TI 2008-096/4
Tinbergen Institute Discussion Paper

Bayesian Averaging over Many Dynamic Model Structures with Evidence on the Great Ratios and Liquidity Trap Risk

Rodney W. Strachan[1]
Herman K. van Dijk[2]

[1] *The University of Queensland, Australia;*
[2] *Econometric Institute, Erasmus University Rotterdam, and Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Bayesian Averaging over Many Dynamic Model Structures with Evidence on the Great Ratios and a Liquidity Trap Risk.

Rodney W. Strachan[1*]
Herman K. van Dijk[2]
[1]School of Economics,
The University of Queensland,
St Lucia 4072, Australia.
email: r.strachan@uq.edu.au
[2]Econometric and Tinbergen Institutes,
Erasmus University Rotterdam,
Rotterdam, The Netherlands
email: hkvandijk@few.eur.nl

October 2, 2008

*Corresponding author.

ABSTRACT

A Bayesian model averaging procedure is presented that makes use of a finite mixture of many model structures within the class of vector autoregressive (VAR) processes. It is applied to two empirical issues. First, stability of the "Great Ratios" in U.S. macro-economic time series is investigated, together with the effect of permanent shocks on business cycles. Second, the linear VAR model is extended to include a smooth transition function in a (monetary) equation and stochastic volatility in the disturbances. The risk of a liquidity trap in the USA and Japan is evaluated. Although this risk found to be reasonably high, we find only mild evidence that the monetary policy transmission mechanism is different and that central banks consider the expected cost of a liquidity trap in policy setting. Posterior probabilities of different models are evaluated using Markov chain Monte Carlo techniques.

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

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 (VAR) processes. We demonstrate the operational implications of our approach by investigating two empirical issues. First, the stability of the "Great Ratios" in U.S. consumption, investment and income is investigated, together with the presence and effects of permanent shocks for the duration of the business cycle. Second, the VAR model is extended to include a smooth transition function in a (monetary) equation and stochastic volatility in the disturbances. The risk of a liquidity trap in the USA and Japan is evaluated as is the evidence that central banks incorporate the expected cost of a liquidity trap in setting policy. We take evidence that the transmission mechanism is different when there is a high probability of hitting a liquidity trap in the near future as evidence that the monetary authority has incorporated the risk of the liquidity trap in its reaction function.

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), only a few 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) 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.

Summarizing, 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 forecasting and 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), Terui and van Dijk (2002) and more recently by Francesco, Van Dijk and Verbeek (2007). Some explanation for this phenomenon in particular cases was provided by Hendry and Clements (2002). 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.

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

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

We focus in this paper 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 whose moments exist. 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, the forms of those equilibrium relationships, the dynamic responses to disequilibria, and the type of deterministic processes that may be present and the lag structure. In order to obtain these results we make use of the prior developed in Strachan and Inder (2003). This prior uses manifolds and orthogonal groups and their measures. With these techniques 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 for prior elicitation and an efficient Markov chain Monte Carlo technique for simulating from the posterior, 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. In other words, our forecasts are based on a finite mixture of model structures.

Third, we demonstrate the proposed methodology with an empirical investigation of two economic issues. First, the stability of the 'Great Ratios' − as discussed in King, Plosser, Stock and Watson (1991) (hereafter KPSW) −, the relative weights of permanent and transitory components in US consumption, investment and income, the importance of permanent shocks for the presence of business cycles, and, finally, the credibility of alternative paths of responses to a possible productivity shock are investigated. Second, the linear structural VAR model is extended to include a smooth transition function in a (monetary) equation and stochastic volatility in the disturbances. Within this extended VAR model, the risk of a liquidity trap in the USA and Japan is estimated and we investigate the evidence that the expected cost of a liquidity trap influences monetary authorities' policy setting decisions.

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), Strachan and van Dijk (2003), Strachan and Inder (2004), Villani (2005), Koop, Potter and Strachan (2008), Koop, Léon-Gonzalez and Strachan (2008a and 2008b) using diffuse type of priors. Cogley and Sargent (2005), Primiceri (2005) and Sims and Zha (2006) specify a VAR with stochastic volatility. We extend the analysis of these different lines of research 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 basic models of interest in this paper - the vector autoregressive models, the general structural features of interest, and the restrictions they imply. These models are used in the first empirical application but we extend them in the second to account for a wider range of behaviours. 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 highest posterior density intervals for features of interest like impulse responses. We demonstrate the approach in Sections 4 and 5 with an investigation of the two empirical economic issues mentioned before. The first application employs the models from Sections 2 and 3 directly. The second application builds upon these models to permit stochastic volatility and a smooth transition in central bank reaction functions. In Section 6 we summarize conclusions and discuss possibilities for further research.

2 A Set of Vector Autoregressive Model Structures.

Since the influential work by Sims (1980), the class of vector autoregressive (VAR) models has enjoyed much success in macroeconometrics: it can incorporate a wide range of short and long run dynamic, structural and deterministic behaviour.

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 feature into the VAR model and allows the separation of long run and short run behaviour. For 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 l initial observations 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_j $j = 1, \ldots, l$ are $n \times n$ and β^+ and α' are $n \times r$ and assumed to have rank r, and if $r = n$ then $\beta^+ = I_n$. The $1 \times n$ vector of errors ε_t are assumed to be $iidN(0, \Omega)$.[3] We define the deterministic terms $d_{i,t}\theta_i$, $i = 1, 2$, formally below.

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

$$Z_0 = Z_1\beta\alpha + Z_2\Phi + E = ZA + E. \quad (3)$$

Vectorising this expression we have

$$z_0 = za + e \quad (4)$$

where $z_0 = vec(Z_0)$, $z = (I_n \otimes Z)$, $a = vec(A)$ and $e = vec(E)$.

Next, we specify the restrictions of interest, combinations of which define (in our notation) different model structures of interest which we may compare or weight using posterior probabilities. The restrictions refer to the *number* of equilibrium relations, to the *structural (over)identification* restrictions of these relations, to particular types of *deterministic* processes and to the *lag* length.

[3]Throughout the paper, we denote the Normal distribution with mean m and covariance matrix c by $N(m, c)$.

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 are 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 ourselves to two cases: no restriction on β $(o = 0)$; and $\beta = H\psi$ $(o = 1)$ where H is an $n \times s$ matrix and 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. For example, KPSW study stability of the 'Great Ratios' of consumption to income and investment to income as cointegrating relations. If we have logs of consumption (c_t), investment (i_t) and income (inc_t), and there are two cointegrating relations, then the 3×2 (with $s = r = 2$ in this case) matrix

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

defines the two cointegrating relations among $y = (c_t, i_t, inc_t)$ by

$$y_t H = (c_t, i_t, inc_t) \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} = (c_t - inc_t, i_t - inc_t).$$

We investigate this example further in Section 4.

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

We set the range of l by setting a minimum, $L_{\min}$, and a maximum number of lags, L. For example, we may have $l = 0, 1, ..., L$ so $L_{\min} = 0$ for a total of $N_L = L - L_{\min} + 1$ lags considered.

Each model will be identified by M_ξ where $\xi = (r, o, d, l)$ and $\xi \in \Xi$, the set of all ξ considered. For example, the least restricted model will be $M_{(n,0,1,L)}$, while the most restricted model will be $M_{(0,1,5,0)}$. 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, $M_{(n,0,3,6)}$, and the second having a constant and a trend, $M_{(n,0,1,6)}$. 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 $M_{(2,1,3,6)}$. Thus, with $n = 3$ in our application, we deal with a case of $4 \times 2 \times 5 \times N_L = 40N_L$ models.[4,5] While we do allow for a range of *lags* of differences, as these have little economic importance for the studies we look at, and for space considerations, we later denote some models by the shorter notation (r, o, d) and by this we mean that the reported results will have been averaged over all lag lengths.

[4]This reduces to $26(N_L - 1)$ models when we account for impossible models and observationally equivalent models. See Subsection 3.1 below for further discussion on this point.

[5]Generally, if we consider L different lag lengths the number of observationally distinct models is $L\,(1 + 5\,(n + s))$.

3 Priors, Posteriors and Model Averaging.

In this section the priors and resultant posterior are presented beginning 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 Ω and a. Conditional upon β the model in (1)-(4) is linear in the equation parameters a. This fact makes it relatively straightforward to elicit priors on Ω and a, however we adopt a transformation that improves the sampling scheme and so give the full prior after we have given careful consideration to the prior for β, before then presenting the method of posterior analysis.

3.1 The Prior.

In this paper we wish to treat all models as *a priori* equally likely, however this is not a straightforward issue.[6] The priors for the individual elements of $\xi = (r, o, d)$ are not independent, as certain combinations are either impossible, meaningless (such as, for example, $r = 0$, that is we have no stable relations, 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[7]. However, the researcher must carefully consider how she wishes to treat observationally equivalent models.

It seems sensible to regard this set of observationally equivalent models as just one model and then assign equal prior probabilities to all these models. For example, at $r = 0$ the models with $d = 2$ and $d = 3$ are observationally equivalent. If we were to treat these two models as one model, they each 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.[8]

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

[7] 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.

[8] This issue could be viewed as a conflict between the desire to be uninformative across statistical models and the desire to be uninformative across economic models.

Alternatively we could specify all possible combinations of the indices in ξ 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 $k(M_\xi) = p(r)\,p(o)\,p(d)\,p(l)$. Next, set $k(M_\xi) = 0$ for impossible combinations and for each set of combinations of ξ that imply observationally equivalent models, we set $k(M_\xi) = 0$ for all but one of the combinations. Finally we compute the prior model probabilities as $p(M_\xi) = k(M_\xi)/\Sigma_\xi k(M_\xi)$ where in the denominator we have summed $k(M_\xi)$ over all ξ.

To demonstrate the assignment of prior probabilities we use the first application in this paper. As we have $n = 3$, $r \in [0, 1, 2, 3]$ so we use $p(r) = (n+1)^{-1} = 0.25$, with $d \in [1, 2, 3, 4, 5]$ we set $p(d) = 0.2$, and with $l \in [0, 1, \ldots, L]$ we set $p(l) = \frac{1}{L+1}$. 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 $p(o) = 0.5$ for $o \in [0, 1]$.

For each model implied by a particular value of ξ, we need to specify a prior for the parameters in the model. We use a proper inverted Wishart with scale matrix $\underline{S} = I_n 10$ and degrees of freedom $\underline{\nu} = 10$ as the prior for Ω and this is rather uninformative. As a changes dimensions across the different versions of ξ implied by different models and each element of the vector a has the real line as its support, the Bayes factors for different models will not be well defined if an improper prior on a, such as $p(a|\beta, M_\xi) \propto 1$ were used.[9] For this reason a weakly informative proper prior for a must be used. We defer giving an expression for the full prior to the end of the next subsection, but the prior for a conditional upon (Ω, β, M_ξ) has zero mean and covariance matrix $\underline{V} = \Omega \otimes \eta^{-1} I_{n(k_i+r)}$.[10] 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).

As β and α appear as a product in (2), r^2 restrictions need to be imposed

[9] 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).

[10] 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).

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 $\overline{\beta} = \left[I_r \; \overline{\beta}_2' \right]'$ and a prior is then specified for $\overline{\beta}_2$.[11]

We also note that 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 so 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 4).

Assuming that c is known, Kleibergen and van Dijk (1994 & 1998) (compare also Bauwens and Lubrano, 1996) 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 Markov Chain Monte Carlo (MCMC) methods 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: 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, although the posterior again has no first or higher-order moments of $\overline{\beta}_2$.

As is indicated before, a flat prior on $\overline{\beta}_2$ cannot be employed to obtain posterior probabilities for M_{ξ} since the dimensions of $\overline{\beta}_2$ depend upon ξ. We do not, however, 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[12]

[11] 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.

[12] The authors would like to thank Soren Johansen for making this point to one of the author's. Villani (2005) also makes use of a prior on $\mathfrak{p}$.

$G_{r,n-r}$ (James, 1954), $\mathfrak{p} \in G_{r,n-r}$.

We save the technical discussion for the Appendix, but to implement this approach, we specify β to be semi-orthogonal, i.e., $\beta'\beta = I_r$, and specify a Uniform distribution for β (for some background information, see Strachan (2003), Strachan and Inder (2004) and Strachan and van Dijk (2003)). The support of all $n \times r$ dimensional semiorthogonal matrices is the Steifel manifold which we denote by $V_{r,n}$.

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 $p(\beta|M_\xi) = \mathfrak{c}_\beta^{-1}$ where[13] $\mathfrak{c}_\beta = \int_{G_{r,n-r}} d\beta$ 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 the group of $r \times r$ orthogonal matrices, see the Appendix and the reference given there.[14]

For the cases in which we impose identifying restrictions discussed in Section 2 of the form $\beta = H\psi$ $(o = 1)$, we impose $\psi \in V_{r,s}$ and impose the Uniform prior on $V_{r,s}$. This implies that we are uninformative 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 $H \rightarrow H(H'H)^{-1/2}$. This transformation is innocuous since the space of H, which is the important parameter, is unchanged by this transformation.

As β is semiorthogonal, the posterior distribution will be nonstandard regardless of the form we choose for the prior. Therefore, to obtain an expression for the posterior useful for obtaining draws of β, we make use of the fact that 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 U such that $\beta\alpha = \beta U U^{-1}\alpha = \beta^*\alpha^*$. 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. We give β^*

[13] 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 $d\beta$.

[14] More recently, the topic of invariance to rescaling of the data has been raised in conversations with colleagues. Our prior is not invariant and no uniform, invariant prior exists. Such invariance gives us the virtue of being able to say that the probability of being in this region is the same after rescaling, no matter what the region. No prior in the literature has this virtue except Strachan (2003). However, while Strachan (2003) can be used for BMA only if the data dependence is ignored. While it might be worth further investigation, we do not consider invariance further here except to note that we are yet to see a Bayesian cointegration study in which it is an important issue.

a Normal prior with zero mean and covariance matrix $n^{-1}I_{nr}$. We can easily transform back to the parameters of interest via $\beta = \beta^* U^{-1}$ and $\alpha = \alpha^* U$. The prior for β^* resembles that of Geweke (1996) except that our prior implicitly specifies, in addition to a proper prior for U, that the marginal prior for $\beta = \beta^* U^{-1}$ is Uniform. The efficiency of this approach is discussed in Koop, León-González and Strachan (2008a).

We let a^* denote the vector a with the elements of α replaced by the corresponding elements of α^* and let $b^* = vec(\beta^*)$. The prior for a^* is $N\left(0, \Omega \otimes \eta^{-1} I_{(k_i+r)}\right)$. The full prior distribution for the parameters in a given model is then

$$\begin{aligned} p\left(\Omega, a^*, b^* | M_\xi\right) \quad \propto \quad & \exp\left\{-\frac{\eta}{2} a^{*\prime}\left(\Omega^{-1} \otimes I_{(k_i+r)}\right) a^* - \frac{n}{2} b^{*\prime} b^*\right\} \\ & \times |\Omega|^{-(n(k_i+2r+1)+\underline{\nu}+1)/2} \exp\left\{-\frac{1}{2} tr \Omega^{-1} \underline{S}\right\}. \end{aligned}$$

3.2 Posterior Analysis.

An expression for the posterior distribution of the parameters for any model given the data is obtained by combining the prior, $p\left(\Omega, a^*, b^* | M_\xi\right)$, with the likelihood for the data $L\left(y | \Omega, a^*, b^*, M_\xi\right)$ where y represents all data. That is,

$$p\left(\Omega, a^*, b^* | M_\xi, y\right) \propto p\left(\Omega, a^*, b^* | M_\xi\right) L\left(y | \Omega, a^*, b^*, M_\xi\right) = k\left(\Omega, a^*, b^*, M_\xi | y\right). \tag{5}$$

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 transformation $\beta\alpha = \beta U U^{-1} \alpha = \beta^* \alpha^*$ and the fact that, conditional upon b^*, the model in (3) and (4) is linear.

As the model is linear conditional upon b^*, standard results show that the posterior for a^* will be

$$a^* \sim N\left(\overline{a}, \overline{V}\right) \tag{6}$$

where $\overline{a} = \left(I_n \otimes \left(Z'Z + \eta I_{(k_i+r)}\right)^{-1} Z'\right) z_0$ and $\overline{V} = \left[V^{-1} + \underline{V}^{-1}\right]^{-1} = \Omega \otimes \left(Z'Z + \eta I_{(k_i+r)}\right)^{-1}$.

Next, in the equation (3) we vectorise $Z_1 \beta^* \alpha^*$ to obtain $vec\left(Z_1 \beta^* \alpha^*\right) = z_1 b^*$ where $z_1 = \left(\alpha^{*\prime} \otimes Z_1\right)$. Thus we can rewrite the expression in (4) as

$\widetilde{z}_0 = z_1 b^* + e$ where $\widetilde{z}_0 = vec\left(Z_0 - Z_2\Phi\right)$, and use standard results again to show the posterior for b^* will be

$$b^* \sim N\left(\overline{b}^*, \overline{V}_{b^*}\right) \tag{7}$$

where $\overline{b}^* = \overline{V}_{b^*}\left(\alpha^*\Omega^{-1} \otimes Z_1'\right)\widetilde{z}_0$ and $\overline{V}_{b^*} = \left[\left(\alpha^*\Omega^{-1}\alpha^{*\prime} \otimes Z_1'Z_1\right) + nI_{nr}\right]^{-1}$.

We use the following scheme at each step i to obtain draws of (a^*, Ω, β^*).

1. Initialize $(a^*, \Omega, b^*) = \left(a^{*(0)}, \Omega^{(0)}, b^{*(0)}\right)$.
2. Draw $\Omega|a^*, b^*$ from $IW\left(\underline{S} + \eta A'A + E'E, T + n\left(k_i + 2r + 1\right) + \nu\right)$
3. Draw $a^*|\Omega, b^*$ from $N\left(\overline{a}, \overline{V}\right)$
4. Draw $b^*|\Omega, a^*$ from $N\left(\overline{b}_{\beta^*}, \overline{V}_{\beta^*}\right)$.
5. Repeat steps 2 to 4 for a suitable number of replications.

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

$$p\left(M_i|y\right) = \frac{m_i p\left(M_i\right)}{\sum_{\xi\in\Xi} m_\xi p\left(M_\xi\right)} = \frac{m_i/m_0 p\left(M_i\right)}{\sum_{\xi\in\Xi} m_\xi/m_0 p\left(M_\xi\right)} \tag{8}$$

where the summation in the denominator is over all elements of Ξ and the marginal likelihood m_0 is for some model M_0. The marginal likelihood for a model is given by

$$m_\xi = \int_{R^{(k_i+r)n}} \int_{\Omega>0} \int_{G_{r,n-r}} k\left(\Omega, a^*, \beta^*, M_\xi|y\right)\left(db^*\right)\left(d\Omega\right)\left(da^*\right), \tag{9}$$

where $a^* \in R^{(k_i+r)n}$, Ω is positive definite (denoted $\Omega > 0$). The expression in (8) suggests two ways to compute the model probabilities. We could either compute the m_ξ directly and use the first expression, or we could compute the ratio m_ξ/m_0 for each model and use the second expression.

If M_0 nests within all of the models in the model set (M_0 need not actually be in the model set considered) then we can use the Savage-Dickey density ratio (SDDR) to estimate m_0/m_i (Verdinelli and Wasserman (1995) and see

Koop, León-González and Strachan (2005) for an example of an application of this approach). To demonstrate briefly, the model $M_0 = M_{(0,1,5)}$ nests within all models at the point $a^* = 0$, and the SDDR can be computed as the ratio of the marginal posterior to the marginal prior at the point $a^* = 0$. Thus

$$\frac{m_0}{m_i} = \frac{p(a^* = 0|M_\xi, y)}{p(a^* = 0|M_\xi)}.$$

Given our earlier choices for the prior, the expression $p(a^* = 0|M_\xi) = \left(\frac{2\pi}{n}\right)^{-nr/2}$. Given sequences of draws $\left(\Omega^{(i)}, b^{*(i)}\right)$, $i = 1, \ldots, K$ from the posterior and $\left(\Omega^{(j)}, b^{*(j)}\right)$, $j = 1, \ldots, K$ from the prior[15], the marginal posterior density for a^* can be approximated by

$$\widehat{p}(a^* = 0|M_\xi, y) = K^{-1} \sum_{i=1}^{K} p\left(a^* = 0|\Omega^{(i)}, b^{*(i)}, M_\xi, y\right).$$

Alternatively we could directly estimate m_i using, for example, the approach of Gelfand and Dey (1994). This approach is attractive if the dimension of the integral to be approximated is not large. In our case we can reduce the dimension as the posteriors of a^* and Ω have standard conditional forms and so we can readily integrate these out of the full joint posterior to obtain an expression for $p(b^*, M_\xi|y) \propto \mathfrak{g}_\xi k(b^*|M_\xi, y)(db^*)$. We can write $m_\xi = \mathfrak{g}_\xi \mathfrak{c}_\xi$ where $\mathfrak{g}_\xi$ is known and given in the Appendix and $\mathfrak{c}_\xi = \int k(b^*|M_\xi, y)\, db^*$ is the only unknown term to be estimated.

To estimate the marginal likelihood, we must estimate the term $\mathfrak{c}_\xi$. We approximate this integral using the method proposed by Gelfand and Dey (1994) which uses the relation

$$\frac{1}{\mathfrak{c}_\xi} = \int \frac{q}{k} \frac{k}{\mathfrak{c}_\xi} (db^*).$$

in which $q = q(b^*)$ is a proper known density and $k = k(b^*|M_\xi, y)$. As we have we have a sequence of draws $b^{*(i)}$, $i = 1, ..., J$, from the posterior distribution for b^*, we can estimate $\mathfrak{c}_\xi$ by

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

[15] It is relatively straightforward to show that the conditional priors are all proper and of standard Normal and inverted Wishart forms.

As our choice for q, we use a truncated Normal with mean zero and covariance matrix $I_{nr}\frac{1}{n}$. The truncation is such that the density is zero for $b^{*\prime}b^{*} > \chi^{*}_{nr}$ where $\Pr\left(\chi_{nr} > \chi^{*}_{nr}\right) = 0.01$ and χ_{nr} is a Chi-squared random variable with nr degrees of freedom. Except for the truncation, this the same density as the prior and implies a Uniform density for β.[16] With a non-diagonal covariance matrix this density for β^{*} would imply a Matrix Angular Central Gaussian distribution (Chikuse, 1990) for β. Our choice of q implies the ratio k/q has the form of a kernel for a 1-1 poly-t density but over a compact support. This density will have fat tails and so q/k will tend to be stable, however the truncation further ensures the stability.

Alternative approaches exist for estimating $\mathfrak{c}_{\xi}$. For the computation of the posterior probabilities, we need only draws of β^{*} to approximate $\mathfrak{c}_{\xi}$. 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 Green (1995). Kleibergen and van Dijk (1998) and Kleibergen and Paap (2002) develop MCMC schemes in the simultaneous equations model and the VECM. Strachan (2003) employs this approach when β has been identified using restrictions related to those of the ML estimator of Johansen (1992). Alternatively one may use the Adaptive Radial based method of Bauwens, Bos, van Dijk and van Oest (2004) or the neural network mixture method of Hoogerheide, Kaashoek and van Dijk (2006). Bauwens and Lubrano (1996) and Strachan and Inder (2004) demonstrate other approaches.

3.3 Bayesian Model Averaging with MCMC.

In this section we outline how we implement Bayesian model averaging to provide unconditional inference. One of the advantages of our approach over previous approaches is that for all model specifications we consider, as shown in the Appendix, the posterior will be proper and all finite moments of $b^{*} = vec\left(\beta^{*}\right)$ (or β) exist. 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

[16] The symmetric truncation for the symmetric density has no implications for the distribution on the cointegrating space or β.

b^*.

Suppose we have an economic object of interest ϕ which is a function of the parameters for a given model $(a^*, \Omega, b^*|M_\xi)$, $\phi = \phi\,(a^*, \Omega, b^*|M_\xi)$. Examples include estimates of impulse responses, forecasts, or loss functions. We wish to report the unconditional (upon any particular model) expectation of this object. That is, we wish to report an estimate of

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

where $E\,(\phi|y, M_\xi)$ is the expectation of ϕ from model ξ. To obtain this estimate, denote the i^{th} draw of the parameters from the posterior distribution for model M_ξ as $\left(a^{*(i)}, \Omega^{(i)}, b^{*(i)}\right)$ and so the i^{th} draw of ϕ as $\phi^{(i)} = \phi\left(a^{*(i)}, \Omega^{(i)}, b^{*(i)}|M_\xi\right)$. 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 $E\,(\phi|y, M_\xi)$ from each model by

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

These estimates are then averaged as

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

in which $\widehat{p}\,(M_\xi|y)$ is an estimate of $p\,(M_\xi|y)$.

4 The Great Ratios.

In this subsection we provide empirical evidence on the role of permanent shocks in logarithms of 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 from quarter one 1947 up to and including the fourth quarter of 2007[17], we report evidence upon the

[17] The data are seasonally adjusted, quarterly observations covering the period from the first quarter 1947 to the last quarter of 2007, on Personal Consumption Expenditures, Gross Private Domestic Investment, and GDP (Source: Bureau of Economic Analysis).

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.

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 $i_t - inc_t$ will both be stationary $I(0)$ processes. We therefore allow the rank, r, to vary over all possible values, $r \in [0, 1, \ldots, n]$ 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 130 models[18] may be summarized as $r \in [0, 1, 2, 3]$, $o \in [0, 1]$, $d \in [1, 2, 3, 4, 5]$ and $l \in [0, 1, \ldots, 5]$.[19]

Beginning with the support for the alternative models in the model set, the modal model with posterior probability of 18%, has eight lags of differences ($l = 4$), two stochastic trends ($r = 1$), the great ratios do form the cointegrating relations ($o = 1$) and neither the equilibrium relations nor the levels contain deterministic trends ($d = 5$). The posterior probabilities of the models (averaged over lags and computed using the method of Gelfand and Dey (1994)) are given in Table 1. These results show that both with and without the overidentifying restrictions, the weight of support is upon there being two common stochastic trends in y_t ($p(r = 1|y) = 66\%$), with some support for only one stochastic trend ($p(r = 2|y) = 34\%$). This gives a log odds ratio of 0.66 which as evidence for two rather than one stochastic trend is, according to Kass and Raftery (1995), is not worth more than a bare mention. This result gives some support to the first feature suggested by the model

[18] Simply multiplying up the cardinality of each set of (r, o, d, l) would produce 240 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.

[19] All models with lags below 3 had zero posterior probability.

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 has slightly stronger support with a posterior probability of 45%, although again this is not strong evidence. These two conclusions do not disagree with the findings of Centoni and Cubadda (2003) (hereafter CC) who use a data set to April 2001. Finally, we find strong evidence that the equilibrium relations are $I(0)$ with no linear deterministic trends as $p(d = 4|y) = 55\%$.

Table 2 lists the probabilities and cumulative probabilities of the five most likely models. Although these models account for two thirds of all the posterior probability over this model set (and the top 20 models account for 99% of the posterior probability mass) there is not strong support for any one of these models. The results in Table 2 do suggest support for particular structural features: there are four lags of differences; no deterministic trends; there is cointegration; and the great ratios are as likely as not to be stable.

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

Just Identified Models ($o = 0$)					
r	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
0	0.000	0.000		0.000	
1	0.000	0.001	0.000	0.109	0.164
2	0.000	0.001	0.000	0.074	0.021
3	0.000		0.000		0.000
Over Identified Models ($o = 1$)					
1	0.000	0.001	0.000	0.087	0.294
2	0.000	0.000	0.000	0.246	0.000

Table 2: Posterior probabilities, $P(M_\xi|y)$, of the top five models.

d	l	r	o	$P(M_\xi\|y)$	Cumulative probabilities
5	4	1	1	0.1797	0.1797
4	4	1	0	0.1422	0.3219
4	4	2	1	0.1275	0.4494
5	4	1	0	0.1139	0.5633
4	4	2	0	0.1064	0.6697

Effects of 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.

KPSW estimate the proportion of variance 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 0.57 for c_t, 0.14 for i_t and 0.18 for inc_t. Table 3 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 income 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, while important, *the single permanent shock is not the main determinant of business cycle fluctuations.*

Table 3: Estimated variance decompositions into permanent components in the frequency domain.

Estimation method	c_t	i_t	inc_t
Averaged over all models	0.346	0.341	0.348
CC model $M_{(2,0,3)}$	0.344	0.345	0.346
KPSW model $M_{(2,1,3)}$	0.461	0.390	0.454
Best model $M_{(2,0,2)}$	0.341	0.339	0.354

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.

In each case there is a slightly positive long run response to a permanent shock in each series. More interestingly, we see that the form of the densities are generally symmetric but leptokurtic. The leptokurtosis occurs at both short and longer horizons and so is not a result of a few divergent paths which would show up only at longer horizons. The leptokurtosis results from mixing over normals and demonstrates well the effect of accounting for both model and parameter uncertainty. That is, the form of the density at each horizon reflects the effect that low probability events have on the tail behaviour. From the upper panels, we see that the bulk of the mass does not continue to increase as rapidly at longer horizons and, in fact, settles down to a consistent shape. This is most evident for consumption which seems to have stopped increasing its spread after around 30 periods.

The form of these densities are important for giving a full account of the uncertainty associated with the responses. In each case the fat tails in the densities derive from models with low posterior probabilities. Neglecting these models and using only the best model (effectively assuming model certainty) would produce very different estimates of the distributions of impulse responses, forecasts and the resulting expected loss from a particular action.

5 The Risk of a Liquidity Trap in the USA and Japan and Evidence of its Importance for Monetary Policy.

5.1 Introduction.

In recent decades some industrialized nations, in particular Japan, seem to have reached a state of the economy where inflation, interest rates and economic growth are all low. Some illustrative data (interest rates[20], r_t; prices[21], p_t; and real per capita GDP, g_t) are given in Figure 5 for the USA and Japan. With low or negative inflation and already low interest rates, monetary policy to activate the economy by lowering interest rates even further may not be effective or possible anymore. Earlier authors have characterized this state of the economy as the liquidity trap and - clearly - Central Bank authorities would wish to avoid such a state.

The literature on the liquidity trap is extensive. Discussions of this issue date from Keynes (1936) and Hicks (1937)[22] who focussed upon the form of the IS-LM model and a positive lower bound on long interest rates. More recent work, however, tends to focus upon the importance of the zero lower bound (ZLB) on short maturity interest rates. Eggertsson and Woodford (2003a,b and 2004) demonstrate how the ZLB on interest rates complicates the conduct of monetary policy in a low inflation environment and the role of fiscal policy in such a situation. Summers (1991) identified a trade-off, due to the ZLB, between the aims of achieving a zero-inflation target and stable output. Fuhrer and Madigan (1997) use simulation to provide evidence on the importance of a ZLB on US interest rates in a low inflation environment in contrast to a high inflation environment, and conclude that the optimal rate of inflation should be positive rather than zero. Other interesting studies include Reifschneider and Williams (2000), Orphanides and Wieland (1998), and Wolman (1998).

Our investigation has two stages. In the first, we generalize the VAR model to allow for both cointegration and multivariate stochastic volatility (VECM-SV). With this model we aim to provide estimates of the probability of encountering the LT for the US and Japan between 1975 and quarter three of 2006 using a mixture of forecast distributions generated by averaging over a range of models. In the second stage, we introduce a smooth transition function in a monetary equation (VECM-SV-ST) to allow inference on changes in the transmission mechanism as the forecast distribution of interest rates puts more mass near zero. We use this model to address

[20] The interest rate is the overnight Federal Funds Rate for the US and the Money Market (call money) rate for Japan.

[21] For each country, p_t is taken to be the log of the CPI for the US and Japan.

[22] Boianovsky (2003) outlines early discussion by Hicks (1937) who attributed the concept of the LT to Keynes and focussed on a lower bound on long (bond) rates. More recent discussions tend to focus upon the ZLB for short rates. Boianovsky (2003) gives an interesting overview of the development of the term and concept of the "liquidity trap".

two aims: 1) to again provide estimates of the probability of encountering the LT; and 2) to obtain inference on the whether there is a change in the monetary policy transmission mechanism when the risk of encountering the liquidity trap becomes significant.

We define the LT as occurring if interest rates fall below some low level, $\boldsymbol{r}$, for two or more consecutive quarters in the next twelve months. That is, if we denote the interest rate by r_t, then the event LT is defined as $LT = \{r_{t+l}, r_{t+l+1} : (r_{t+l} < \boldsymbol{r}) \cap (r_{t+l+1} < \boldsymbol{r}), l \in 1, 2, 3\}$.[23] The probability of this event at time t is $p_{LT,t} = \Pr(LT)$. The reasoning behind this definition of the LT is that we assume $\boldsymbol{r}$ is a boundary for what the central bank believes is an absorbing state under current strategies. Once interest rates fall below this level, the bank or the government must adopt different strategies, such as appropriate fiscal stimulus, to escape the LT.

In the first stage of the analysis we use the VECM-SV and assume the boundary $\boldsymbol{r}$ is known to be either 0.25% or 0.5%. In the second stage we generalize the model to the VECM-SV-ST and estimate $\boldsymbol{r}$ as an unknown parameter in the model. In that the computation of $p_{LT,t}$ relies on the entire forecast distribution, not just the mean forecast, our approach resembles the approach of Orphanides and Wieland (1998). We impose the ZLB in our model by working with the log of the interest rates and so this approach implies a nonlinear reaction function similar to that used in, for example, Fuhrer and Madigan (1997). We compute the probability, $p_{t,i}$, at each time t that the interest rate i periods in the future will be below $\boldsymbol{r}$ for each $i = 1, 2, 3, 4$. We can then compute $p_{LT,t}$ from these values of $p_{t,i}$.

In the first stage of the work the VECM-SV does not permit any change in the monetary policy transmission mechanism and we compute $p_{LT,t}$ with known $\boldsymbol{r}$. The evidence suggests the $p_{LT,t}$ significantly increased in the US and particularly in Japan around the turn of the century but has fallen more recently.

[23] It might seem more accurate to simultaneously account for the projected paths of output and inflation. However, the liquidity trap is clearly defined within the IS-LM model as the when the rate of interest hits its floor on the left part of the LM curve. Further, concern for the liquidity trap derives from the inability of the rate of interest to act as a stabilizer for the economic system when it hits its lower bound (Boianovsky, 2003). Interest rates can always be increased. However, if the interest rate is at its lower bound – regardless of where is the distribution of income and price growth – they can be lowered no further and this imposes a constraint on policy options. Finally, the location of the distribution of income and price growth is very closely related to the location of the interest rate and early work found the differences in the approaches were not significant.

While $p_{LT,t}$ may be low, we might expect the cost of LT, l_{LT}, is high, thus the expected cost $(p_{LT,t}l_{PT})$ would be large enough to prompt banks to react. An example of a strategy that might produce such a response is the forward-looking adjustment to the Taylor Rule discussed in Reifschneider and Williams (2000). This change in behaviour, which we model in the second stage of the study, will show up in the equation for r_t. In the second stage, the results from the VECM-SV-ST show that allowing the bank to react to the possibility of the LT results in significantly lower probability estimates over the full period for each country and estimated values of $\boldsymbol{r}$ of 0.28% for the US and 0.38% for Japan.

5.2 Cointegrating VAR with stochastic volatility and smooth transition.

As evidenced from the literature (Cogley and Sargent 2005, Primiceri 2005 and Sims and Zha 2006), it is important to appropriately model heteroscedasticity for these variables. Therefore values of $p_{t,f}$ are estimated from a reduced form vector error correction model as in (1), but with multivariate stochastic volatility given by

$$\Delta y_t = z_{1,t}\beta\alpha + z_{2,t}\Phi + u_t\Sigma_t A_t^{-1} \tag{10}$$

where u_t is a vector of independent standard Normal variables. The specification for the covariance matrix is similar to that of Primiceri (2005). We denote the covariance of $\varepsilon_t = u_t\Sigma_t A_t^{-1}$ at time t as Ω_t and decompose Ω_t as $A_t'\Omega_t A_t = \Sigma_t'\Sigma_t$ where $\Sigma_t = diag\left\{\sigma_{1,t}, \sigma_{2,t}, \sigma_{3,t}\right\}$ and A_t is given as

$$\Delta y_t A_t = (\Delta g_t, \pi_t, \Delta lr_t)\begin{bmatrix} 1 & \varrho_{1,t} & \varrho_{2,t} \\ 0 & 1 & \varrho_{3,t} \\ 0 & 0 & 1 \end{bmatrix} = \begin{pmatrix} \Delta g_t \\ \pi_t + \varrho_{1,t}\Delta g_t \\ \Delta lr_t + \varrho_{2,t}\Delta g_t + \varrho_{3,t}\pi_t \end{pmatrix}.$$

We do not give economic interpretations to these expressions beyond assuming that the central bank reaction function is embedded in the mean equation for the interest rates (see for example, Garratt *et al.* 2003).

Define the log structural variances $h_{i,t} = \ln(\sigma_{i,t})$ for $i = 1, 2, 3$ and collect the parameters into 3×1 vectors $h_t = (h_{1,t}, h_{2,t}, h_{3,t})'$ and $\varrho_t = \left(\varrho_{1,t}, \varrho_{2,t}, \varrho_{3,t}\right)'$. We assume these parameter vectors evolve according to the state equations

$$h_t = h_{t-1} + \nu_t \text{ and } \varrho_t = \varrho_{t-1} + \chi_t. \tag{11}$$

The vectors ν_t and χ_t are assumed independent of each other and Normally distributed with zero mean. ν_t has 3×3 covariance matrix Ψ and the first element of χ_t has variance s_1 and is independent of the remaining elements which have 2×2 covariance matrix S_2.

Recent evidence suggests it is less important to allow time variation in the reduced form coefficients if heteroscedasticity is appropriately modelled. While providing evidence themselves, Sims and Zha (2006) point to the results for the US by Primiceri (2005) and the contrast between the results of Cogley and Sargent (2001) and Cogley and Sargent (2005) in support of this claim for the US. We therefore assume constant reduced form mean equation coefficients.

In the second stage of the study we again use a VAR model, but we augment the equation for the log interest rates with parameters that will be different from zero if the central does react to the possibility of the LT, and the same variables multiplied by the probabilities of the liquidity trap.

This augmentation in the reduced form equation comes from an augmented reaction function. The modeling strategy assumes the central bank obtains forecasts of interest rate distributions while ignoring the possibility of the LT. If the forecast distributions imply a high enough value for $p_{LT,t}$ such that the expected costs of the LT is significant, the bank will then incorporate this into its interest rate setting strategy via a shift in the reaction function. For example, assume that ignoring the possibility of the LT, the interest rate rule produces $\ln r_{NoLT,t}$. Estimates of $p_{LT,t}$ are then obtained from the forecast densities of $r_{NoLT,t}$.

Next, assuming $p_{LT,t} = 1$, the interest rate rule produces $\ln r_{LT,t}$. Combining the two rules we therefore have the rule for setting the interest rate, r_t, as

$$\begin{aligned} \ln r_t &= (1 - p_{LT,t}) \ln r_{NoLT,t} + p_{LT,t} \ln r_{LT,t} \\ &= \ln r_{NoLT,t} + p_{LT,t} (\ln r_{LT,t} - \ln r_{NoLT,t}) . \end{aligned}$$

As $p_{LT,t}$ is a continuous bounded variable, this specification implies a smooth transition function for the reaction function where the transition function is $p_{LT,t}$ which itself is a function of forecast densities of $r_{NoLT,t}$. The resulting mean equation for the log interest rates in the reduced form VAR will be functions of the reaction functions that produce $\ln r_{LT,t}$ and $\ln r_{NoLT,t}$. We augment the model (10) with this specification of the monetary policy equation. Denoting by a subscript 3 the coefficients in the equation for $\ln r_t$, the resulting equation in the VECM-SV-ST for the interest rates will be

$$\ln r_t = z_{1,t}\beta\alpha_3 + z_{2,t}\Phi_3 + p_{LT,t}\left[z_{1,t}\beta\alpha_3^{\#} + z_{2,t}\Phi_3^{\#}\right] + u_t\Sigma_t A_t^{-1}.$$

From the form above, we have simple testable hypotheses to establish the evidence for or against the hypothesis that monetary policy responds to the likelihood and extent of the event LT. If the monetary authority does not behave differently when faced with the liquidity trap then $E_t(\ln r_{LT,t}) = E_t(\ln r_{NoLT,t})$, which implies $\alpha_3^{\#} = 0$ and $\Phi_3^{\#} = 0$.

An important determinant of the role of expectations formation is whether shocks are permanent or transitory. We wish to remain uninformative on the exact specification of the model and so, to allow for the proportion of variability in the variables that is due to permanent shocks to cover the full range from zero (implied by no stochastic trends) to one (implied by three stochastic trends in the system), we consider models with $r = 0, 1, 2,$ and 3. Deterministic processes and the lag structure affect the forecasting performance of the model which is important in this application. We therefore allow $d = 3, 4,$ and 5 and $l = 0, 1, 2,$ and 3.

The full (general) model now has the form

$$\begin{aligned}\Delta y_t &= z_{1,t}\beta\alpha + z_{2,t}\Phi + p_{LT,t}\left[z_{1,t}\beta\alpha^{\#} + z_{2,t}\Phi^{\#}\right] + u_t\Sigma_t A_t^{-1} \\ &= \widetilde{z}_t\widetilde{A} + u_t\Sigma_t A_t^{-1} \qquad (12)\end{aligned}$$

where $\alpha^z = (0, 0, \alpha_3^{z\prime})'$ and $\Phi^z = (0, 0, \Phi_3^{z\prime})'$, $\widetilde{z}_t = (z_{1,t}\beta, z_{2,t}, p_{LT,t}z_{1,t}\beta, p_{LT,t}z_{2,t})$, $\widetilde{A} = [\alpha', \Phi', \alpha^{z\prime}, \Phi^{z\prime}]'$. This augmentation of the monetary policy equation implies the model has the form of a Seemingly Unrelated Regression model, we will refer to it as the SUR model. We collect all the mean coefficients into $\widetilde{a} = (a_0', a_1', a^{z\prime},)'$ where $a_1 = \left(vec(\alpha)', vec(\Phi)'\right)'$, $a^z = \left(vec(\alpha_3^z)', vec(\Phi_3^z)'\right)'$, and a_0 is the vector of zero elements in α^z and Φ^z. Collecting the nonzero elements into $a = (a_1', a^z)'$, so that $\widetilde{a} = (a_0', a')'$.

5.3 Priors and posteriors.

We describe the sampling scheme for $\boldsymbol{r}$, ϱ_t, h_t, and a. Given $\boldsymbol{r}$, ϱ_t, h_t, and a, the probabilities in the vector $p_{LT} = (p_{LT,1}, \ldots, p_{LT,T})'$ can then be computed directly and used to update the posteriors for $\boldsymbol{r}$, ϱ_t, h_t, and a.

We use the same priors for the variances of the state equations, Ψ, s_1 and S_2, as Primiceri (2005), and our priors for the initial values ϱ_0 and h_0

are Normal with the mean equal to the OLS estimate from the first 20% of the sample and covariances $0.004I_3$ and I_3. A good description of the method of drawing ϱ_t and h_t using the Kalman filter is given in Primiceri (2005) and so we refer readers to that paper for a full explanation. Briefly, for ϱ_t, measurement equations are constructed for $\varrho_{1,t}$ and $(\varrho_{2,t}, \varrho_{3,t})$ from $\left(\Delta y_t - \widetilde{z}_t\widetilde{A}\right) A_t = u_t\Sigma_t$ and for h_t the measurement equation is obtained by squaring and taking logs of the elements of the above equation. That is, if the i^{th} element of $\Delta y_t - \widetilde{z}_t\widetilde{A}$ is $\widehat{y}_{i,t} = \exp\{h_{i,t}\} u_{i,t}$, then the measurement equation for $h_{i,t}$ is given by[24] $\ln\left(\widehat{y}^2_{i,t}\right) = 2h_{i,t} + \ln\left(u^2_{i,t}\right)$. As the error $\ln\left(u^2_{i,t}\right)$ is not Normal, the mixture of Normals approximation of Kim *et al.* (1998) is used to implement the Kalman Filter.

For the vector a we again use a conditional Normal prior with zero mean but with covariance $\underline{V} = I_n \otimes \eta^{-1} I_{(r+k_i)}$. This prior is consistent with the one in Section 3 in that it imposes shrinkage towards random walks and conditions upon the covariance of the error, in this case $E\left(u'_t u_t\right) = I_n$. In deriving the posterior for a, we begin by postmultiplying (12) by $A_t\Sigma_t^{-1}$ and vectorising to obtain $\widetilde{z}_{0,t} = \widetilde{z}_t\widetilde{a} + \widetilde{e}_t$ where $\widetilde{e}_t = u'_t$, $\widetilde{a} = vec\left(\widetilde{A}\right)$, $\widetilde{z}_{0,t} = \Sigma_t^{-1} A'_t \Delta y'_t$ and $\widetilde{z}_t = \left(\Sigma_t^{-1} A'_t \otimes \widetilde{z}_t\right)$. Stacking the vectors $\widetilde{z}_{0,t}$, $\widetilde{z}_t$ and $\widetilde{e}_t$ as $\widetilde{z}_0 = \left(\widetilde{z}'_{0,1}, \ldots, \widetilde{z}'_{0,T}\right)'$, $\widetilde{z} = \left(\widetilde{z}'_1, \ldots, \widetilde{z}'_T\right)'$ and $\widetilde{e} = \left(\widetilde{e}'_1, \ldots, \widetilde{e}'_T\right)'$ we obtain a form $\widetilde{z}_0 = \widetilde{z}\widetilde{a} + \widetilde{e}$ which is similar to (4). Combining this form with the Normal prior for $\widetilde{a}$ given above, we obtain the conditional Normal posterior with mean $\overline{a} = \overline{V}\widetilde{z}'\widetilde{z}_0$ and covariance matrix $\overline{V} = \left[\eta I_{n(r+k_i)} + \widetilde{z}'\widetilde{z}\right]^{-1}$. Recall some elements of $\widetilde{a}$ are known to be zero and are collected into the vector a_0. We draw the remaining non-zero elements of $\widetilde{a}$ conditional upon $a_0 (= 0)$ using well known results for the conditional Normal distribution. That is, with the partition $\widetilde{a} = (a'_0, a')'$ and the conformable partitions of $\overline{a} = (\overline{a}'_0, \overline{a}'_.)'$ and

$$\overline{V} = \begin{bmatrix} \overline{V}_{00} & \overline{V}_{0.} \\ \overline{V}_{.0} & \overline{V}_{..} \end{bmatrix},$$

then the posterior for a (conditional upon $a_0, \boldsymbol{r}, \varrho_t, h_t$) is Normal with mean $\overline{a}_. - \overline{V}_{.0}\overline{V}_{00}\overline{a}_0$ and covariance matrix $\overline{V}_{..} - \overline{V}_{.0}\overline{V}_{00}^{-1}\overline{V}_{0.}$.

For the threshold parameter $\boldsymbol{r}$ we specify a Uniform distribution over the range from $(0\%, 1\%]$.[25] Using a random walk Metropolis Hastings (MH) al-

[24] In fact we use the offset adjustment such that the dependent variable is $\ln\left(\widehat{y}^2_{i,t} + 0.001\right)$ in place of $\ln\left(\widehat{y}^2_{i,t}\right)$.

[25] Proper priors are required for $\boldsymbol{r}$ and b as there is a point of local nonidentification at

gorithm to obtain draws of $\ln(\boldsymbol{r})$ results in low acceptance rates. To improve the acceptance rates we use a Metropolis Hastings scheme in which draws are obtained from a candidate density that approximates the posterior. This candidate is a Rao-Blackwellized estimate from a preliminary run using the random walk MH scheme.

To obtain model probabilities and the probability of the restriction $a^z = 0$, we use the Savage-Dickey density ratio to compute the Bayes factors. Details and examples of these techniques in a VECM are provided in Koop, Leon-Gonzales and Strachan (2005).

5.4 The Results.

The estimated probabilities for the range of stochastic and deterministic processes (r and d) presented in Table 4 suggest there exists model uncertainty. The modal model probabilities are $\Pr(d = 5, l = 3, r = 3|y) = 0.18$ for the US and $\Pr(d = 3, l = 1, r = 3|y) = 0.21$ for Japan. The estimated probability of the restriction $a^z = 0$ is almost one in each case suggesting that either central banks do not concern themselves with the risks associated with LT, or the evidence is weak due to there being too few observations where $p_{LT,t}$ would be large enough to be informative about a^z.
Conditioning upon the model with $a^z = 0$, the estimated probabilities of $p_{LT,t}$ are plotted in Figure 5 for $\boldsymbol{r} = 0.25\%$, $\boldsymbol{r} = 0.5\%$ with the interest rate r_t. This figure also plots the estimated $p_{LT,t}$ from the model without the restriction $a^z = 0$ and the value of $\boldsymbol{r}$ estimated from the model in (12). The situation in the US and Japan is interesting. The results for Japan indicate, not surprisingly, that the country met our definition of the LT from the beginning of 1999 and there appears to be a chance of escaping this situation emerging in 2006Q3. The probability of the LT in the US increases significantly after 2001Q1 to be between 10% and 20%, and begins to fall again after the middle of 2005.
Table 4: Posterior probabilities of structural features for real business cycle model. Note that the cells for observationally equivalent models have been merged.

the points $\boldsymbol{r} = 0$ and $b = 0$.

	The US			Japan		
r	$d=3$	$d=4$	$d=5$	$d=3$	$d=4$	$d=5$
0	0.08	0.10		0.04	0.06	
1	0.03	0.13	0.13	0.04	0.26	0.04
2	0.04	0.03	0.08	0.04	0.12	0.06
3	0.15		0.24	0.29		0.05

Table 5 reports the average estimates of $p_{LT,t}$ over various periods and with different assumptions about $\boldsymbol{r}$. The first column identifies the period over which values are averaged. The second column gives the average interest rate for that period. The third and forth columns give the estimates of $p_{LT,t}$ for $\boldsymbol{r} = 0.25\%$ and $\boldsymbol{r} = 0.50\%$ respectively. The final column gives estimated $p_{LT,t}$ when the model allows the central bank to react differently to the potential LT and $\boldsymbol{r}$ is estimated. The actual estimates in each case and for all t, are shown in Figure (6). The results in Table 5 and Figure (6) clearly indicate that for the US and Japan, the probability of LT has increased since 1994. This change has coincided with a fall in the level of interest and inflation rates (see Figure (5)). However, when we permit the central bank to react to the risk of the LT, we see that the probabilities are noticeably lower and slightly pre-empt the rise in the risk of the LT, $p_{LT,t}$. The fall in $p_{LT,t}$ when we allow banks to react to the risk of LT suggests that the banks did alter their behaviour, however slightly, to mitigate the risk of a LT.

Table 5: Average estimated $p_{LT,t}$ for $\boldsymbol{r} = 0.25\%$, $\boldsymbol{r} = 0.5\%$ and $\boldsymbol{r}$ estimated. Values are averaged over the periods in the first column and the second column gives the average interest rate, r_t, for each period.

	r_t	$p_{LT,t}\|\boldsymbol{r}=0.25\%$	$p_{LT,t}\|\boldsymbol{r}=0.5\%$	$p_{LT,t}\|\boldsymbol{r}$
	The US (estimated $\boldsymbol{r}=0.35\%$)			
1975-1994	7.99%	0.84%	1.17%	0.86%
1995-2001	5.29%	0.93%	1.37%	1.46%
2002-2004	1.38%	3.98%	7.33%	7.75%
2005-2006Q3	3.92%	2.89%	4.20%	2.58%
	Japan (estimated $\boldsymbol{r}=0.24\%$)			
1975-1994	6.01%	0.95%	1.35%	1.14%
1995-2001	0.39%	46.90%	82.33%	60.42%
2002-2004	0.00%	97.80%	98.17%	92.32%
2005-2006Q3	0.03%	96.40%	97.78%	92.89%

In our models we have allowed for heteroscedasticity via a multivariate stochastic volatility specification. As this is a significant departure from the standard VAR/VEC models usually considered one might question the importance of this extension for our results. We therefore estimated $p_{LT,t}$ again with the restriction that the errors are homoscedastic. The results are presented in Figure (7) for Japan for $\boldsymbol{r} = 0.25\%$ (solid line) and $\boldsymbol{r} = 0.5\%$ (dashed line).

The first instance where the Japanese interest rate falls below 0.5% for two consecutive periods (i.e., we first observe LT for $\boldsymbol{r} = 0.5\%$) is 1995Q4/Q5 and first fell below 0.25% for two periods in 1998Q4/Q5. We take $p_{LT,t} > p = 50\%$ as an indication that LT will occur over the next year. The homoscedastic models do not indicate LT will occur for $\boldsymbol{r} = 0.5\%$ until 1998Q4 and for $\boldsymbol{r} = 0.25\%$ until 2000Q1. The heteroscedastic models first estimate a probability of LT greater than 50% 1995Q3 for $\boldsymbol{r} = 0.5\%$ and 1998Q4 for $\boldsymbol{r} = 0.25\%$. That is, the homoscedastic models do not indicate LT will occur until at least a year after it has occurred, while the heteroscedastic models clearly indicate that LT will occur correctly or one quarter early. The results change as we change p, but the poor relative performance of the homoscedastic models remains. These results suggest that modelling the volatility is important for estimating events that occur in the tails of the distributions, such as LT.

Due to the rare nature of the event LT, any evidence that it matters for central banks will be very weak. Recent work on IS-LM models based upon optimizing behaviour such as Krugman (1998) and McCallum (2000), focuses on the lower bound on short rates and suggests this bound could be zero. Our work does not aim to provide direct evidence for or against this result. Rather we provide evidence (albeit weak) that central banks in the US and Japan react differently when setting rates and faced with an increase in the possibility that rates will go 'too low'. The formal evidence suggests the central banks do not respond to the increased risk of the LT (as the probability $a^z = 0$ is one in all cases). However, allowing banks to respond (by letting $a^z \neq 0$) noticeably affected the risk for the US and Japan. Allowing when $a^z \neq 0$ the probability of a LT for Japan reduces, although not significantly and not always.

6 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 structure. 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 defined by structural features such as cointegration, deterministic processes, short-run dynamics and overidentifying restrictions upon the cointegrating space. We apply the techniques 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 and to the evidence and relevance of the liquidity trap for central bank behaviour for the US and Japan.

The method presented in this paper has already found applications in several other 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 uncertainty. 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 Green (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 as Primiceri (2005) does for the VAR. 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 (see Harvey, Trimbur and van Dijk (2007)). 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.

Acknowledgements

A preliminary version of this paper has been presented at, among others, the ESEM-2007, ECB 2007, MCMSki IMS-ISBA meeting in Bormio 2008, and the US Federal Reserve Bank in 2008. We would like to thank Luc Bauwens, Geert Dhaene, David Hendry, Lennart Hoogerheide, Soren Johansen, Helmut Lutkepohl, Christopher Sims, Mattias Villani and Anders Warne for helpful discussions on the topic of this paper. Of course, any remaining errors remain the responsibility of the authors. Strachan acknowledges the assistance from The University of Queensland and van Dijk acknowledges financial support from the Netherlands Organization of Scientific Research.

7 References.

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.

Bauwens, L., C.S. Bos, H.K. van Dijk & R.D. van Oest, (2004): "Adaptive radial-based direction sampling: some flexible and robust Monte Carlo integration methods," *Journal of Econometrics* **123(2)**, 201-225.

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., *Eco-*

nomic 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.

Bofinger, P. (2001): *Monetary Policy: Goals, Institutions, Strategies, and Instruments.* Oxford University Press.

Boianovsky, M. (2003): "The IS-LM Model and the Liquidity Trap Concept: From Hicks to Krugman", *Brazilian Association of Graduate Programs in Economics,* Proceedings of the 31th Brazilian Economics Meeting

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

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.

Cogley, T., and T. J. Sargent (2001): "Evolving Post-World War II US Inflation Dynamics," *NBER Macroeconomics Annual* **16**, 331–373.

Cogley, T., and T. J. Sargent (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics* **8**, 262-302.

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.

Eggertsson, G. and M. Woodford (2003a): "Optimal Monetary Policy in a Liquidity Trap," *International Workshop on Overcoming Deflation and Revitalizing the Japanese Economy*, Economic and Social Research Institute, Tokyo, September 18, 2003.

Eggertsson, G. and M. Woodford (2003b): "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, 2003:1, pp. 139-211.

Eggertsson, G. and M. Woodford (2004): "Optimal Monetary and Fiscal Policy in a Liquidity Trap," *NBER International Seminar on Macroeconomics*, Reykjavik, Iceland, June 18-19, 2004

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.

Fuhrer, J.C. and B.F. Madigan (1997): "Monetary Policy when Interest Rates are Bounded at Zero," *The Review of Economics and Statistics*, Vol. 79, No. 4: 573-585.

Garratt, A, K. Lee, M. H. Pesaran, and Y. Shin (2003): "A Long Run Structural Macroeconomic Model of the UK," *Economic Journal* **113**, 412-455.

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.

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

Harvey, A.C. T.M. Trimbur and H. K. van Dijk (2007): “Trends and cycles in economic time series: A Bayesian approach,” *Journal of Econometrics* **140**, 618-649.

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

Hicks, J. (1937): “Mr. Keynes and the ‘Classics’: A Suggested Interpretation,” *Econometrica* **5**, 147-59.

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

Hoogerheide, Johan F. Kaashoek and Herman K. van Dijk (2006) “On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks”, *Journal of Econometrics* In Press.

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.

Keynes, J.M. (1936): *The General Theory of Employment, Interest and Money*, Volume VII of The Collected Writings of John Maynard Keynes (1973), Macmillan, London.

Kim, S., Shephard, N. and Chib, S. (1998), “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models”, *Review of Economic Studies*, **65**, 361–393.

King, M. (1999): "Challenges for Monetary Policy: New and Old," *in New Challenges for Monetary Policy: A Symposium Sponsored by the Federal Reserve Bank of Kansas City.* FRB Kansas City: 11-57.

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. Léon-Gonzalez and R. W. Strachan (2008a): "Efficient posterior simulation for cointegrated models with priors on the cointegration space" forthcoming in *Econometric Reviews.*

Koop G., R. Léon-Gonzalez and R. W. Strachan (2008b): "Bayesian inference in a cointegrating panel data model" forthcoming in *Advances in Econometrics, Volume* 23.

Koop G., R., S. Potter and R. Strachan (2005): "Re-examining the Consumption-Wealth Relationship: The Role of Model Uncertainty," *Journal of Money, Credit and Banking*, Vol. 40, No. 2–3, 341-367.

Krugman, P. (1998): "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity,* no. 2: 137-205.

Krugman, P. (2000): "Thinking About the Liquidity Trap," *Journal of the Japanese and International Economics* **14** (Dec): 221-37.

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.

McCallum, B.T. (1989): *Monetary Economics.* New York: Macmillan.

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.

Orphanides, A. and V. Wieland (1998): "Price Stability and Monetary Policy Effectiveness when Nominal Interest Rates are Bounded at Zero," *Federal Reserve Board Finance and Economics Discussion Series No 35*, August.

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.

Primiceri, G. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* **72**, 821-852.

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

Reifschneider, D. and J.C. Williams (2000): "Three Lessons for Monetary Policy in a Low Inflation Era," *Journal of Money, Credit and Banking* **32**, 936-966.

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.

Sims, C. A. and T. Zha (2006) "Where there Regime Switches in US Monetary Policy?," *American Economic Review*, forthcoming.

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 (2006) "Model uncertainty and Bayesian Model averaging in vector autoregressive processes," Econometric Institute Report EI 2006-08, 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.

Summers, L. (1991): "How Should Long-Term Monetary Policy be Determined?" *Journal of Money Credit and Banking* **23**, 625-631.

Svensson, Lars E.O. (2000): "How Should Monetary Policy be Conducted in an Era of Price Stability?," NBER Working Papers 7516, National Bureau of Economic Research, Inc.

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.

Verdinelli, I. and L. Wasserman (1995): "Computing Bayes Factors using a generalization of the Savage-Dickey density ratio," *Journal of the American Statistical Association* **90**, 614-618.

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

Wolman, A.L. (1998): "Real Implications of the Zero Bound on Nominal Interest Rates," *Federal Reserve Bank of Richmond*, mimeo.

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

8 Appendix

In this appendix we provide the theorems used in the paper. For background and proofs we refer to the unpublished report Strachan and van Dijk (2006). We also sketch the line of reasoning leading to some theorems.

To integrate (9) with respect to (Ω, a^*, b^*) we first analytically integrate (5) with respect to (a^*, Ω) as these parameters have conditional posteriors of standard form. This integration gives us the following.

Theorem 1 *The marginal posterior for* (b^*, M_ξ) *is*

$$\begin{aligned} p(b^*, M_\xi|y) &\propto \mathfrak{g}_\xi k(b^*|M_\xi, y)(db^*), \qquad (13)\\ k(b^*|M_\xi, y) &= |I_r\eta + \beta^{*\prime} D_0 \beta^*|^{-T/2} |I_r\eta + \beta^{*\prime} D_1 \beta^*|^{(T-n)/2} \exp\left\{-\frac{n}{2} tr \beta^{*\prime}\beta^*\right\}. \end{aligned}$$

The expressions for $\mathfrak{g}_\xi$, D_0 *and* D_1 *are*

$$\begin{aligned}
\mathfrak{g}_\xi &= (2\pi)^{-r/2}\,\eta^{n\left(k_\xi+r\right)/2} n^{-nr/2}\left|\underline{S}+S_{00}\right|^{-T/2}\left|I\eta+Z_2'Z_2\right|^{-n/2} \\
D_1 &= Z_1'M_{22}Z_1,\;\; M_{22}=I_T-Z_2\left(Z_2'Z_2+\eta I_{k_i}\right)^{-1}Z_2' \text{ and} \\
D_0 &= D_1-S_{01}S_{11}^{-1}S_{10}
\end{aligned}$$

where

$$S_{10}=Z_1'M_{22}Z_0 \text{ and } S_{00}=Z_0'M_{22}Z_0.$$

Proof. See, for example, Zellner (1971) or Bauwens and van Dijk (1990)[26].■

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*

$$\underline{dg_r^n}=\pi^{-(n-r)r}\Pi_{j=1}^{r}\frac{\Gamma\left[(n+1-j)/2\right]}{\Gamma\left[(r+1-j)/2\right]}\left|I_r+\overline{\beta}_2'\overline{\beta}_2\right|^{-n/2}\left(d\overline{\beta}_2\right) \tag{14}$$

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

Next we provide a theorem that linear identifying restrictions with a flat prior give zero weight to the chosen linear restrictions. The Jacobian defined by (14) 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*

$$\begin{aligned}
\left(d\overline{\beta}_2\right) &= \pi^{(n-r)r}\Pi_{j=1}^{r}\frac{\Gamma\left[(r+1-j)/2\right]}{\Gamma\left[(n+1-j)/2\right]}\left|I_r+(c\beta)'^{-1}\beta'c_\perp'c_\perp\beta\,(c\beta)^{-1}\right|^{n/2}\left(dg_r^n\right) \\
&= J\;dg_r^n. \qquad (15)
\end{aligned}$$

Proof. Invert (14) 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.

[26] Remark: From the expression (13) 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.

Theorem 4 *Given r, use of the normalisation $\overline{\beta}_2 = c_\perp \beta \left(c\beta\right)^{-1}$ results in a transformation of measures for the transformation $\overline{\beta}_2 \in R^{(n-r)r} \longrightarrow \mathfrak{p} \in G_{r,n-r}$ 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 $\overline{\beta}_2 \longrightarrow \mathfrak{p}$ 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 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. From the expression for $k\left(b^*|M_\xi, y\right)$ above, we can see the marginal posterior for b^* is a polynomial times the kernel for a Normal. The expectation with respect to the divergent Lebesgue measure of the polynomial is finite as it is the kernel of a 1-1 poly-t (Drèze, 1977). The measure with respect to a convergent measure will then be finite. As all moments of a Normal exist, the expectation of this polynomial with respect to kernel of the Normal - a convergent measure - will be finite. Taking the density as the expectation of a polynomial with respect to a Normal distribution also tells us that $|b^*|^c \, k\left(b^*|M_\xi, y\right)$ for any $c \geq 0$ will be finite and so all moments will exist. ■

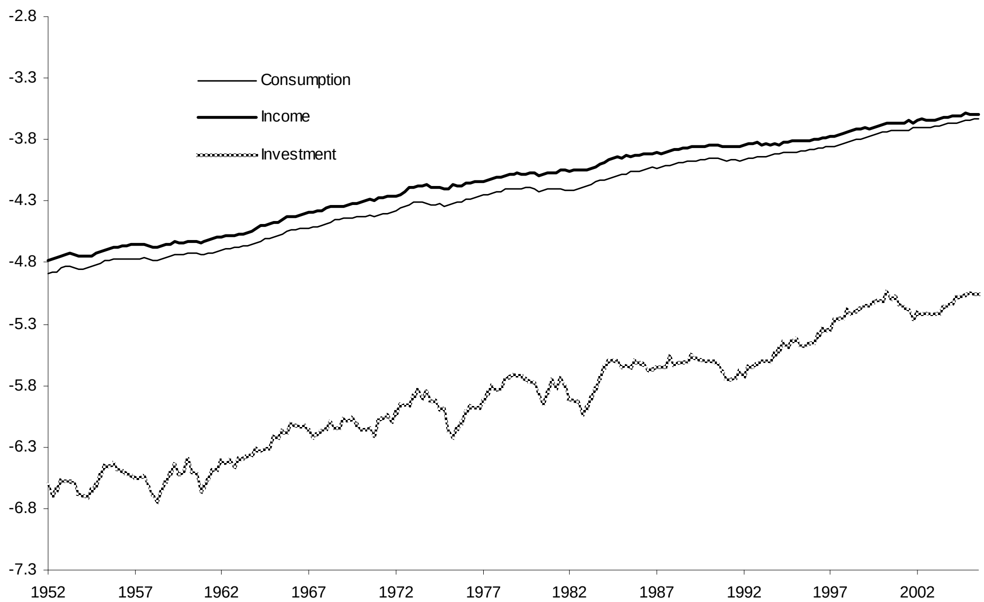

Figure 1: Logarithms of U.S. consumption (c_t), investment (i_t) and income (inc_t). The data are seasonally adjusted, quarterly observations covering the period from the first quarter 1951 to the second quarter of 2005, on Personal Consumption Expenditures, Gross Private Domestic Investment, and GDP (Source: Bureau of Economic Analysis).

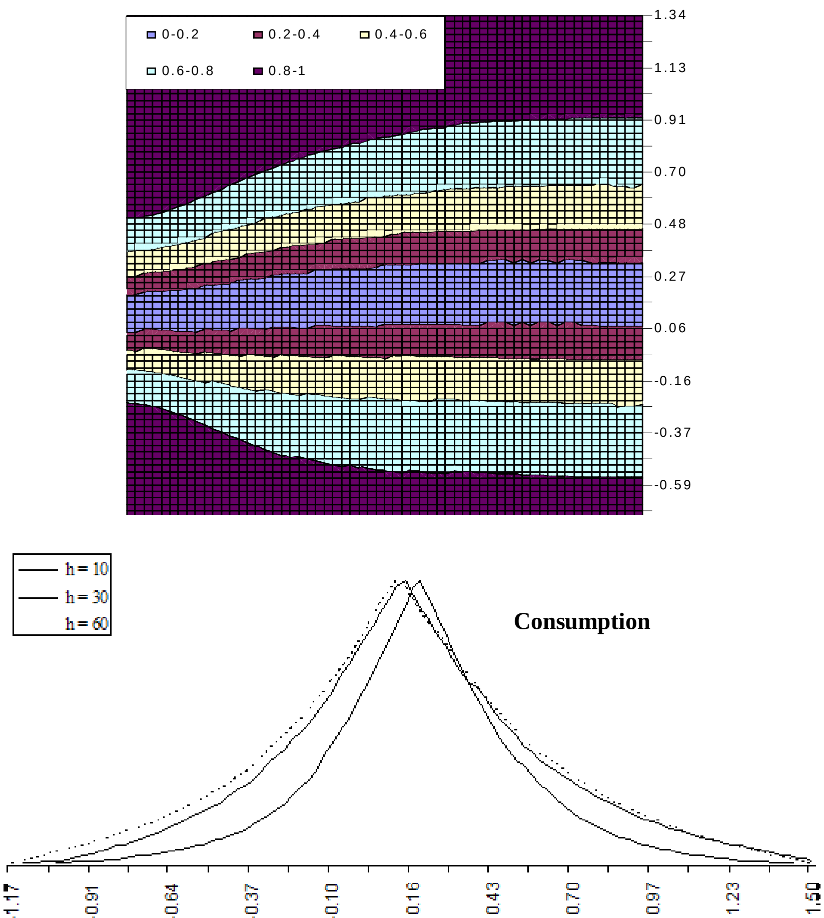

Figure 2: This figure shows the densities over 60 periods of the impulse responses of **consumption** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse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

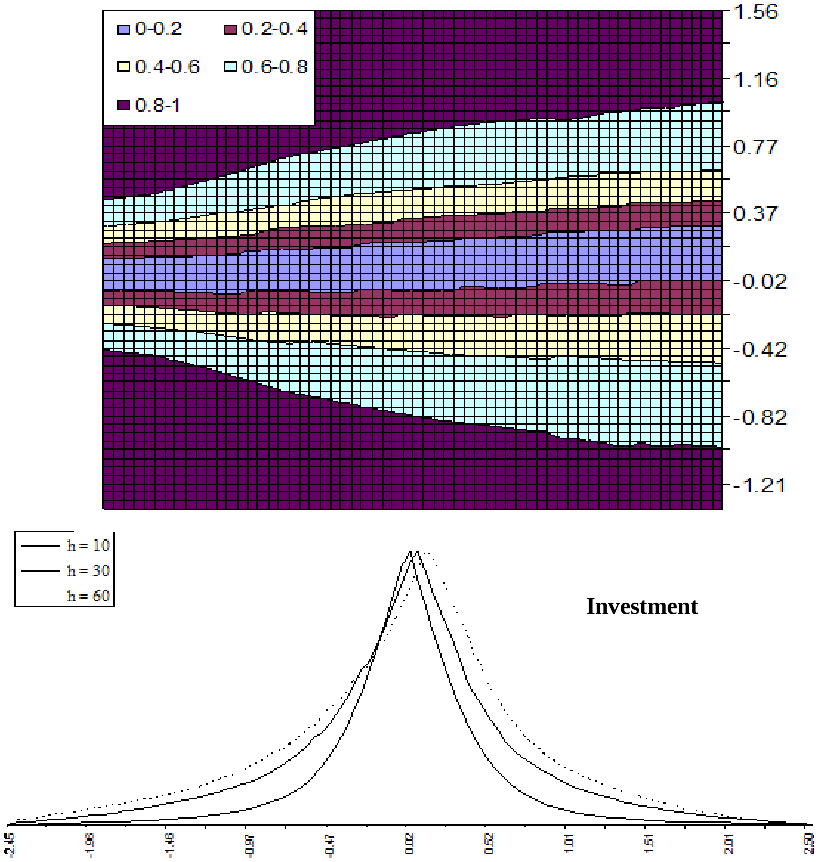

Figure 3: This figure shows the densities over 60 periods of the impulse responses of **investment** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse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

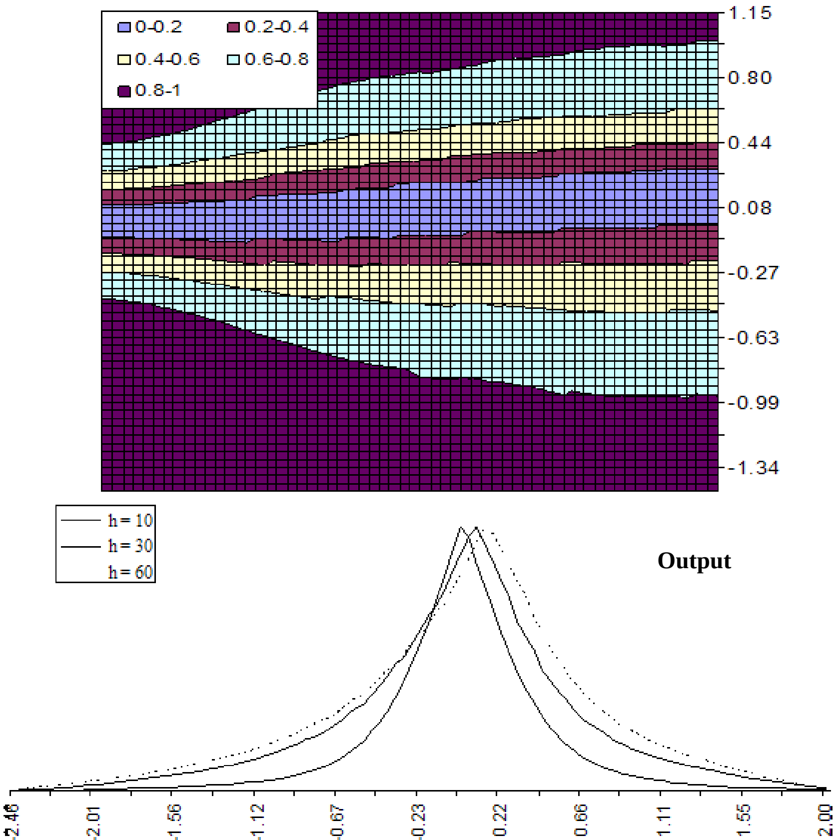

Figure 4: This figure shows the densities over 60 periods of the impulse responses of **income** to a permanent shock. The upper panel shows the 20% (0-0.2), 40% (0.2-0.4), 60% (0.4-0.6) and 80% (0.6-0.8) highest posterior density intervals. The lower panel shows the density profiles for the impulse response at $h = 10$, 30 and 60 periods into the future.

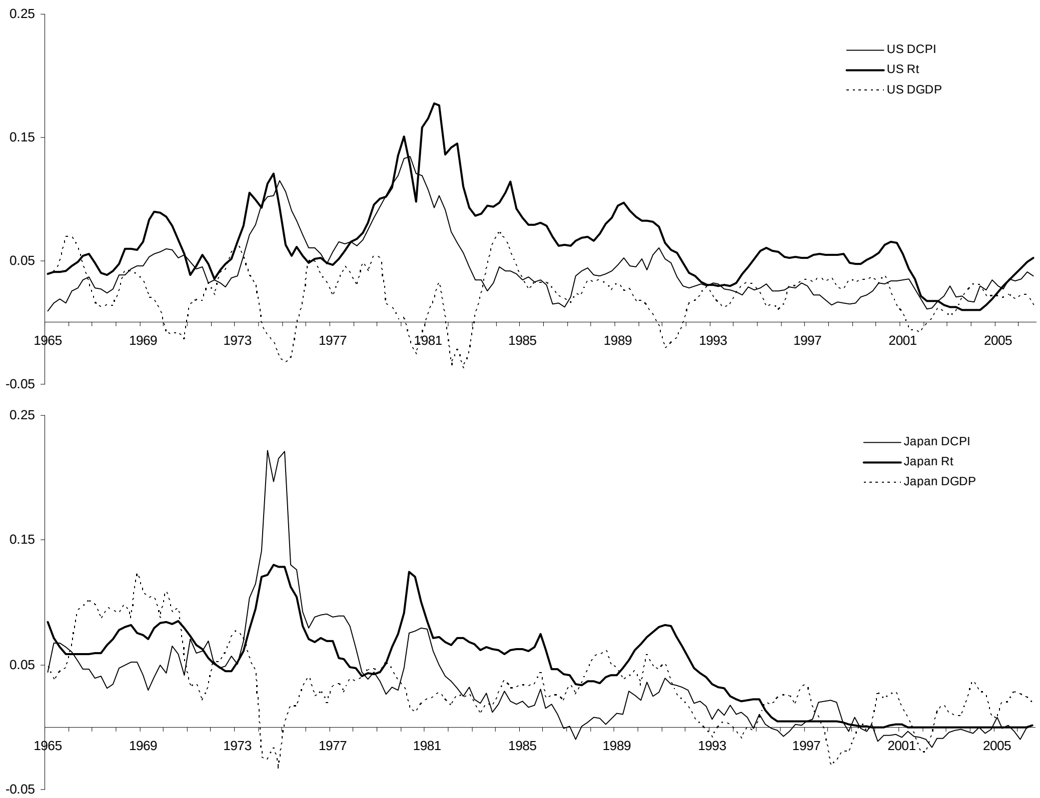

Figure 5: Plot of annual inflation (DCPI or DRPI), annual growth in real per capita GDP (DGDP) and interest rates (Rt) for the US and Japan.

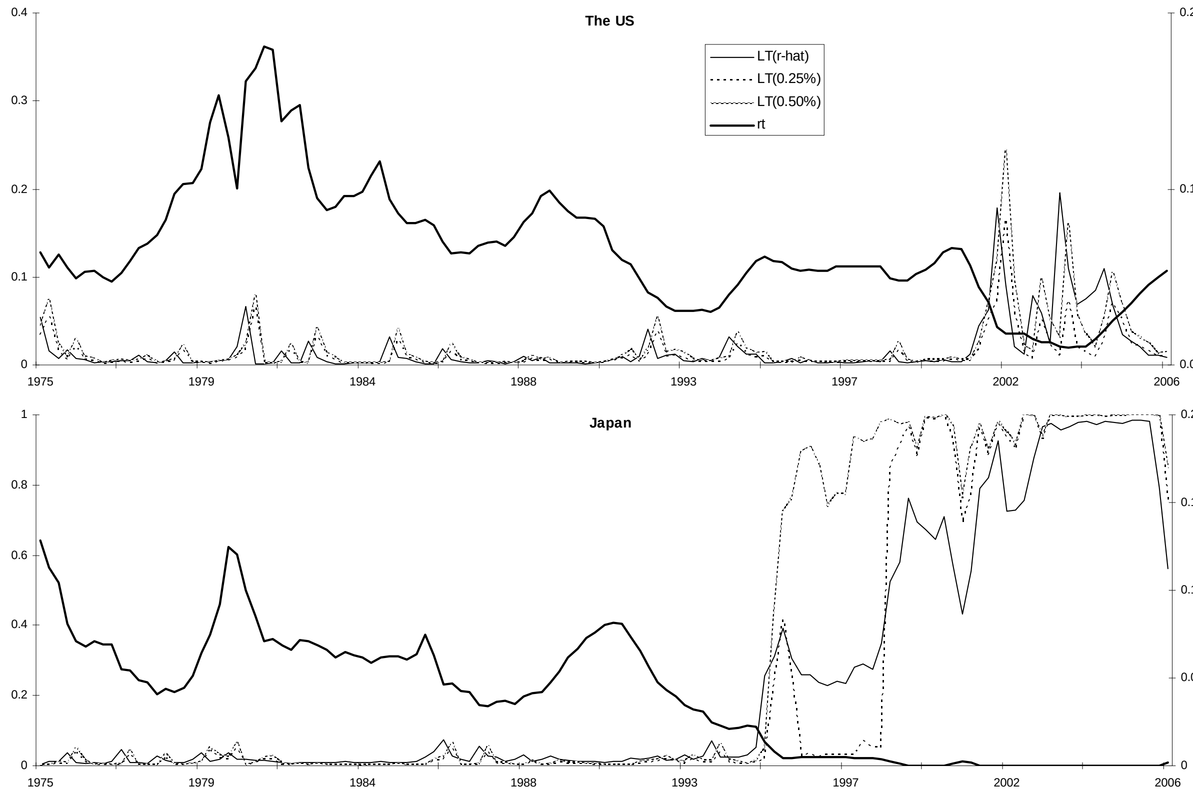

Figure 6: In each panel is plotted: the estimated probability of LT (Left hand scale) when $\boldsymbol{r}$ is estimated, LT(r-hat); the estimated probabilities of LT when $\boldsymbol{r} = 0.25\%$ and 0.5%, LT(0.25%) and LT(0.50%); and the interest rate, rt (Right hand scale). Note that the scale is different on the left for the US to that for Japan.

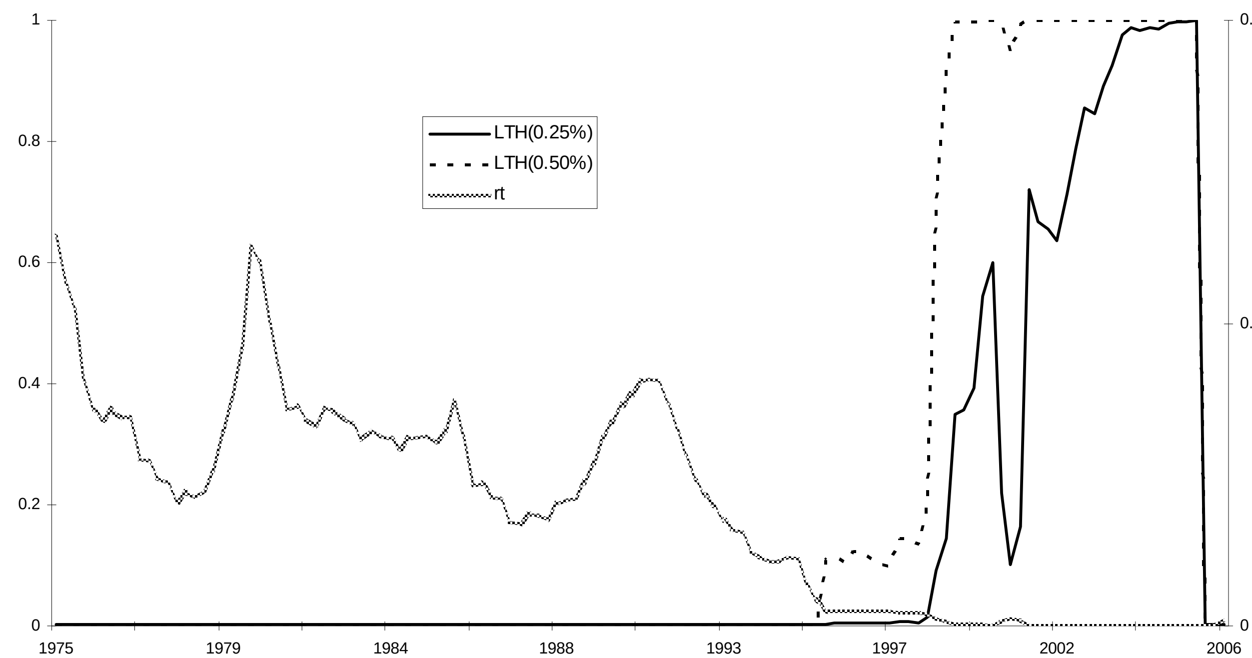

Figure 7: This figure plots the estimated probability of LT (Left hand scale) for Japan for $\boldsymbol{r} = 0.25\%$ and 0.5% without stochastic volatility The estimates without SV are the dashed lines LTH(0.25%) and LTH(0.50%). The interest rate is rt (Right hand scale).