayesian approach to obtaining 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 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, short-run dynamics and overidentifying restrictions upon the cointegrating space. The estimates are derived using a mixture of analytical integration and MCMC. We apply the methodology to investigating the importance and effect of permanent shocks in US macroeconomic variables, with a focus upon the support for the behaviour implied by the model KPSW.

The method presented in this paper has already found applications in several areas. Koop, Potter and Strachan (2005) investigate the support for the hypothesis that variability in US wealth is largely due to transitory shocks. They demonstrate the sensitivity of this conclusion to model incertainty. Koop, León-González and Strachan (2006) develop methods of Bayesian inference in a flexible form of cointegrating VECM panel data model. These methods are applied to a monetary model of the exchange rate commonly employed in international finance. Other current work includes investigating the impact of oil prices on the probability of encountering the liquidity trap in the UK and stability of the money demand relation for Australia.

More recent work is looking to develop methods of inference in very large model sets (as occurs in, say, models with the additional dimension of an unknown number of regime shifts) using the reversible jump methodology proposed by Greene (1995).

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

6 Acknowledgements

Preliminary versions of this paper have been presented at, among others, the ESEM-2002, the EC^2 2002, Tokyo University, CORE, the Swedish central bank, University of Leuven. We would like to thank Luc Bauwens, Geert Dhaene, David Hendry, Soren Johansen, Helmut Lutkepohl, Christopher Sims, and Mattias Villani for helpful discussions on the topic of this paper. We are heavily indebted to the editor and two anonymous referees for several constructive suggestions on an earlier version (Strachan and van Dijk, 2004) that have greatly improved the paper. Of course, any remaining errors remain the responsibility of the authors. Strachan acknowledges the assistance of funding from the University of Liverpool grant number 4128 and study leave from the University of Leicester, van Dijk acknowledges financial support from the Erasmus Research Institute for Management.

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 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," *Biometrika* **44**, 533–534.

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

Canova, F. and C. Matteo (2004): "Forecasting and turning point predictions in a Bayesian panel VAR model," *Journal of Econometrics* **120**, 327-359.

Centoni, M. and Cubbada, G. (2003): "Measuring the business cycle effects of permanent and transitory shocks in cointegrated time series," Economics Letters 80, 45-51.

Chang, Y. (2000): "Vector Autoregressions with Unknown Mixtures of I(0), I(1), and I(2) Components," *Econometric Theory* **16**, 905-26.

Chang, Y. and P. C. P. Phillips (1995): "Time Series Regression with Mixtures of Integrated Processes," *Econometric Theory* **11**, 1033-94.

Chikuse, Y. (1990): "The Matrix Angular Central Gaussian distribution," *Journal of Multivariate Analysis* **33**, 265-274.

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

Davidson, R. and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*, , New York, Oxford University Press.

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

Doan, T. ,R. Litterman, and C Sims (1984): "Forecasting and conditional projections using realistic prior distributions," *Econometric Reviews* **3(1)**:1-100.

Draper, D. (1995): "Assessment and propagation of model uncertainty (with discussion)," *Journal of the Royal Statistical Society Series B* **56**, 45-98.

Drèze, J.H. (1977): "Bayesian Regression Analysis using Poly-t Densities," *Journal of Econometrics* **6**, 329-354.

Elliott, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica* **66**, 149-158.

Elliott, G. and J. H. Stock (1994): "Inference in Time Series Regression When the Order of Integration of a Regressor Is Unknown," *Econometric Theory* **10**, 672-700.

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

Fernández, C., E. Ley, and M. Steel (2001): "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics* **16**, 563-576.

Gelfand, A.E., and D. K. Dey (1994): "Bayesian model choice: asymptotics and exact calculations," *Journal of the Royal Statistical Society Series B* **56**, 501–504.

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

Greene, Peter J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika* **82**, 711-732.

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

Hodges, J. (1987): "Uncertainty, policy analysis and statistics," *Statistical Science* **2**, 259-291.

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," *Journal of Econometrics* **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.

Koop, G. (1991): "Cointegration tests in present value relationships: A Bayesian look at the bivariate properties of stock prices and dividends," *Journal of Econometrics* **49**, 105-140.

Koop, G. (1994): "An objective Bayesian analysis of common stochastic trends in international stock prices and exchange rates," *Journal of Empirical Finance* **1**, 343-364.

Koop G., R. León-González and R. Strachan (2006): "Bayesian Inference in a Cointegrating Panel Data Model," Department of Economics Working Paper 06/02, University of Leicester.

Koop G., R., S. Potter and R. Strachan (2005): "Re-examining the Consumption-Wealth Relationship: The Role of Model Uncertainty," Department of Economics Working Paper 05/03, University of Leicester.

Leamer, E. (1978): *Specification Searches.* New York: Wiley.

Lettau, M. and S. Ludvigson (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review* **94**, 276-299.

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.

Maddala, G.S. and I.M.Kim (1998): *Unit roots, Cointegration and Structural Change*, Cambridge: Cambridge University Press.

Min, C. and A. Zellner (1993): "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics* **56**, 89-118.

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.

Ni, S. X. and D. Sun (2003): "Noninformative Priors and Frequentist Risks of Bayesian Estimators of Vector-Autoregressive Models," *Journal of Econometrics* **115**, 159-197.

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

Phillips, P. C. B. (1996): "Econometric Model Determination," *Econometrica* **64**, 763–812

Poirier, D. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: The MIT Press.

Raftery, A. E., D. Madigan, and J. Hoeting (1997): "Bayesian model averaging for linear regressionmodels," *Journal of the American Statistical Association* **92**, 179–191.

Sala-i-Martin, X., G. Doppelhoffer, and R. Miller (2004): "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach," *American Economic Review* **94**, 813-835.

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

Sims, C. A. and T. Zha (1999) "Error Bands for Impulse Responses," *Econometrica* **67**, 1113-1156.

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," *Journal of Econometrics* **123**, 307-325.

Strachan, R and H. K. van Dijk (2003): "Bayesian Model Selection with an Uninformative Prior," *Oxford Bulletin of Economics and Statistics* **65**, 863-876.

Strachan, R. and H.K. van Dijk (2004) "Valuing structure, Model Uncertainty and Model Averaging in Vector Autoregressive Processes," Econometric Institute Report EI 2004-23, Erasmus University Rotterdam.

Strachan, R and H. K. van Dijk (2005): "Weakly informative priors and well behaved Bayes Factors," Econometric Institute Report EI 2005-40, Erasmus University Rotterdam.

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.

Villani, M. (2005): "Bayesian reference analysis of cointegration," *Econometric Theory* **21**, 326-357.

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

8 Appendix

In this appendix we provide the technical details for statements in the paper. First we must introduce some notation for matrix spaces and measures on these spaces. For an introduction to these concepts see Muirhead (1982) and for a more intuitive discussion see Strachan and Inder (2004). We assume throughout this appendix that $d_{1,t} = \{\}$ such that $\beta^+ = \beta$.

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 $C \in O(r)$. 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,[16] $G_{r,n-r}$. That is $\mathfrak{p} = sp(V) \in G_{r,n-r}$ and 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

[16]The Grassman manifold, $G_{r,n-r}$, is the collection of all possible $r-$ dimensional planes in the $n-$ dimensional real space. Thus $G_{r,n-r} \subset R^n$.

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 $[b_1, b_2, \ldots, b_r] \in V_{r,n}$, $r < n$, 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 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 $\int_{G_{r,n-r}} \underline{dg_r^n} = 1$.

Theorem 2 *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*

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

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

Proof. 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 an 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

$$\begin{aligned} dv_r^n &= dg_r^n dv_r^r \\ \text{or } \underline{dv_r^n} &= \underline{dg_r^n}\, \underline{dv_r^r}. \end{aligned} \tag{14}$$

James[17] obtains the volume of $G_{r,n-r}$ as

$$\begin{aligned} \mathfrak{c}_\beta &= \int_{G_{r,n-r}} dg_r^n = \frac{\int_{V_{r,n}} dv_r^n}{\int_{O(r)} dv_r^r} \\ &= \pi^{(n-r)r} \Pi_{j=1}^r \frac{\Gamma\left[(r+1-j)/2\right]}{\Gamma\left[(n+1-j)/2\right]}. \end{aligned} \tag{15}$$

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

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 (14) 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

$$\underline{dv^n_r} = \pi^{-(n-r)r}\Pi^r_{j=1}\frac{\Gamma\left[(n+1-j)/2\right]}{\Gamma\left[(r+1-j)/2\right]}\left|I_r + \overline{\beta}'_2\overline{\beta}_2\right|^{-n/2} d\overline{\beta}_2\left(\underline{dv^r_r}\right) \tag{16}$$

(Phillips, 1989, 1994).

Equating (14) and (16) gives the result. Another, slightly more general proof for the same result is presented in Chikuse (1998).■

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 (14) and the discussion in the proof of Theorem 2 provides a natural solution to this question. That is use $\beta \in V_{r,n}$ 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 (14), 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 $\mathfrak{p}$.

Next we provide a proof that linear identifying restrictions with a flat prior give zero weight to the chosen linear restrictions. The Jacobian defined by (13) 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*

$$\begin{aligned} \left(d\overline{\beta}_2\right) &= \pi^{(n-r)r}\Pi_{j=1}^{r}\frac{\Gamma\left[(r+1-j)/2\right]}{\Gamma\left[(n+1-j)/2\right]}\left|I_r+(c\beta)'^{-1}\beta' c_\perp' c_\perp\beta\,(c\beta)^{-1}\right|^{n/2}(dg_r^n) \\ &= J\ dg_r^n. \end{aligned} \tag{17}$$

Proof. Invert (16) and replace $\overline{\beta}_2$ by $c_\perp\beta\,(c\beta)^{-1}$.■

The following proof demonstrates the claim in Section 3.2 that assuming we know which rows of β are linearly independent so as to impose linear identifying restrictions makes this assumption *a priori* impossible.

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: Essentially, the Jacobian for $\overline{\beta}_2 \to \mathfrak{p}$ places infinitely more weight in the direction where $c\beta$ is singular. Thus, normalisation of β by choice of c with a flat prior on $\overline{\beta}_2$ implies infinite prior odds against this normalisation.

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 (17) 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.

To support the use of model averaging in this application, we provide here proofs that the posterior will be proper and all finite moments of β exist. Since $\mathfrak{g}_\omega$ (in (8)) is finite for the class of priors considered, for the Bayes factor to be 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:*

$$\begin{aligned} p(\beta | M_\omega, y) &\propto |\beta' D_0 \beta|^{-T/2} |\beta' D_1 \beta|^{(T-n)/2} \\ &= k_\beta(\beta | M_\omega, y) \end{aligned} \tag{18}$$

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

Theorem 6 *The marginal posterior density for β conditional upon ω in (18) 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

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

for $m = 0, 1, 2, \ldots$ is bounded above almost everywhere by the finite integral $\mathcal{M} \int_{-1}^{1} |b_{ij}|^m db_{ij}$. 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 2 above, dg_r^n is integrable and therefore bounded above almost everywhere by some finite constant, $\mathcal{M}_1$.

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, $\mathcal{M}_2$. Thus $k_{\beta}(\beta)\, dg_r^n$ has a finite upper bound, $\mathcal{M} = \mathcal{M}_1\mathcal{M}_2$. 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).■

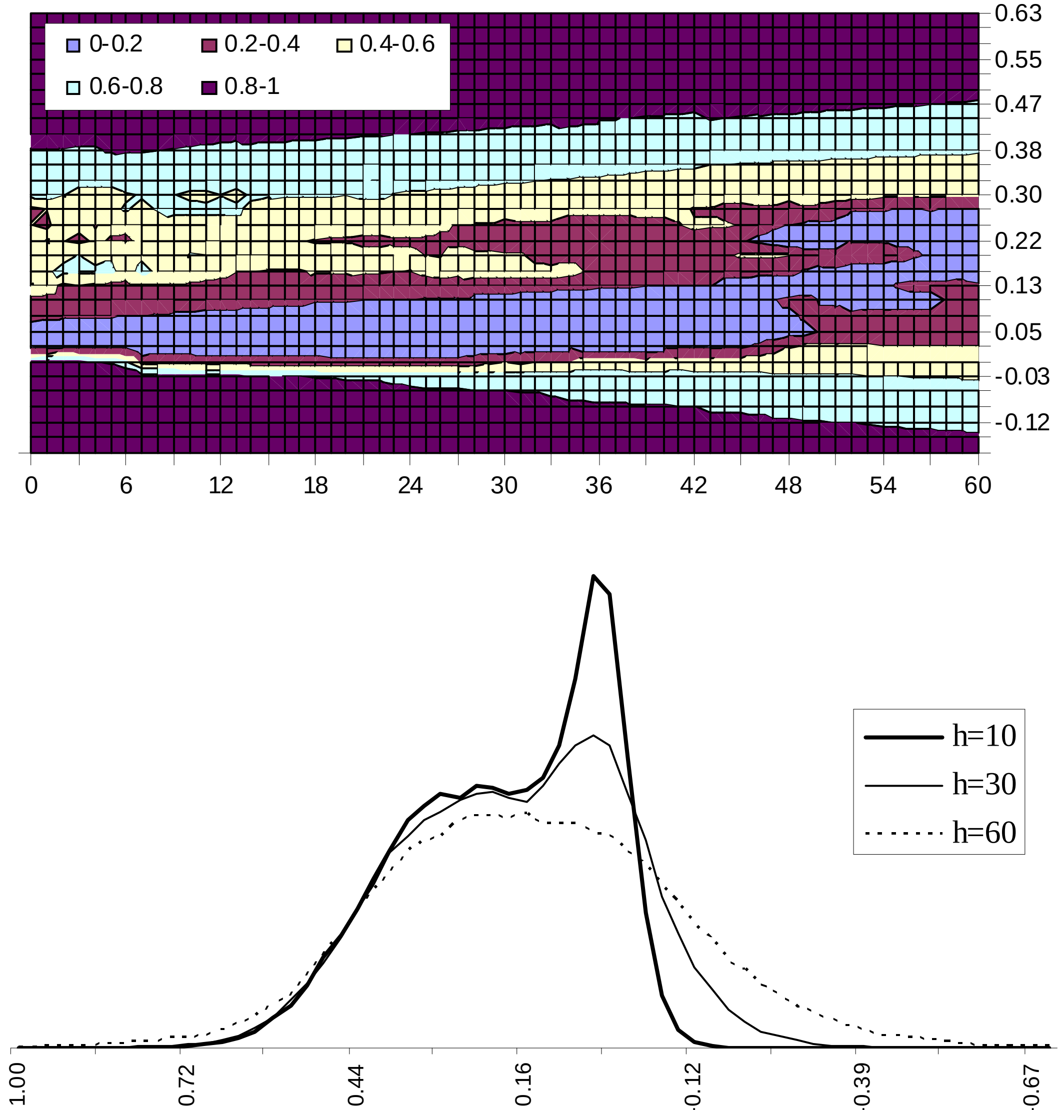

Figure 1: This figure shows the densities over 60 periods of the impulse responses of **consumption** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse response at $h = 10$, 30 and 60 periods into the future.

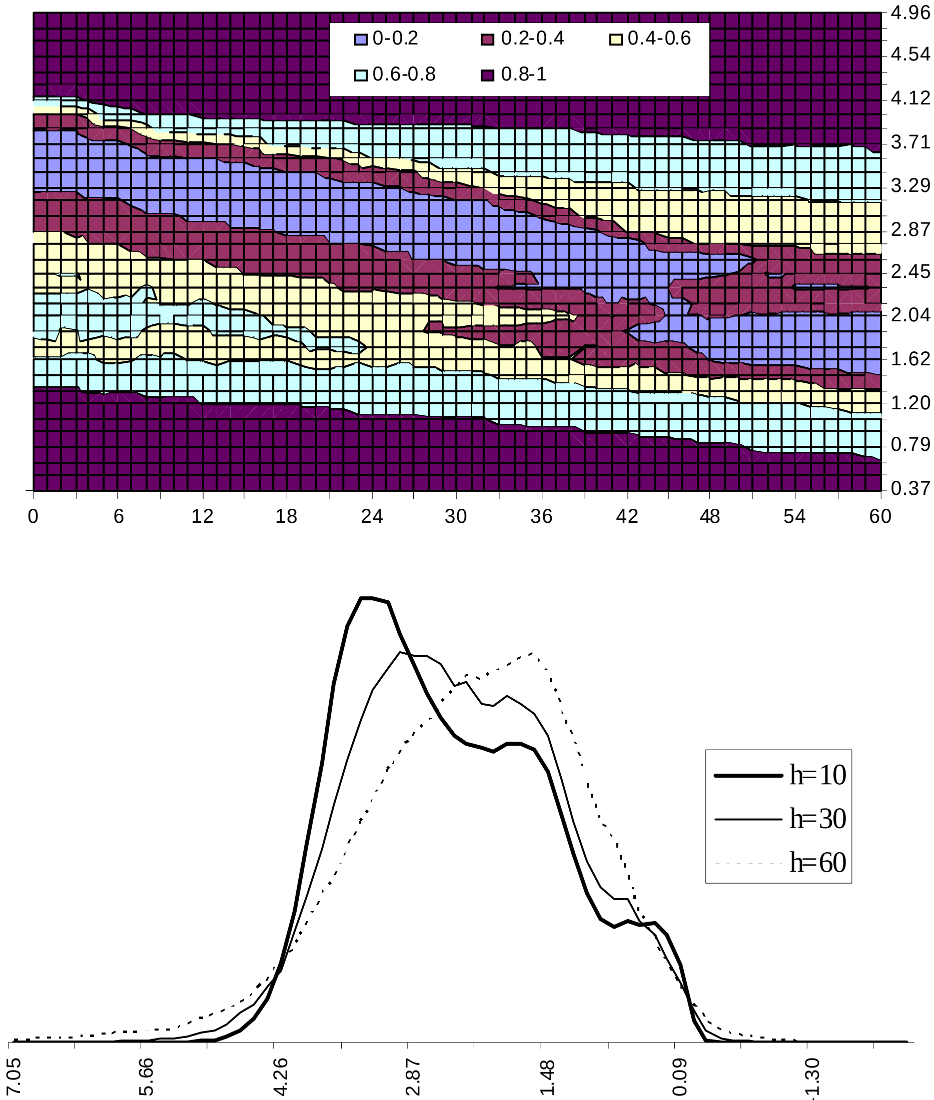

Figure 2: This figure shows the densities over 60 periods of the impulse responses of **investment** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse response at $h = 10$, 30 and 60 periods into the future.

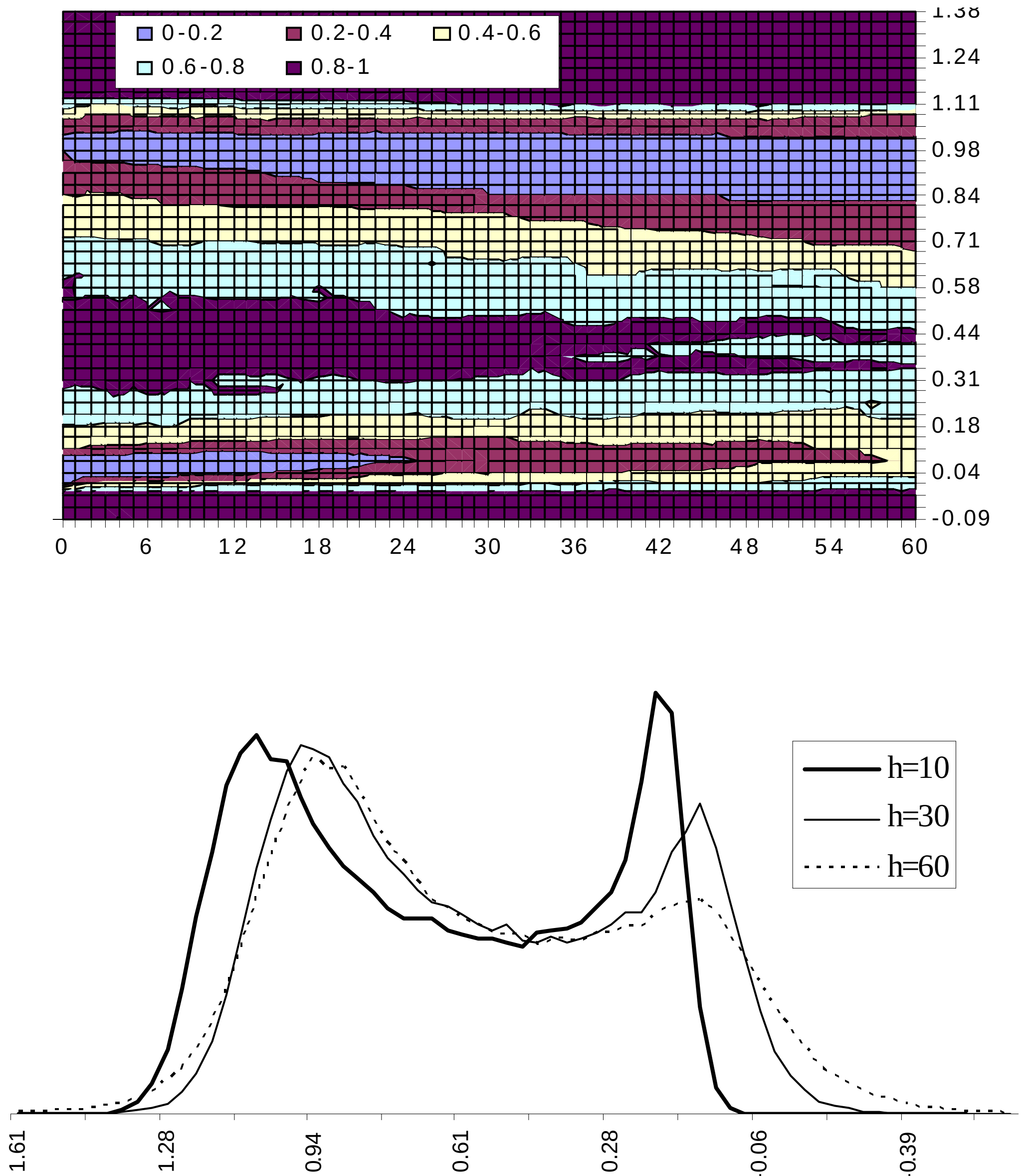

Figure 3: This figure shows the densities over 60 periods of the impulse responses of **income** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse response at $h = 10$, 30 and 60 periods into the future.