WORKING PAPERS IN ECONOMICS & ECONOMETRICS

Evidence on a Real Business Cycle model with Neutral and Investment-Specific Technology Shocks using Bayesian Model Averaging.

Rodney W. Strachan
Research School of Economics
College of Business and Economics
Australian National University
E-mail: rodney.strachan@anu.edu.au

Herman K. van Dijk
Econometric and Tinbergen Institutes
Erasmus University Rotterdam

JEL codes: C11, C32, C52

Working Paper No: 522
ISBN: 086831 522 2

May 2010

Evidence on a Real Business Cycle model with Neutral and Investment-Specific Technology Shocks using Bayesian Model Averaging.

Rodney W. Strachan[1]

Research School of Economics
The Australian National University
email: rodney.strachan@anu.edu.au

Herman K. van Dijk

Econometric and Tinbergen Institutes
Erasmus University Rotterdam
email: hkvandijk@few.eur.nl

[1]Corresponding author. Strachan is also a Fellow of the Rimini Centre for Economic Analysis.

ABSTRACT

The empirical support for a real business cycle model with two technology shocks is evaluated using a Bayesian model averaging procedure. This procedure makes use of a finite mixture of many models within the class of vector autoregressive (VAR) processes. The linear VAR model is extended to permit cointegration, a range of deterministic processes, equilibrium restrictions and restrictions on long-run responses to technology shocks. We find support for a number of the features implied by the real business cycle model. For example, restricting long run responses to identify technology shocks has reasonable support and important implications for the short run responses to these shocks. Further, there is evidence that savings and investment ratios form stable relationships, but technology shocks do not account for all stochastic trends in our system. There is uncertainty as to the most appropriate model for our data, with thirteen models receiving similar support, and the model or model set used has significant implications for the results obtained.

Key Words: Posterior probability; Real business cycle model; Cointegration; Model averaging; Stochastic trend; Impulse response; Vector autoregressive model.

JEL Codes: C11, C32, C52

1 Introduction.

In this paper we evaluate the robustness, in face of model uncertainty, of the empirical support for a real business cycles (RBC) model with two types of technology shocks: investment-specific technology shocks; and neutral technology shocks. We use sets of VAR models and take into account model uncertainty using a Bayesian model averaging approach. Our work is distinguished from most other model averaging papers since averaging over systems of variables (rather than single equation models) implies averaging over features of the model rather than averaging over sets of regressors. Although averaging over models of systems adds a level of complexity, the approach we propose makes such as exercise feasible and the results suggest the exercise is worthwhile.

The RBC model investigated in this paper is based upon one described by Fisher (2006). This model has several clear and empirically testable implications for the econometric model, but at the same time gives no direction on other features of the econometric model. For example, the economic model suggests that the Great Ratios (e.g., consumption to income, investment to income) are stationary, that only investment-specific technology shocks have permanent effects on the real investment good price, and only technology shocks affect productivity in the long run. However, little direction is given on the form of deterministic trends or the lag structure required to produce the short run dynamics in the various processes. We measure model uncertainty using Bayesian model averaging (BMA) and develop the method of model averaging over reduced form vector autoregressive (VAR) models, although these are rearranged into vector error correction models (VECM) to more easily parameterize restrictions. The model uncertainty derives from uncertainty over the number of stochastic trends present in the system, the form of the deterministic trends, lag length, the form of the reduced form equilibrium (cointegrating) relations and long run restrictions on responses to shocks.

Applied economists have become comfortable accounting for parameter uncertainty in inference, but model uncertainty is less commonly taken into account. It is rare, however, that a well specified economic model will find unequivocal empirical support when confronted with a wide range of alternative econometric models, only a few of which derive from the economic model.

This paper proposes evaluating the evidence on features of econometric models rather than models themselves. An economic model will often suggest features that an econometric model should have. No one feature need be associated with only one econometric model and few econometric models will have all of the features implied by the theory. By considering the unconditional[2] evidence on individual features of the model, it is possible to identify those features that have stronger empirical support. The joint evidence for those features implied by the economic model will indicate the empirical support for the economic model. In this sense, our analysis de-couples the dependence of the economic model from any one econometric model.

The idea underlying BMA 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), an increasing number of recent applications suggest its relevance for macroeconometrics (Fernández, Ley and Steel, 2001, Sala-i-Martin, Doppelhoffer and Miller, 2004, Koop and Potter, 2003 and Wright, 2008). There are several arguments for model averaging and only a few are mentioned here. At the simplest level, it is often attractive to report inferences robust to model specification. A large body of applied work has demonstrated the that averaging results in gains in forecasting accuracy (Bates and Granger (1969), Diebold and Lopez (1996), Newbold and Harvey (2001), Terui and van Dijk (2002), Ravazzolo, van Dijk and Verbeek (2007) and Wright (2008)). Some explanation for this phenomenon in particular cases was provided by Hendry and Clements (2002). Methodologically, averaging over models addresses to some degree the well understood pre-test problem (see, for example, Poirier, 1995, pp. 519-523).

In macroeconomic analysis it is not only the regressors that tend to differ between models, but also the structure[3] or features of the model. The majority of work using model averaging techniques have used single equation, linear regression models. This paper differs from most other model averaging papers as it averages over systems that have interesting and complex features, and so it does not only average over alternative regressors. Averag-

[2] Here 'unconditional evidence' means that the empirical evidence does not depend upon a single model or, in particlular, the other features in that model.

[3] We prefer to use the word 'features' rather than 'structures' to avoid confusing our work with structural VAR analyisis. We consider the restrictions on the reduced form model including those required to identify structural shocks in an SVAR.

ing over systems increases the computational complexity and requires careful consideration of the prior distributions.

This paper makes three contributions. First, the argument is developed for using the model specification and prior of Strachan and Inder (2004) with new results. We show in this paper how to obtain posterior inference from model averages in which the economically and econometrically important features may have weights other than zero or one. In other words, the inferences are based on a finite mixture of models. Second, this paper demonstrates how to estimate probabilities and parameters in models that incorporate restrictions on the responses described in Fisher (2006) in the presence of cointegration. Prior equality and inequality conditions are included in the parameter space of structural VARs and we demonstrate how to compute the posterior probabilities of such restrictions. An important component of this contribution is an approach to estimating a model with restrictions on the long run response matrix in a Beveridge-Nelson decomposition of the Wald representation of the process. Subject to cointegration, such restrictions imply highly nonlinear restrictions on the parameters in the mean equations and this complicates sampling and estimation of posterior probabilities. Third, the proposed methodology is demonstrated with an empirical investigation of a RBC model. Important in this model are the long run responses of investment prices and productivity to technology shocks and that technology follows stochastic rather than deterministic trends.

The structure of the paper is as follows. In Section 2 the important features of the economic model used by Fisher (2006) are outlined. We do not develop the model or discuss the underlying theory as that is well covered in Fisher's paper and others in the area. Empirically testable features implied by the economic model are identified. In the Section 3 the basic econometric models, the reduced form VECMs, of interest in this paper are introduced, including characterizations of the features implied by the economic model. We present the priors, the likelihood and the sampling scheme used to simulate the posterior in Section 4. The tools for inference in this paper, posterior probabilities, are introduced. The evidence for the alternative restrictions are presented in Section 5 as are estimates of important functions. In Section 6 we summarize conclusions and discuss possibilities for further research.

2 A Real Business Cycle Model.

In this section we outline the features of the real business cycle model of Fisher (2006), which is in turn based upon the competitive equilibrium growth model of Greenwood, Hercowitz, and Krusell (1997) with two simplifications: capital is not separated into equipment and structures; and technologies are given stochastic rather then deterministic trends. The general model was developed in Kydland and Prescott (1982) and detailed in King, Plosser and Rebelo (1988), and an interesting early econometric analysis is provided in King, Plosser, Stock and Watson (1991). The reader is directed to these earlier papers for the development of the model as we focus upon certain features that imply empirically testable restrictions upon our reduced form econometric model.

The model suggests that a system of consumption, C_t, investment, X_t, and output, $W_t = C_t + X_t$, will share a balanced growth path since each will be driven by shocks to two technologies: an investment specific technology, V_t; and neutral technology, A_t. We denote the logs of C_t, X_t, and W_t by c_t, x_t, and w_t respectively.

The resource constraint and Cobb-Douglas production technology are given by

$$C_t + X_t \leq A_t K_t^{\lambda} H_t^{1-\lambda}, \quad 0 < \lambda < 1$$

and period $t+1$ capital stock is given by

$$K_{t+1} \leq (1-\delta) K_t + V_t X_t, \quad 0 < \delta < 1.$$

Fisher (2006) specifies technology as having stochastic rather than deterministic trends. The log of investment-specific technology, $v_t = \ln(V_t)$, and the log of neutral technology, $\mathrm{a}_t = \ln(A_t)$, are assumed to be simple random walks, possibly with drifts, and with independent innovations. In the empirical analysis we evaluate the evidence on the importance of deterministic and stochastic trends as well as the relative contribution to business cycle volatility of permanent and transitory shocks.

An implication of the production technology and the resource constraint is that we can represent the log real price of an investment good in consumptions goods by $p_t = -v_t$ and

$$p_t = p_{t-1} - \nu - \varepsilon_{\nu,t}.$$

Since $\nu \geq 0$, this justifies the downward trend we see in the price of an investment good. Neutral technology evolves by the process

$$\mathrm{a}_t = \gamma + \mathrm{a}_{t-1} + \varepsilon_{a,t}$$

where $\gamma \geq 0$ and $\varepsilon_t = (\varepsilon_{v,t}, \varepsilon_{a,t})'$ has zero mean and constant covariance matrix.

A first implication of this model is that the variables c_t, x_t, and w_t will all be integrated of order one due to a common stochastic trend given by $\omega a_t + (1-\omega)\, p_t$ and the differences between any two will be stationary. This is not an unusual result in the balanced growth literature (see, for example, King, Plosser, Stock and Watson, 1991) and it implies that we can treat the relations

$$c_t - w_t \text{ and } x_t - w_t$$

as valid cointegrating relations.

Denote by $h_t = \ln(H_t)$ the log number of hours worked which is assumed to have no unit root, although it may have a trend. The log price of an investment good, p_t, and labour productivity, $a_t = \ln(W_t/H_t) = w_t - h_t$, are assumed to have unit roots but p_t should not cointegrate with the other variables. Since h_t is assumed to be $I(0)$, and c_t, w_t and x_t are all assumed to be $I(1)$ sharing a common stochastic trend, the above assumptions imply that a_t will be $I(1)$ and the relations

$$c_t - w_t + h_t = c_t - a_t \text{ and } x_t - w_t + h_t = x_t - a_t$$

will be $I(0)$ and form valid cointegrating relations. None of the assumptions preclude the above $I(0)$ relations having deterministic trends, but this is not a feature we would expect to find.

Two important final restrictions apply to the long run responses of the real price of investment, p_t, and labour productivity, a_t, to technology shocks. Fisher (2006) assumes that the long run responses of investment prices and productivity only respond in the long run to the technology shocks, and investment prices respond only to investment-specific technology shocks. That is, the long run response of p_t to an investment-specific technology shock will be nonzero, in fact negative, but its long run response to all other shocks will be zero. Second, the long run response of a_t to both an investment-specific and a neutral technology shock will be nonzero, but the long run response of a_t to any other shock will be zero. These restrictions identify the investment-specific technology shock, the I-shock, and the neutral technology shock, the N-shock.

Fisher imposes an additional restriction that, in the long run, the investment specific shocks lower the price of an investment good by an amount of known proportion to the amount that it will raise labour productivity. This assumption implies a linear restriction on the long-run responses of p_t and a_t to an investment-specific shock. The proportion is given as a function of the elasticity parameter in the production function; specifically the proportion is $\frac{1-\lambda}{\lambda}$. Imposing this restriction then requires that we know λ and Fisher uses a value of $\lambda = 1/3$ for the simulation experiment and $\lambda = 1/4$ for the econometric analysis. This restriction is not necessary to identify the shocks and we do not impose it. However, we use the relationship (detailed in Fisher (2006)) between the long run responses to estimate λ. The resulting estimates suggest that the values Fisher used were very reasonable. As we report the full posterior distribution of λ from a range of models, we are able to characterize more fully the uncertainty associated with λ.

3 A Set of Vector Autoregressive Models.

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 $y_t = (p_t, a_t, h_t, c_t, x_t)$, $t = 1, \ldots, T$, conditioning on $l+1$ initial observations is

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

where $\Delta y_t = y_t - y_{t-1}$. The $1 \times n$ vector of errors u_t are assumed to be $iidN(0, \Omega)$.[4] 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. We define the deterministic terms $d_t\mu$ below.

Next, we specify the model set which is defined by the combinations of restrictions imposed upon the VAR. The restrictions refer to particular types of *deterministic* processes (indexed by d), the *lag* length (l), the *number* of cointegrating relations (r), the *(over)identification* restrictions on the cointegrating space (o), and the long run restrictions identifying technology shocks (s).

The number cointegrating relations, r, determines the dimensions of β and α and the number of stochastic trends in the system as $n-r$, where $r =$

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

$0, 1, \ldots, n$. Different overidentifying restrictions on β are denoted by o, where $o \in \{0, 1, 2\}$. If $o = 0$ then no overidentifying restrictions are imposed on β. If $o = 1$ then it is assumed that p_t has a unit root but does not cointegrate with the other variables in the system. If $o = 2$ then the restriction implied when $o = 1$ is imposed and, further, that hours worked, h_t, and the great ratios of consumption to income and investment to income are stationary. The restrictions $o = 1$ and $o = 2$ imply a model specification in which $\beta = H_1\psi$ or $\beta = H_2\psi$ respectively for appropriate H_1 and H_2. The restriction implied when $o = 2$ is stronger than that under $o = 1$ since $sp\,(H_2) \subset sp\,(H_1)$ and these overidentifying restrictions imply maximum cointegrating ranks such that if $o = 1$ then $r \leq 4$ and if $o = 2$ then $r \leq 3$. Clearly then some models, such as $(r = n, o = 1)$ or $(r = n - 1, o = 2)$, are *a priori* impossible and will be assigned zero prior probability.

We allow for five different lag lengths such that $l \in \{0, 1, 2, 3, 4\}$. The deterministic processes are denoted by $d \in \{1, 2, 3, 4, 5\}$ and these processes, given in the table below,are the five most commonly used combinations (see, for example, Johansen, 1995):

d	$y_t\beta$	y_t
1	linear trend	quadratic drift
2	linear trend	linear drift
3	non-zero mean	linear drift
4	non-zero mean	no drift
5	zero mean	no drift

Some models implied by the deterministic processes will be observationally equivalent. For example, if $r = 0$ then the models with $d = 2$ or $d = 3$ will be observationally equivalent as will the models with $d = 3$ and $d = 4$ when $r = n$. The treatment of *a priori* impossible and observationally equivalent models is explained in the next section when the prior is outlined.

Finally, the long run restriction to identify the technology shocks is employed. As discussed in the previous subsection, this restriction implies that the long run response of p_t is nonzero only for the investment-specific technology shocks and that the long run response of a_t is nonzero only for the investment-specific technology shock and the neutral technology shocks. This restriction can be parameterized using the standard Beveridge-Nelson form of the Wald representation of the VECM as

$$\Delta y_t' = Cu_t' + C^*(L)\,\Delta u_t' \text{ where } C = \beta_\perp \left(\alpha_\perp \Gamma \beta_\perp\right)^{-1} \alpha_\perp.$$

The restriction on C implies the matrix will have the following zero entries:

$$C = \begin{bmatrix} c_{11} & 0 & 0 & & 0 \\ c_{21} & c_{22} & 0 & \cdots & 0 \\ * & * & * & & * \\ & \vdots & & \ddots & \\ * & * & * & & * \end{bmatrix}$$

where the asterisks (*) imply no restriction is imposed.

It seems reasonable to assume that as this is a long run restriction, imposing it has no implications for the short run dynamics and so does not imply any restrictions on Γ or Ω, but must imply restrictions on α and β. Therefore it is necessarty to recover the values of α and β after imposing the long run restriction and these new restricted values are denoted by $\widetilde{\alpha}$ and $\widetilde{\beta}$. With the ordering of the variables in the system given above, a Cholesky decomposition of $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$ where $\Omega^{1/2}$ is lower triangular, imposes no further identifying restrictions on the model, however these identifying restrictions imply that the long run restrictions will be testable overidentifying restrictions. Without the over identifying restrictions on the matrix C we do not have an interpretation for the shocks identified by the Cholesky decomposition of Ω. This is not a problem as we are interested in the technology shocks and so only consider responses from models with $s = 1$.

Next we explain how, once the restriction is imposed on C, the restricted values of α and β are recovered. Note that there is no information in C on the orientation of α or β within the spaces they span, only the space of α and the space of β can be obtained from C. This can be seen as, for any full rank $r \times r$ matrices κ_a or κ_b, we can write C as

$$C = \beta_\perp \left(\alpha_\perp \Gamma \beta_\perp\right)^{-1} \alpha_\perp = \beta_\perp \kappa_b \left(\kappa_a \alpha_\perp \Gamma \beta_\perp \kappa_b\right)^{-1} \kappa_a \alpha_\perp.$$

The space of $\beta_\perp$ defines the space of β and, similarly, the space of $\alpha_\perp$ defines the space of α. It is possible to recover the space of $\beta_\perp$ and the space of $\alpha_\perp$ from C, however, no further information on β or α can be retrieved. Writing $\alpha = \left(\alpha\alpha'\right)^{1/2} V$ and $\beta = U \left(\beta'\beta\right)^{1/2}$ where $VV' = U'U = I_r$, then the matrices $\left(\alpha\alpha'\right)^{1/2}$ and $\left(\beta'\beta\right)^{1/2}$ can be thought of as the norms of the matrices α and β. These results imply that a restriction on C does not restrict Ω, $\alpha\alpha'$ or $\beta'\beta$, but will restrict $sp\left(\beta\right) = sp\left(U\right)$ and $sp\left(\alpha'\right) = sp\left(V'\right)$.

Denote by $\widetilde{C}$ the matrix C with the restrictions imposed. To recover the restricted $sp\left(\beta\right)$ and $sp\left(\alpha'\right)$ from $\widetilde{C}$, first observe that as $\widetilde{C}$ has rank $n - r$, it may be written as $\widetilde{C} = \gamma\eta'$ where γ and η are $n \times (n-r)$ full rank $(n-r)$ matrices. Necessarily $sp(\gamma) = sp\left(\widetilde{\beta}_\perp\right)$ and $sp(\eta) = sp\left(\widetilde{\alpha}'_\perp\right)$. A singular value decomposition of $\widetilde{C}$ into $\widetilde{C} = \widetilde{U}_\perp S \widetilde{V}'_\perp$ will then give semi-orthogonal matrices $\widetilde{U}_\perp$ and $\widetilde{V}'_\perp$ such that

$$\begin{aligned} sp\left(\widetilde{U}_\perp\right) &= sp(\gamma) = sp\left(\widetilde{\beta}_\perp\right) \text{ and} \\ sp\left(\widetilde{V}'_\perp\right) &= sp(\eta) = sp\left(\widetilde{\alpha}'_\perp\right). \end{aligned}$$

From $\widetilde{U}_\perp$ and $\widetilde{V}'_\perp$ construct semi-orthogonal matrices $\widetilde{U}$ and $\widetilde{V}$ such that $sp\left(\widetilde{U}\right) = sp\left(\widetilde{\beta}\right)$ and $sp\left(\widetilde{V}'\right) = sp\left(\widetilde{\alpha}'\right)$ and use these to construct $\widetilde{\alpha}$ and $\widetilde{\beta}$ as $\widetilde{\alpha} = (\alpha\alpha')^{1/2}\widetilde{V}'$ and $\widetilde{\beta} = \widetilde{U}(\beta'\beta)^{1/2}$.

Since C has rank $n-r$, the zero restrictions on C and the assumption of nonzero responses of p_t and a_t stated above implies C must have at least rank two, the restriction can only apply if $r \in \{0, 1, \ldots, n-2\}$. This is consistent with the two technology shocks entering the system as stochastic trends. The index for this long run restriction is s and we set $s = 0$ when we do not impose the restrictions and $s = 1$ when the restrictions are imposed.

In summary, each model will be defined by the combination of the deterministic process (d), lags of differences (l), cointegrating rank (r), overidentifying restrictions on the cointegrating space (o), and whether or not the long run responses are restricted (s). Each model will be identified by M_i where $i = (d, l, r, o, s)$ and $i \in \Xi$, the set of all i considered. As an example of some models we will use, suppose we allow a linear drift and nonzero mean in the cointegrating relations $(d = 3)$, two lags of differences $(l = 2)$, stationary great ratios and hours worked $(r = 3, o = 2)$ and the long run identifying restrictions for the technology shocks are imposed $(s = 1)$. This model would be denoted as $M_{(3,2,3,2,1)}$. In total we average using 508 models in our application.[5]

[5] There are 900 models implied by the restrictions. However, this reduces to 508 models when we exclude *a priori* impossible models, meaningless models and only consider one in a set of observationally equivalent models.

4 Priors, Posteriors and Model Averaging.

In this section the priors and resultant posterior are presented. We begin with discussion of the prior model probabilities taking into account that some models are impossible and others that are observationally equivalent. Next we consider the priors for the covariance matrix and the mean equation parameters excluding (β, α), that is: μ; Γ_1; $\ldots$; Γ_l. For notational convenience we collect these mean equation parameters into a $k_i \times n$ matrix $\Phi = [\mu' \quad \Gamma_1' \quad \cdots \quad \Gamma_l']'$ and vectorize into $\phi = vec(\Phi)$. Conditional upon β, the model in (1) is linear in the equation parameters $vec(\alpha)$ and ϕ. This fact makes it relatively straightforward to elicit priors on Ω and ϕ, however we adopt a transformation that improves the sampling scheme. For this reason we give the full prior after we have given careful consideration to the prior for β, before then presenting the method of posterior analysis.

4.1 The Prior.

Ideally all models would be treated as *a priori* equally likely, however this is not a straightforward issue in VECMs.[6] The priors for the individual elements of $i = (d, l, r, o, s)$ are not independent, as certain combinations are either impossible (such as when $r = n$ and $o = 2$), meaningless (such as, for example, $r = 0$ with $o = 1$) or observationally equivalent to another combination (such as the models with $r = n$ and $d = 1$ or 2). The prior probability for impossible and meaningless models is set to zero. However, the researcher must carefully consider how she wishes to treat observationally equivalent models. Treating these models as just one model and then assigning equal prior probabilities to all models biases 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.[7] Alternatively, these could be treated as separate models. A choice must be made and in this paper, observationally equivalent models are treated as one model.

A referee has raised the interesting question as to whether it is appropriate to specify independent priors for d, l, r, o, and s. One might expect, for

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

[7] 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.

example, that a strong deterministic process such as $d = 1$ might reduce the prior expectation of finding stochastic trends in the processes. This might imply that the probabilitiy $\Pr(r < n|d)$ may decrease as d increases. Similarly a shorter lag length, l, might be associated with a higher prior probability of finding (more) stochastic trends. We do not pursue this idea further, but note that it might be a worthwhile topic for investigation.

For each model we use a proper prior for Ω that is an inverted Wishart with scale matrix $\underline{S} = I_n 10$ and degrees of freedom $\underline{\nu} = n + 1$ as this prior is rather uninformative. We specify a weakly informative proper prior for α, however, defer specification of the full prior to the end of the next subsection, but the prior for $vec(\alpha)$ conditional upon (Ω, β, M_i) (and hyperparameters discussed below) has zero mean and covariance matrix $\frac{1}{\eta}\underline{V}_a$ where $\underline{V}_a = \Omega \otimes I_r$.[8]

For μ and Γ_i, we had initially specified a normal prior with zero mean and covariance matrix $\frac{1}{\eta}\underline{V}_0$ where $\underline{V}_0 = \Omega \otimes I_{k_i}$, however a referee pointed out that it would make more sense that coefficient matrices for higher lags are more likely to be near zero. This suggests using the well known Litterman prior (Litterman, 1980, 1986, Doan, Litterman and Sims, 1984). As we average over models with different numbers of lags we feel we already allow the data to choose shorter lags, however, as we have already mentioned, shrinkage tends to improve inference (Ni and Sun (2003)) which suggests a technical reason to prefer the Litterman prior. To express our uncertainty as to which is the correct prior, we specify the prior for Γ_i to be a mixture of two normal zero mean priors. One with covariance matrix $\underline{V}_0$ and the other with the Litterman type prior for a VECM specified in Villani (2001).

The covariance matrix in Villani (2001), which we will denote by $\frac{1}{\eta}\underline{V}_1$, has zero off-diagonals and the variance of each element of Γ_i shrinks toward zero the higher is i and for off diagonal elements of Γ_i. The coefficients for own lags are not quite as heavily shrunk as the coefficients for other variable lags. The full covariance matrix for ϕ can be represented as $\frac{1}{\eta}\underline{V}_\phi$ where $\underline{V}_\phi = (Iu + \underline{V}_0(1-u))(I(1-u) + \underline{V}_1 u)$ where $u \in \{0, 1\}$ with prior probabilities $\Pr(u = 1) = \Pr(u = 0) = 0.5$.[9] The posterior estimate of u will inform us on the data's preference between the two specifications and in this

[8] If an informative prior is used on for the cointegrating space then we recommend the prior for α described in Koop, León-González and Strachan (2008).

[9] Alternatively we could give u a continuous distribution over $[0, 1]$ and mix continuously over the two normals. Either approach seems reasonable.

sense produces an empirical Bayes prior for ϕ. We found the posterior was not very informative on the choice of u (either 0 or 1). The estimated posterior probability that the Litterman prior was adopted, i.e., the estimate of $\Pr(u = 1|y)$, was 0.29 which shows some preference for $\underline{V}_0$, but this does not indicate strong evidence for either covariance matrix. It would seem, therefore, that mixing over the two normals, rather choosing one, is a reasonable approach. However, mixing the densities produced a posterior probability of one that there are no lags in the model. This result seems due to the extra shrinkage implied by the Litterman prior.

The parameter η determines the overall degree of shrinkage that is applied to the mean equation parameters. Further evidence on the influence of this parameter can be found in Strachan and Inder (2004). A gamma prior with mean $E(\eta) = 5$ and a relatively large variance $V(\eta) = 16.67$ is specified for η. These settings provide a reasonable degree of shrinkage towards zero which has been shown to improve estimation (see Ni and Sun (2003)). The posterior distribution of η, by contrast, is very tight with a mean of $E(\eta|y) = 0.001$ and variance $V(\eta|y) = (0.0195)^2$. This result suggests the data prefer less shrinkage, although the Litterman prior already imposes a significant degree of shrinkage. Setting the prior mean of η (and therefore variance) to a larger value did not significantly change the posterior estimates of other objects of interest (such as impulse responses). We concluded that while the bulk of the posterior mass of η is near zero, there is sufficient mass away from zero to give enough shrinkage.

This paper argues, with new results to support this argument, for using the model specification and prior of Strachan and Inder (2004) as a more general and less problematic approach than what is commonly used in these models. The general argument is that any inference that can be achieved with linear identifying restrictions can be achieved with the identifying restrictions in this paper, and this inference can be achieved without encountering the issues that surround the linear restrictions.

As in many reduced rank models, there is a well known identification issue since β and α appear as a product in (1) such that $\beta\alpha = \beta\gamma\gamma^{-1}\alpha = \beta^*\alpha^*$ and (β, α) and (β^*, α^*) are observationally equivalent. What is not often recognized in the cointegration literature is that the space of β and the space of α are fully identified under the likelihood and, without restrictions on α, the data can only inform us about the space of β. Any further restrictions, such as to identify the elements of β and α to permit interpretation, are necessarily part of the prior and will potentially have implications for posterior inference.

In the Bayesian literature it is common to use linear identifying restrictions to impose restrictions to permit interpretation and estimation. 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 $B = c_{\perp}\overline{\beta}$ where $c_{\perp}c' = 0$. For example, if $c = [I_r \ 0]$ then $\overline{\beta} = [I_r \ B']'$ and a prior is then specified for B.[10] We do not impose such restrictions for several reasons: empirically they can have a determining effect on inference (Lopes and Wests (2004)); imposing such restrictions has the unexpected and undesirable result that it makes the assumption supporting the restrictions *a priori* impossible (Strachan and Inder, 2004); these restrictions have been associated with nonexistence of moments, improper posteriors, local nonidentification problems and reducibility of Markov chain methods (Kleibergen and van Dijk (1994 & 1998) and Bauwens and Lubrano (1996)).

A further problem with the priors that use linear identifying restrictions is that the posterior is improper at any point where the restriction $Rvec(\alpha) = \mathfrak{r}$ for any known R and $\mathfrak{r}$ (see Appendix I). This issue has not been discussed in the literature but has the same implications as local nonidentification. Is it important to stress that this result holds for *any* known R and $\mathfrak{r}$ and so implies an almost everywhere covering of the support.

This paper uses a specification of Strachan and Inder (2004) that permits estimation with minimal restrictions. To implement this approach, specify β to be semi-orthogonal, i.e., $\beta'\beta = I_r$, and specify a Uniform distribution for β (for background information, see Strachan (2003), Strachan and Inder (2004), Strachan and van Dijk (2003) and Villani (2005)). This approach does not preclude achieving interpretable coefficients, B, by imposing such identifying restrictions as these can be imposed ex-post once a draw or an estimate of β is obtained. As many choices of identifying restrictions can be imposed to permit as many interpretations of the coefficients is desired. The difference is that these restrictions are imposed on draws or estimates from the posterior and not in the prior. This approach avoids problems of local non-identification, all moments exist and sampling is simplified. We make the argument, and we think this to be a compelling argument, that there is nothing gained but much potentially is lost from imposing the linear normalization a priori. Using the normalization of Strachan and Inder (2004),

[10] 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.

coefficients from whatever linear normalization the researcher wishes can be retrieved ex-post.[11]

This approach to identification is closer to the identifying restrictions used in classical models with reduced rank structures. For example, the well known Johansen method of identifying the cointegrating vectors uses a similar approach. as do other nonlinear models. See, for a further example, the multi-mode model discussed in Magnus and Neudecker (1988).

For the cases in which identifying restrictions discussed in Section 2 of the form $\beta = H\psi$ $(o = 1)$ are imposed, set ψ where $\psi'\psi = I_s$ and give ψ a Uniform prior. For computational and mathematical simplicity, we convert H to be semiorthogonal by the transformation $H \rightarrow H\left(H'H\right)^{-1/2}$. 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 choosen for the prior. Therefore, to obtain an expression for the posterior useful for obtaining draws of β, we use an approach proposed in Koop, León-González and Strachan (2010). Note that the matrices α and β always occur in a product form as $\beta\alpha$ such that it is possible to introduce any full rank square $r \times r$ matrix κ such that $\beta\alpha = \beta\kappa\kappa^{-1}\alpha = \beta^*\alpha^*$ without affecting the posterior.The matrices α^* and α have the same support, however, β is semiorthogonal with the Stiefel manifold (see Muirhead, 1982 or James, 1954) as its support while β^* has as its support the nr dimensional real space. The matrix β^* is given a Normal prior with zero mean and covariance matrix $n^{-1}I_{nr}$. Transforming back to the parameters of interest is straightforward via $\beta = \beta^*\kappa^{-1}$ and $\alpha = \alpha^*\kappa$. The prior for β^* resembles that of Geweke (1996) except that our prior implicitly specifies, in addition to a proper prior for κ, that the marginal prior for $\beta = \beta^*\kappa^{-1}$ is Uniform. The efficiency of this approach is discussed in Koop, León-González and Strachan (2010).

The prior is specified to impose an important inequality constraint. The VECM is assumed to be balanced, in that all stochastic elements are $I\left(0\right)$,

[11]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. As we prove in Appendix II, no proper prior in the literature is invariant. 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.

and so the support under the prior is restricted to exclude explosive unit roots: call this restricted region the stationary region. Imposing a restriction on the support of the parameters implies the prior must be renormalized so that it integrates to one. The renormalizing constant, p_s, is the probability mass in the stationary region under the prior described above over the unrestricted support.

To give this a more formal explanation let the vector of all parameters in the model that appear in the likelihood, i.e., $\mathfrak{p}, \alpha, \phi$, and Ω , be denoted by θ and the unrestricted support is Θ, $\theta \in \Theta = G_{r,n-r} \times R^{n(k_i+r)} \times R_+^{n(n+1)/2}$ (where $R_+^{n(n+1)/2}$ denotes the blunt, one-sided cone that forms the support of all $n \times n$ positive definite symmetric matrices) and let the full prior be denoted as $p(\theta)$. Next, denote the stationary region as $\Theta_S \subset \Theta$ and an indicator function for this region as $1(\theta \in \Theta_S)$. Then the renormalizing constant $p_S = \int_{\Theta_S} p(\theta)\, d\theta$ and the prior with the restriction imposed is $p_S(\theta) = p(\theta)/p_S$.

Let $a^* = \left(vec(\alpha^*)', \phi'\right)'$, $b^* = vec(\beta^*)$ and

$$\underline{V} = \begin{bmatrix} \underline{V}_a & 0 \\ 0 & \underline{V}_\phi \end{bmatrix}.$$

Introduce θ^* as the vector containing the elements of β^*, a^*, and Ω. The full prior distribution for the parameters in a given model is then

$$\begin{aligned} p(\theta^*, \eta, u|M_i) \quad \propto \quad & \exp\left\{-\frac{\eta}{2} a^{*\prime} \underline{V}^{-1} a^* - \frac{n}{2} b^{*\prime} b^*\right\} n^{nr/2} \\ & \times 1(\theta \in \Theta_S)/p_S \\ & \times |\Omega|^{-(\underline{\nu}+n+1+r+uk_i)/2} \exp\left\{-\frac{1}{2} tr \Omega^{-1} \underline{S}\right\} \\ & \times \eta^{\frac{n(k_i+r)+1}{2}} \exp\left\{-\frac{5\eta}{6}\right\}. \end{aligned}$$

The indicator, $1(\theta \in \Theta_S)$, in the prior is expressed in terms of θ rather than θ^* as we have defined the support for θ. We could define the support for θ^* under the restriction, Θ_S^*, however since is $1(\theta \in \Theta_S) = 1(\theta^* \in \Theta_S^*)$, this is not necessary.

4.2 Posterior Analysis.

An expression for the posterior distribution of the parameters for any model given the data is obtained by combining the prior, $p(\theta^*, \eta, u|M_i)$, with the likelihood for the data $L(y|\theta^*, M_i)$ where y represents all data. That is,[12]

$$p(\theta^*, \eta, u|M_i, y) \propto p(\theta^*, \eta, u|M_i)\, L(y|\theta^*, M_i) = k(\theta^*, \eta, u, M_i|y)\,. \quad (2)$$

As the sampler uses a Gibbs sampling scheme, it is necessary to present the conditional posterior for each parameter.

In the following results, we gather together terms to keep expressions notationally concise. Collect $y_{t-1}\beta$ and the vector $z_{2,t} = (d_{2,t}, \Delta y_{t-1}, \ldots, \Delta y_{t-l})$ into the vector $z_t = (y_{t-1}\beta^*, z_{2,t})$, and define the $k_i \times n$ matrix $\Phi = (\mu', \Gamma_1', \ldots, \Gamma_l')'$ and the $(r + k_i) \times n$ matrix $A = [\alpha^{*\prime}\ \Phi']'$.

As the model is linear conditional upon b^*, standard results show that the posterior for a^* conditional on all other parameters will be normal with mean $\overline{a}$ and covariance matrix $\overline{V}$ constructed as

$$\begin{aligned} \overline{a} &= \overline{V}\left(\Omega^{-1} \otimes I_{k_i+r}\right) vec\left(\sum_{t=1}^{T} z_t'\Delta y_t\right) \\ &\text{and} \\ \overline{V} &= \left(\left(\Omega^{-1} \otimes \sum_{t=1}^{T} z_t' z_t\right) + \eta \underline{V}^{-1}\right)^{-1}. \end{aligned}$$

Next, the posterior for b^* conditional upon the other parameters will be normal with mean $\overline{b}$ and covariance matrix $\overline{V}_b$ which are constructed as

$$\begin{aligned} \overline{b} &= \overline{V}_b\left(\alpha^*\Omega^{-1} \otimes I_n\right) vec\left(\sum_{t=1}^{T} y_{t-1}'\left(\Delta y_t - z_{2,t}\Phi\right)\right) \\ &\text{and} \\ \overline{V}_b &= \left[\left(\alpha^*\Omega^{-1}\alpha^{*\prime} \otimes \sum_{t=1}^{T} y_{t-1}' y_{t-1}\right) + nI_{nr}\right]^{-1}. \end{aligned}$$

The posterior for η will be Gamma with degrees of freedom $\overline{\nu}_\eta = n(k_i + r) + 3$ and mean $\overline{\mu}_\eta = 1/\left(a^{*\prime}\underline{V}^{-1}a^* + 5/3\right)/\overline{\nu}_\eta$ (see, for example, Koop (2003)). Finally, u will have a Bernoulli conditional posterior distribution with $\overline{p} = \Pr(u = 1|a^*, \Omega, \beta^*, y)$ equal to

$$\overline{p} = \exp\left\{-\frac{\eta}{2}a^{*\prime}\underline{V}_0^{-1}a^*\right\} / \left[\exp\left\{-\frac{\eta}{2}a^{*\prime}\underline{V}_0^{-1}a^*\right\} + \exp\left\{-\frac{\eta}{2}a^{*\prime}\underline{V}_1^{-1}a^*\right\}\right].$$

We use the following scheme at each step q to obtain draws of $(a^*, \Omega, \beta^*, \eta, u)$:

[12] Note that as η and u are hyperparameters, do not enter the likelihood.

1. Initialize $(\Omega, b^*, a^*, \eta, u) = \left(\Omega^{(0)}, b^{*(0)}, a^{*(0)}, \eta^{(0)}, u^{(0)}\right)$;

2. Draw $\Omega | b^*, a^*, \eta, u$ from $IW\left(\underline{S} + u\eta A'A + \sum_{t=1}^{T} u_t' u_t, T + uk_i + r\right)$;

3. Draw $a^* | \Omega, b^*, \eta, u$ from $N\left(\overline{a}, \overline{V}\right)$;

4. Draw $b^* | \Omega, a^*, \eta, u$ from $N\left(\overline{b}, \overline{V}_b\right)$;

5. Draw $\eta | \Omega, b^*, a^*, u$ from $\text{Gamma}(\overline{\mu}_\eta, \overline{\nu}_\eta)$;

6. Draw $u | \Omega, b^*, a^*, \eta$ from $Bernoulli\left(\overline{p}\right)$;

7. Repeat steps 2 to 6 for a suitable number of replications.

The algorithm described above gives draws from the model without long run or stationarity restrictions. A Metropolis-Hastings algorithm is used to obtain draws from the model subject to the restriction when $s = 1$ (see, for example, Koop pp. 92-99 (2003)) with draws from the unrestricted posterior as the candidate density. As there are fewer parameters in the restricted model than the unrestricted model, we augment the restricted posterior with a normal distribution for the parameters in the first two rows of C that are replaced by zeros in $\widetilde{C}$ (for a similar approach, see Kleibergen and Paap (2002)).

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

$$p\left(M_i | y\right) = \frac{m_i p\left(M_i\right)}{\sum_{j \in \Xi} m_j p\left(M_j\right)} \tag{3}$$

where the summation in the denominator is over all models. The marginal likelihood for any model M_i is given by

$$m_i = \int_{\Theta} k\left(\theta^*, \eta, u | M_i, y\right) d\left(\theta^*, \eta, u\right). \tag{4}$$

There are several ways to compute the posterior probabilities. Rewriting the expression in (3) as

$$p\left(M_i | y\right) = \frac{m_i p\left(M_i\right)}{\sum_{j \in \Xi} m_j p\left(M_j\right)} = \frac{m_i / m_0 p\left(M_i\right)}{\sum_{j \in \Xi} m_j / m_0 p\left(M_j\right)} \tag{5}$$

where the marginal likelihood m_0 is for some model M_0, suggests a way to compute the model probabilities if there is a way to compute the Bayes factor given by the ratio m_i/m_0 for all models.

If M_0 nests within another model M_i in the model set (M_0 need not actually be in the model set considered) then the Savage-Dickey density ratio (SDDR) can be used to estimate m_0/m_i (Verdinelli and Wasserman (1995) and see Koop, León-González and Strachan (2008) for an example of an application of this approach). Each of the models considered in this study collapses to, or nests, a single model at the point where $a^* = 0$. Denote this model, M_0. The SDDR can be computed as the ratio of the marginal posterior to the marginal prior at the point $a^* = 0$. That is,

$$\frac{m_0}{m_i} = \frac{p\left(a^* = 0|M_i, y\right)}{p\left(a^* = 0|M_i\right)}.$$

Given sequences of draws $\left(\Omega^{(q)}, b_\beta^{*(q)}, \eta^{(q)}, u^{(q)}\right)$, $q = 1, \ldots, K$ from the posterior and $\left(\Omega^{(j)}, b_\beta^{*(j)}, \eta^{(j)}, u^{(j)}\right)$, $j = 1, \ldots, K$ from the prior[13], the marginal posterior and prior densities for a^* at the point $a^* = 0$ can be approximated by

$$\begin{aligned}
\widehat{p}\left(a^* = 0|M_i, y\right) &= K^{-1}\sum_{q=1}^{K} p\left(a^* = 0|\Omega^{(q)}, b_\beta^{*(q)}, \eta^{(q)}, u^{(q)}, M_i, y\right) \text{ and} \\
\widehat{p}\left(a^* = 0|M_i\right) &= K^{-1}\sum_{j=1}^{K} p\left(a^* = 0|\Omega^{(j)}, b_\beta^{*(j)}, \eta^{(j)}, u^{(j)}, M_i\right).
\end{aligned}$$

For the model with long run restrictions, $s = 1$, the same form of the SDDR can be used with α and β replaced by $\widetilde{\alpha}$ and $\widetilde{\beta}$.

The SDDR approach described above is used to estimate the Bayes factors for the models with support Θ (i.e., the model without the restriction to the stationary region) as it is easier and faster to sample from this model than from the model with support restricted to the stationary region Θ_S. However the prior is specified to restrict the support for all models to Θ_S. The Bayes factors for the models on Θ_S can be obtained from draws from the prior and posterior on Θ.

[13] It is relatively straightforward to show that the conditional priors are all proper and of standard Normal and inverted Wishart forms. Therefore sampling from the prior is not difficult.

Let m_i be the marginal likelihood for the model on Θ and let m_{i_S} be the marginal likelihood for the same model with the support restricted to Θ_S. Denote by $p(\theta^*, \eta, u|y)$ and $p(\theta^*, \eta, u)$ the posterior and prior respectively over the support Θ (used to compute m_i/m_0 above) and let the indicator function for the stationary region be $1(\theta^* \in \Theta_S)$. As shown in Klugkist and Hoijtink (2007),

$$\frac{m_{i_S}}{m_i} = \frac{\int p(\theta^*, \eta, u|M_i, y)\, 1(\theta \in \Theta_S)\, d(\theta^*, \eta, u)}{\int p(\theta^*, \eta, u|M_i)\, 1(\theta \in \Theta_S)\, d(\theta^*, \eta, u)}.$$

The above expression implies that the Bayes factor m_{i_S}/m_i can be estimated by

$$\frac{\frac{1}{M}\Sigma_{i=1}^{M} 1\left(\theta^{(i)} \in \Theta_S\right)}{\frac{1}{N}\Sigma_{i=1}^{N} 1\left(\theta^{(i)} \in \Theta_S\right)}$$

where the numerator uses draws from the posterior and denominator uses draws from the prior. We can use this simple expression because the restriction from Θ to Θ_S is an inequality restriction. To compute the posterior probabilities of the models on Θ_S, multiply m_i/m_0 by m_{i_S}/m_i to obtain

$$\frac{m_{i_S}}{m_0} = \frac{m_i}{m_0}\frac{m_{i_S}}{m_i}$$

and again an estimate of $p(M_{i_S}|y)$ can be estimated using (5).

Probability estimates were produced from five runs of 1000 after burn-in iterations with different initial conditions to check convergence and the results are very similar from each run. For example, the probabilities differed at most at the second or, more often, the third decimal place.

4.3 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 the approach in this paper over previous approaches is that for all model specifications considered the posterior will be proper and all finite moments of $b^* = vec(\beta^*)$ (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. To report expectations of these objects, it is necessary that the moments of b^* exist.

Suppose we have an economic object of interest ζ which is a function of the parameters for a given model $(\theta^*|M_i)$, $\zeta = \zeta(\theta^*|M_i)$. Examples include estimates of impulse responses, forecasts, or loss functions. To report the unconditional (upon any particular model) expectation of this object it is necessary to estimate

$$E(\zeta|y) = \sum_{i \in \Xi} E(\zeta|y, M_i)\, p(M_i|y)$$

where $E(\zeta|y, M_i)$ is the expectation of ζ from model i. Denote the q^{th} draw of the parameters from the posterior distribution for model M_i as $\left(\theta^{*(q)}\right)$ and so the q^{th} draw of ζ as $\zeta^{(q)} = \zeta\left(\theta^{*(q)}|M_i\right)$. Using M draws of the parameters from the posterior distribution for each of the J models, first obtain estimates of $E(\zeta|y, M_i)$ from each model by

$$\widehat{E}(\zeta|y, M_i) = \frac{1}{M}\Sigma_{q=1}^{M}\zeta^{(q)}.$$

These estimates are then averaged as

$$\widehat{E}(\zeta|y) = \sum_{j=1}^{J} \widehat{E}(\zeta|y, M_i)\, \widehat{p}(M_i|y)$$

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

5 The Application and the Results.

In this section we provide empirical evidence on the support for the real business cycle model with two technology shocks and the various restrictions that this economic model implies for the econometric model. We begin with the posterior probabilities of the features of the reduced form VECM that are implied by this RBC model. We then report the estimates of objects of interest including impulse response functions.

The variables of interest are: log real price of an investment good measured in consumptions units, p_t; log labour productivity, a_t; log number of hours worked, h_t; log of consumption, c_t; and log investment, x_t. The data, which are seasonally adjusted, start in the first quarter of 1948 and end in the second quarter of 2009. Where appropriate, the data are measured in 1996 dollars deflated using a chain-weighted index of consumption prices.

We measure the investment price using an investment deflator divided by a consumption deflator and we follow the approach using real total investment price from the National Income and Product Accounts (NIPA) for the investment price. Alternative approaches to constructing p_t are discussed quite extensively in Fisher (2006) and Greenwood, Hercowitz, and Krusell (1997). These papers raise the issue of the lack of quality adjustment in the NIPA series. However, in a related study, Fisher (2005) concludes important findings are robust to using the NIPA-based total investment price rather than alternatives that address these issues. Therefore, we do not explore the alternative approaches as we assume that the NIPA based measure will be appropriate. We compute the consumption deflator using a Fisher index and data from the Bureau of Economic Analysis on nondurable goods and services.

Productivity is constructed from nonfarm output per hour measured in consumption units and hours worked. Hours worked is hours of all persons in the nonfarm business sector obtained from the FRED (Federal Reserve Economic Data), which sourced this data from the U.S. Department of Labor: Bureau of Labor Statistics. Consumption is personal consumption expenditures less durable goods and investment is gross private domestic investment in consumption units sourced from the Bureau of Economic Analysis.

The cardinal product of the supports of d, l, r, o, and s produces 900 models, however 392 of these models can be excluded as observationally equivalent to another model, impossible or meaningless leaving only 508 models to estimate. Of these models, there were sixteen models with measurable support. Six models accounted for half the posterior mass and thirteen models accounted for 99.99% of the posterior probability mass. The posterior probabilities of the top thirteen models is presented in Table 1. Table 2 presents the same information as the marginal probabilities of the various features of the VECM. Although relatively few models get any support, it is clear that support is fairly evenly spread over the top thirteen models.

According to the model of Fisher, it might be reasonable to expect the model to contain drift terms ($d = 3$) as the technologies are random walks with drifts, but we would not expect trends in the cointegrating relations ($d = 2$) or quadratic trends in the variables ($d = 1$). The deterministic process seems to be reasonable as quadratic drifts are excluded and there are no trends in the equilibrium or cointegrating relations. The models that contain no drifts ($d = 4$ and 5) receive most of the support with some support for drift terms. This result does not exclude drifting behaviour (random walks can still drift without a drift term) but it does suggest the driving effect of technology is weak. These results do not conflict with Fisher's model as the drifts are specified to be non-negative and so may be zero. The presence of a drift, however, is less important in the economic model than the trending behaviour being stochasic rather than deterministic, and the evidence is conclusively in favour of this feature.

The posterior probability of having three or fewer stochastic trends is zero and the results suggest there is likely to be four. The RBC is driven by technology shocks which, in Fisher and KPSW, are stochastic trends. These are the only stochastic trends described in the economic model and KPSW assume technology is (in the three variable model) the only stochastic trends that enters the system. As the economic model implies there are only two common stochastic trends, there appear to be two extra stochastic trends. CC report evidence of an extra stochastic trend in a three variable system, but they then choose use the single trend model for inference. These extra trends could be entering from unobserved variables such as preference shocks or government expenditures. Fuentes-Albero, Kryshko, Ríos-Rull, Santaeulàlia-Llopis and Schorfheide (2009) (FKRSS) estimate a DSGE model with AR(1) process for these shocks. Although they use a prior that restricts them to the stationary region, they report results indicating these shocks have very high roots. A referee has suggested that this finding (of extra stochastic trends) could result from not modelling structural instability. We agree that this is a potential source and consider this a topic for future work.

The data are uninformative about the form of the cointegrating relations. That is, there is almost the same support for a process with no restrictions (almost 40%) as there is for the processes with restrictions (60%). The 60% probability mass on $o = 1$ and $o = 2$ suggests the assumptions that the price of an investment good is nonstationary and does not cointegrate with any other variable in the system, and that the Great Ratios enter the equilibrium relations, are not unreasonable. Since the posterior probability that $r > 1$ is zero, the evidence suggests that the Great Ratios are not themselves stationary. An important result is that the restrictions on the long run responses of investment price and productivity to the technology shocks (the identifying restrictions for these shocks) also have reasonable support.

Overall the evidence in the estimated probabilities for the features of the econometric model suggested by the RBC of Fisher (2006) are reasonable with the exception of the number of stochastic trends. But the evidence is not decisive and there remains considerable model uncertainty.

Table 1: Posterior probabilities, $P(M_i|y)$, of the top five models.

d	l	r	o	s	$P(M_i\|y)$	Cumulative probabilities
5	0	1	0	0	0.0951	0.0951
4	0	1	0	0	0.0911	0.1862
5	0	1	2	0	0.0858	0.2720
5	0	1	1	0	0.0855	0.3575
4	0	1	2	0	0.0833	0.4408
4	0	1	1	0	0.0772	0.5180
5	0	1	0	1	0.0731	0.5912
5	0	1	1	1	0.0724	0.6635
5	0	1	2	1	0.0697	0.7333
4	0	1	0	1	0.0691	0.8024
4	0	1	2	1	0.0687	0.8711
4	0	1	1	1	0.0646	0.9357
3	0	0	0	0	0.0643	1.0000

Table 2: Posterior probabilities of features for a real business cycle model.

$d=3$	$d=4$	$d=5$	$r=0$	$r=1$
0.0643	0.4540	0.4817	0.0643	0.9357
$o=0$	$o=1$	$o=2$	$s=0$	$s=1$
0.3927	0.2997	0.3076	0.5823	0.4177

A number of interesting outputs can be obtained from the models with $s=1$. Fisher (2006) chooses a value for λ in Section 2 to improve estimation efficiency. As our objective is to evaluate the empirical support for such assumptions, we do not impose a value for λ but rather estimate the posterior distribution. This parameter is estimated using the models where $s=1$ and the posterior density is presented in Figure 1 below. Although some 26.6% of draws fell outside the range $(0,1)$ - 23% of draws fell below zero and 3.6% fell above one - the assumption that $\lambda = 0.25$ seems perfectly reasonable. In fact, conditional on $\lambda \in [0,1]$, the posterior mean is 0.25. This estimate is useful as it permits recovery of neutral technology change if we assume the equality $W_t = A_t K_t^{\lambda} H_t^{1-\lambda}$ (FKRSS).

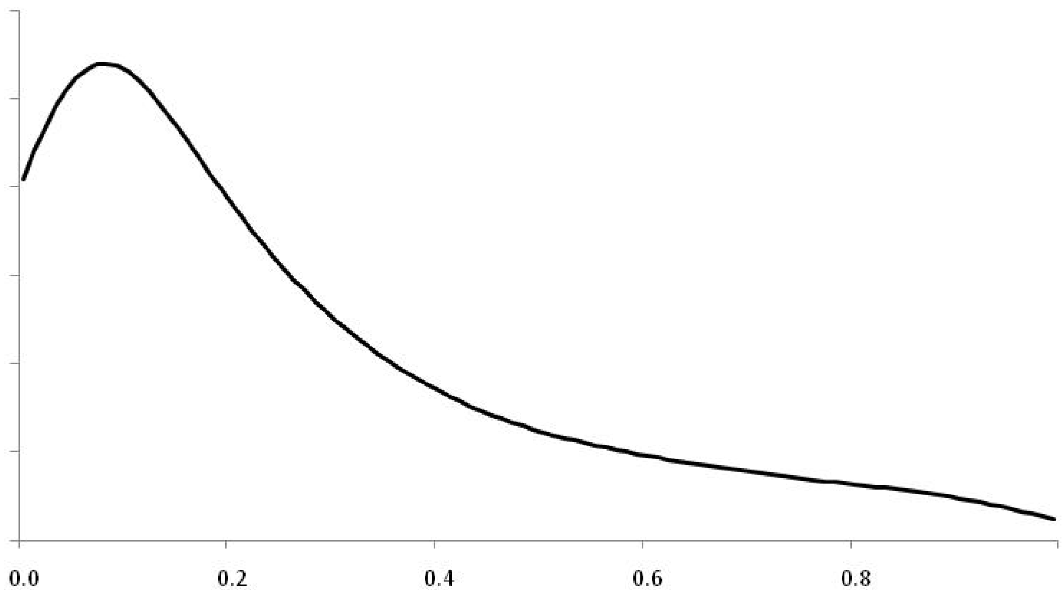

Figure 1: Posterior density of λ estimated from all models with $s = 1$.

An important area of interest in RBC models is the dynamics of w_t, c_t, and x_t, including role of the permanent and transitory shocks in the business cycle. By decomposing the variance into the components due to transitory and permanent shocks, it is possible to gain an impression of the relative importance of these effects for the variability of the consumption, investment and output. As the model set includes models with the same features $(d, l, r, o$ and $s)$ as those used in other studies, specifically King, Plosser, Stock and Watson (1991, hereafter KPSW) and Centoni and Cubadda (2003, hereafter CC), it is possible to compare results across studies. As the model in this paper has two additional variables (p_t and h_t), the results will differ from those in KPSW and CC unless (p_t, h_t) is strongly exogenous to (c_t, x_t, w_t). Therefore, the original results from KPSW and CC are also provided. KPSW and CC use output, w_t, whereas this paper uses productivity, $a_t = w_t - h_t$.

As a_t is a linear function of h_t which is also included in the model, the decomposition for w_t can be readily obtained from the estimation output.

KPSW derive an identification scheme for this decomposition based upon a single productivity shock entering these variables. This model is extended in Fisher to permit two types of permanent shocks, however in both cases the economic model implies that the Great Ratios ($c_t - w_t$ and $x_t - w_t$) will be stationary. As discussed above, results from this study suggest there is uncertainty associated with this aspect of the theory as the evidence suggest more than one stochastic trend entering (c_t, x_t, w_t). However, the equilibrium relations appear to be well described by linear combinations of the Great Ratios. Notwithstanding this ambiguity, it is not evident that the excess of stochastic trends affects estimates of other outputs such as permanent and transitory proportions of the variance over the business cycle .

KPSW estimate the proportion of variance due to transitory shocks in the time domain for the model with one stochastic trend, stationary Great Ratios and a linear deterministic trend with 8 lags of differences. For x_t and w_t they report proportions varying from 0.12 (c_t), 0.88 (x_t) and 0.55 (w_t) at one quarter after the shock to 0.11 (c_t), 0.53 (x_t) and 0.19 (w_t) respectively at 24 quarters after the shock. Our interest is in the proportion of business cycle fluctuations due to permanent shocks and so we follow Centoni and Cubadda (2003) (hereafter CC) who consider the variance decomposition within the frequency domain.

Figure 2 presents the posterior distribution of the transitory component of c_t, x_t and w_t constructed by averaging over all models. This plot shows significant mass at zero for all three variables suggesting a large role for permanent shocks. The exception is investment which has more mass towards one.

With their slightly shorter sample, CC found proportions of variability over an 8-32 quarter period of 0.43 for c_t, 0.86 for x_t and 0.82 for w_t. Table 3 reports the proportions of fluctuations over 8 to 32 quarters that are due to permanent shocks for the three variables using the updated data set and extended model set in this study. For these results the full model set is used, including models with $s = 0$.

Both the CC and KPSW models assign a smaller proportion of the variability in consumption, investment and income to the permanent shocks (productivity shocks in their models) than the other models. The remaining models generally agree with each other, at least in the relative sizes if not the exact values. We estimated using different subsets of the model set and found that adding less likely models to the model set tends to decrease the temporary component in investment. An important influence on the proportion is the imposition of the long run restrictions. The last row of Table 3 reports the estimates for all models with the restrictions on the long run responses ($s = 1$); that is, the restrictions that identify the two technology shocks. These models produce results closer to those of KPSW and CC suggesting that conditioning upon particular models or model sets has a noteable affect upon the results obtained. When model uncertainty is incorporated into the study, the support shifts against the conclusions of CC and KPSW that the permanent component is not an important determinant of business cycle fluctuations for consumption and output. Although the conclusion is consistent for investment.

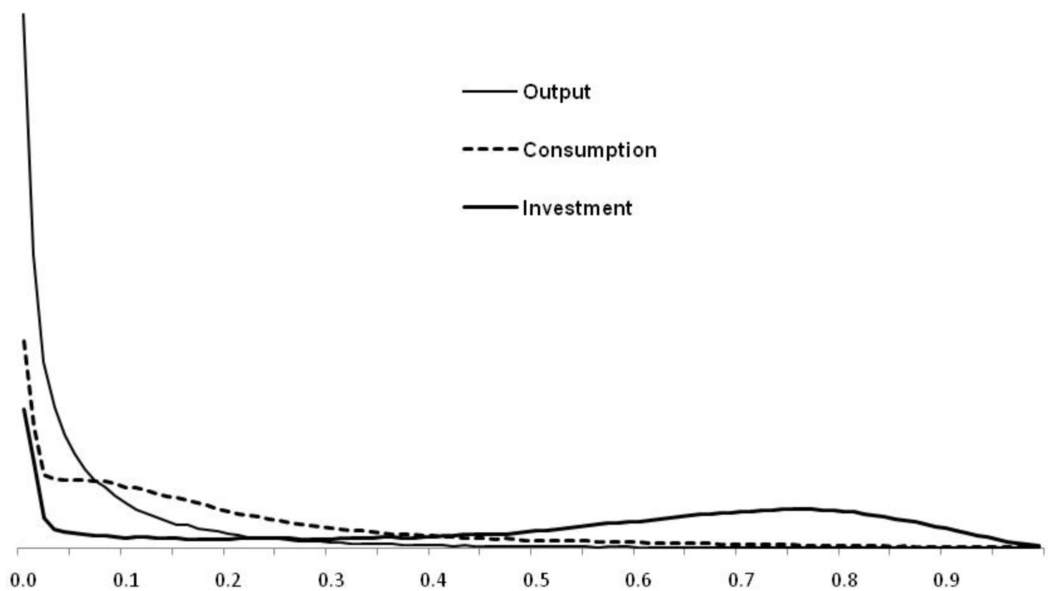

Figure 2: Posterior densities of the transitory component of output, consumption and investment.

Table 3: Estimated proportion of variance due to transitory components in the frequency domain.

Models used	c_t	x_t	w_t
CC model	0.870	0.928	0.439
KPSW model	0.870	0.923	0.440
Best model	0.135	0.693	0.052
Top 5 models	0.180	0.609	0.068
Top 10 models	0.199	0.575	0.075
All models	0.181	0.527	0.068
All models with $s = 1$	0.347	0.885	0.096

We conclude by discussing the responses in a_t, c_t, x_t, and the log Great Ratios $c_t - w_t$ and $x_t - w_t$ to technology shocks. These shocks are identified when the restrictions on long run responses $s = 1$, are imposed. In this study, the restrictions are over-identifying restrictions as a Cholesky decomposition of the covariance matrix has been used to identify structural shocks. Although our interest is in the responses to technology shocks, it is instructive to compare responses to shocks without the overidentifying restrictions. The contrast gives an impression of the importance of the restriction for achieving identification. Figures 3 and 4 show, respectively, the posterior densities, constructed from 12,000 of the parameters in each included model, of responses in productivity to the structural shocks. The response in Figure 3 is to the second structural shock in $e_t = \Omega^{-1/2}u_t$, while the response in Figure 4 is to a neutral technology shock, $\varepsilon_{a,t}$. The lines in these figures are the higher posterior density (HPD) regions for the densities.[14]

These figures clearly suggest that the just identified shock e_t is not a neutral technology shock and the restriction $s = 1$ has important implications for short run dynamics. There is about the same degree of uncertainty for the estimate of the two responses. Note that the scales on the y-axis differ but the mass in both figures is spread over about 5%. An interesting econometric result is that, although the response to e_t is always negative, some of the response paths to $\varepsilon_{a,t}$ appear to converge to zero although it is explicitly allowed to stay away from zero in the long run. As $\varepsilon_{a,t}$ is identified by a

[14] An HPD region is the smallest region in the support with a given probability mass. This implies the HPD regions are always unique (unlike credible intervals) and their borders (shown in the Figures 3 and 4) are contours of the densities. These are more informative that percentiles as they show up, for example, multimodality.

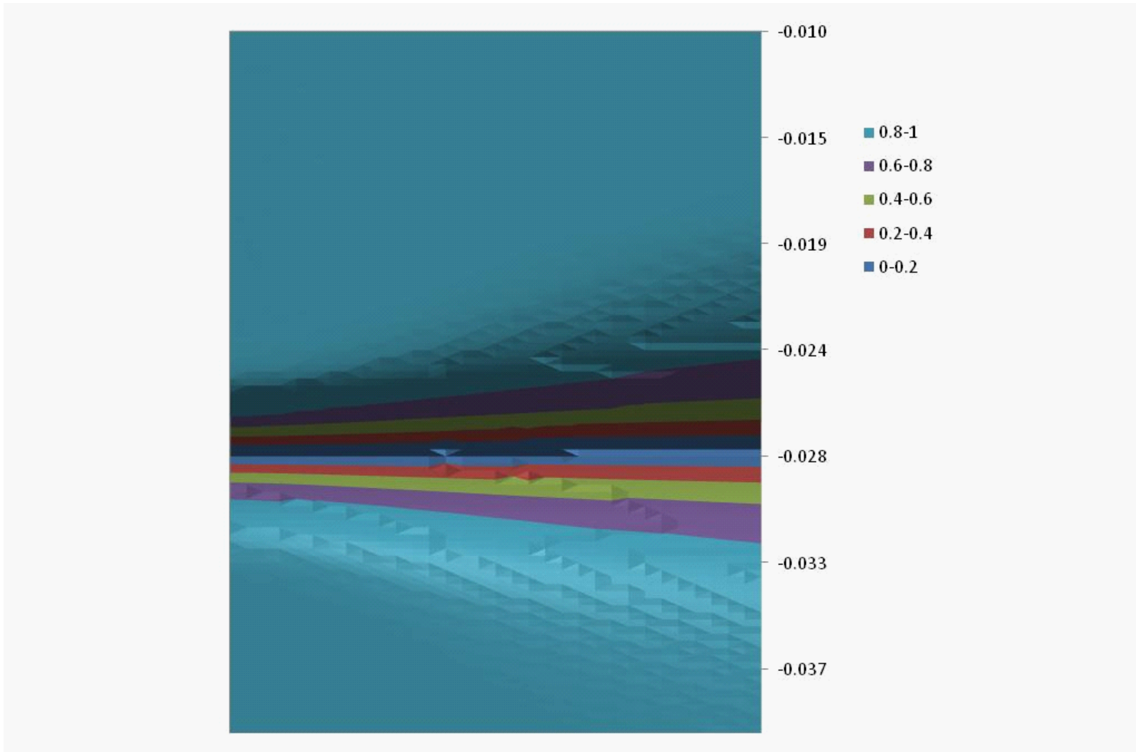

Figure 3: Highest posterior density regions for the response in $c_t - w_t$ to the second structural shock. The different lines denote contours of the posterior densities at each point on the shock horizon. The mass in the region with a higher density is indicated on the key. The impulses are over 34 quarters.

restriction on the model generating e_t, we might expect that as $\varepsilon_{a,t}$ converges to zero then so would e_t. One explanation for this apparent contrast is that the sampling process to obtain draws from the model subject to $s = 1$ does not draw from exactly the same region of the support as when $s = 0$. The Metropolis-Hastings sampler will draw from the region of the support with highest mass under $s = 1$, and this region may be near but need not be the same as for the unrestricted model.

The mean response paths of a_t, c_t, and x,are reported in Figures 5 and 6. Figure 5 shows the mean responses to a one standard deviation investment specific shock, $\varepsilon_{\nu,t}$. The shocks are not scaled as we regard the information on the average size of a shock to be important. The average size of the $\varepsilon_{\nu,t}$

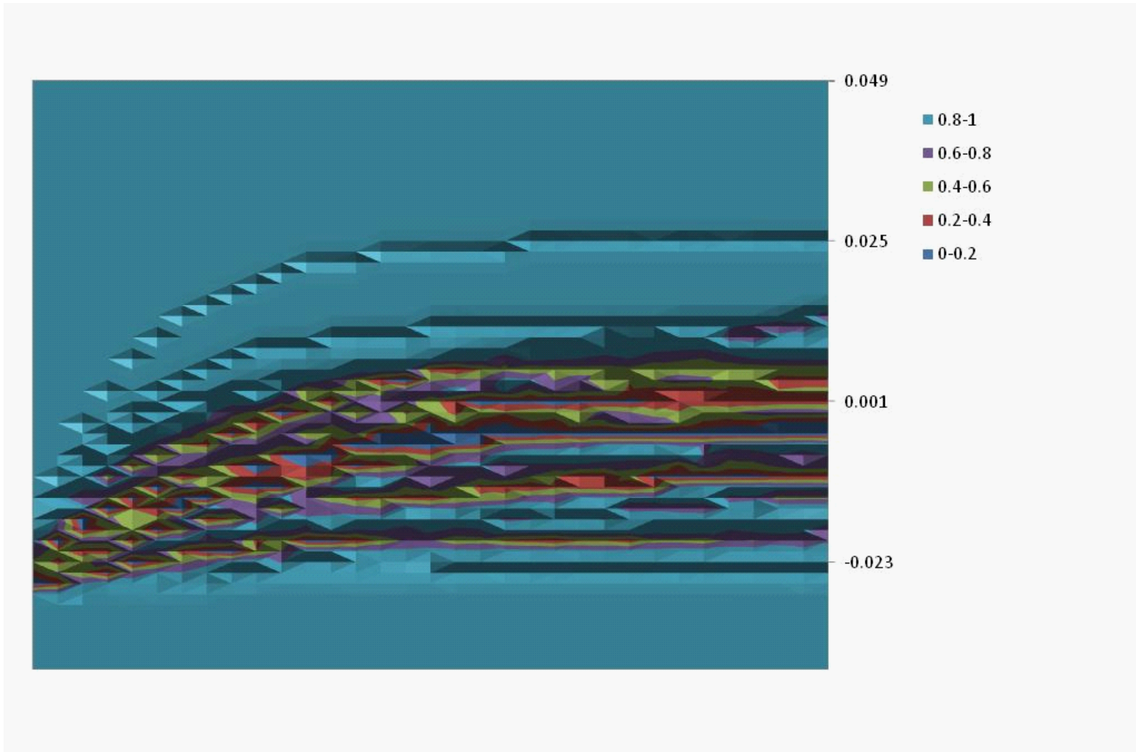

Figure 4: Highest posterior density regions for the response of $c_t - w_t$ to neutral technology shock. The different lines denote contours of the posterior densities at each point on the shock horizon. The mass in the region with a higher density is indicated on the key. The impulses are over 34 quarters.

responses in Figure 5 are much smaller than for the $\varepsilon_{a,t}$ shocks in Figure 6. The response paths of productivity to shocks in $\varepsilon_{\nu,t}$ and $\varepsilon_{a,t}$ are similar in form to the estimated paths in Fisher (2006), although the response to $\varepsilon_{a,t}$ in Figure 6 seems to converge more from the initial shock towards zero. The negative initial response in productivity to an investment specific shock described by Fisher in his simulation experiment does not appear in Figure 5, but these results match his post-1982 empirical estimates.

Fisher does not include consumption and investment in his econometric analysis and so does not report estimated responses for these variables, but he does produce simulated responses. In this sense, the results in this paper provide empirical results to supplement those of Fisher. The estimated responses in investment to the shock in $\varepsilon_{\nu,t}$ and in consumption to the shock in $\varepsilon_{a,t}$ do not conflict with the simulated responses in Fisher. However the responses in c_t to a shock in $\varepsilon_{\nu,t}$ stays negative unlike the simulated response of Fisher, and the response in x_t to the $\varepsilon_{a,t}$ shock is always of the wrong sign. The relative directions of these responses of c_t and x_t meet with Fisher's argument that investment specific shock directly only affects the production of investment goods making current consumption more expensive relative to future consumption. Thus we expect to see lower immediate consumption and higher immediate investment after an investment specific shock than after a neutral shock. These responses of c_t and x_t reported here are greater than we would expect based upon Fisher's simulation.

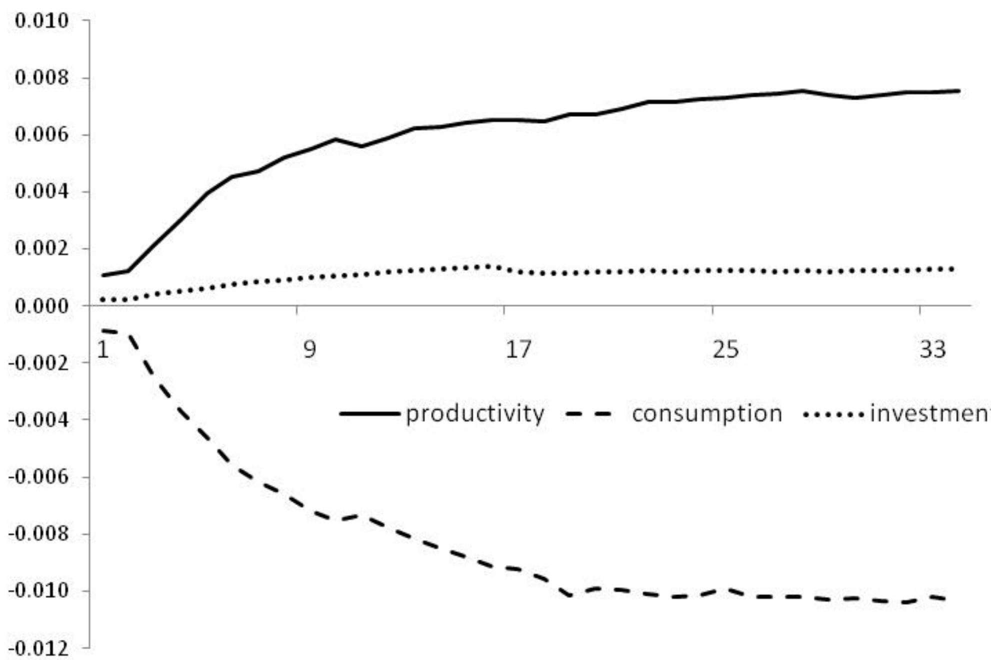

Figure 5: Mean impulse respones of a_t, c_t and x_t to an investment specific technology shock, $\varepsilon_{\nu,t}$.

In addition to considering how technology shocks affect consumption, income and investment, we are often interested in functions of these variables such as the savings rate or investment rate. We extend the results to produce response paths of the ratio of consumption to income and the ratio of invest-

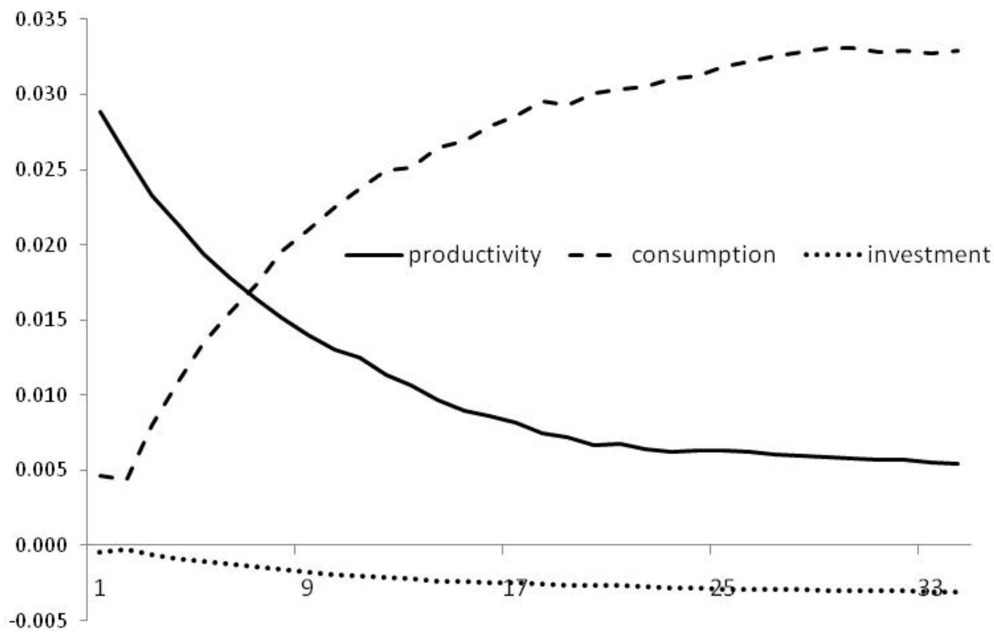

Figure 6: Mean impulse respones of a_t, c_t and x_t to a neutral technology shock, $\varepsilon_{a,t}$.

ment to income to the technology shocks. Figure 7 gives us an indication of how income is allocated to savings/consumption and investment in response to technology shocks. The response of consumption to a neutral technology shock is quite strong with an initial fall in the consumption ratio that slowly reverts back towards zero. Fisher does not give an indication of how much of output is consumed, but his simulation results suggest a much stronger initial response to a neutral shock by output than consumption suggesting the short term pattern reported here is compatible with his model. The responses of output and consumption to an investment specific technology shock are not so clear in Fisher, so it is difficult to judge how compatible is the small response in these results with his model.

The response for investment relative to output to the investment specific shock suggests a slight increase in the allocation to investment over time. However, the full distributions reported below indicate that this response cannot be distinguished from zero at all horizons. Assuming investment is currently less than half of total income, the response of the investment ratio

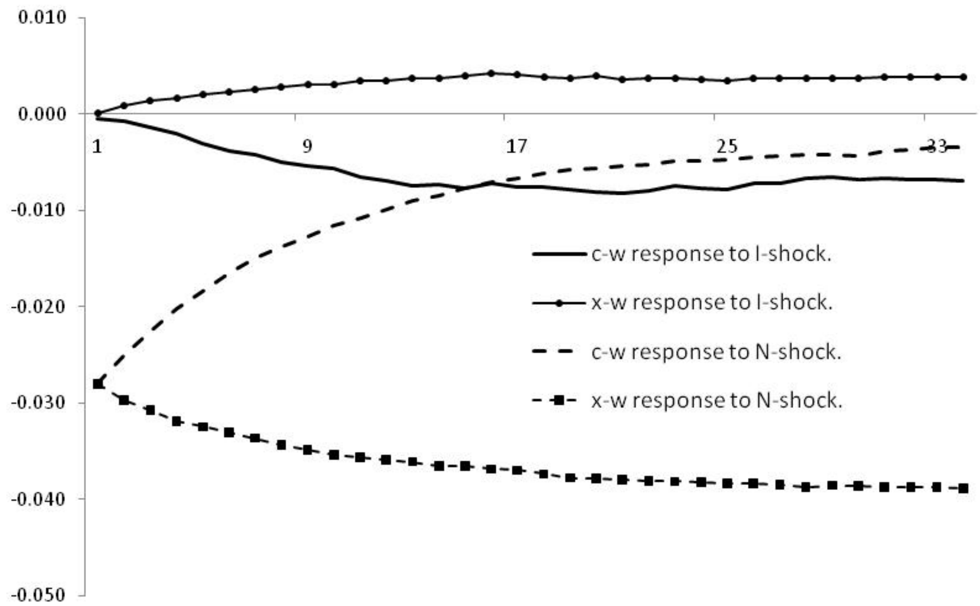

Figure 7: Mean impulse respones of $c_t - w_t$ and $x_t - w_t$ to neutral and investment specific technology shocks, $\varepsilon_{a,t}$ and $\varepsilon_{\nu,t}$.

to a neutral shock matches well with Fisher's simulation results as a larger absolute response by income than investment leads to a fall in the investment ratio. As the response in investment remains subdued (relative to output), the long run response tends to stay slightly negative. In summary, it appears that both a neutral or investment specific technology shock will lead to a notable initial decrease in the proportion of output devoted to consumption, while only neutral shocks lead to a fall in the allocation to investment.

The mean responses reported above do not fully inform us about the uncertainty associated with the responses. Figures 8 to 12 below and 4 above, present the HPD regions for responses in c_t and x_t to both shocks and the responses of $c_t - w_t$ and $x_t - w_t$ to a neutral technology shock. What is immediately obvious is the amount of uncertainty about the location of the response paths. This uncertainty is in part due to the size of the model set or diversity of the models in the model set and this diversity results in what appear to be multiple paths. The number of modes that appear in the densities, the degree of skew and kurtosis suggest that decision making processes that rely upon unimodal or symmetric distributions will give very unreliable answers.

The distribution of the response of c_t to a $\varepsilon_{\nu,t}$ shock becomes very disperse very quickly. Although the mean in Figure 5 decreases monotonically, the proportion of the density above zero fluctuates steadily around 25% reflecting the increasing dispersion and it is clear that considerable mass remains around 0.4%. Figure 9 shows the density for the response of x_t to a shock to $\varepsilon_{\nu,t}$. The density tends to remain around its mode at 0.13%, but has a small amount of mass above 0.5% and a secondary mode rising to about 0.28% after 34 lags. Although the mean in Figure 5 remains low, there is reasonable chance (around 5%) of seeing significantly higher responses. That the mean lies between the two modes at 34 lags, in a low probability area, raises a question about its usefulness as a measure of location.

The pattern in Figure 10 is very similar to the simulated response of consumption to a neutral technology shock produced by Fisher. However, again there is significant uncertainty surrounding the estimate. Figures 11 and 12 largely confirm the earlier results for these shocks but we include them to demonstrate the range of patterns and degrees of dispersion that result when model uncertainty is taken into account.

6 Conclusion.

This paper presents a Bayesian approach to investigating the support for an economic model by considering the empirical support for the features that model implies for a reduced form econometric model. The economic model is the a real business cycle model of Fisher (2006), with reference to other papers that use this class of model such as Greenwood, Hercowitz, and Krusell (1997) and KPSW. An important component of this model is the restrictions of long run responses that are used to identify investment specific and neutral technology shocks. For many of the important features implied by this model we find reasonable support and some, such as the long run identifying restrictions, receive quite strong support. Further, the impulse responses demonstrate that the predictions of the model are quite plausible. The only feature that is strongly rejected is the assumption that the technology shocks are the only sources of instability as stochastic trends.

The methodology is an important contribution of this paper. The ap-

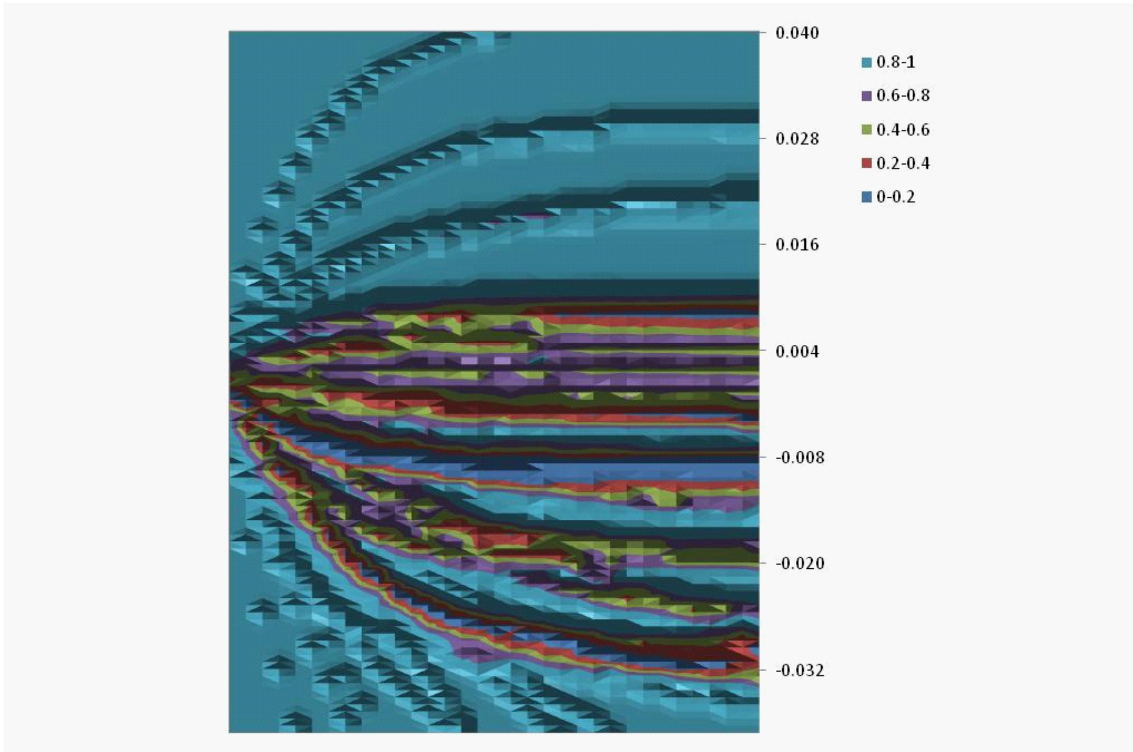

Figure 8: HPD regions for impulse responses of consumption to an investment specific technology shock. See the discussion on Figures 3 and 4 for the interpretation of the lines in this graph.

proach results in unconditional inference on these features of the vector autoregressive model as the effect of any one model on the inference has been averaged out, and so model uncertainty is incorporated into the analysis. Techniques are developed 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.

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 (2008) develop methods

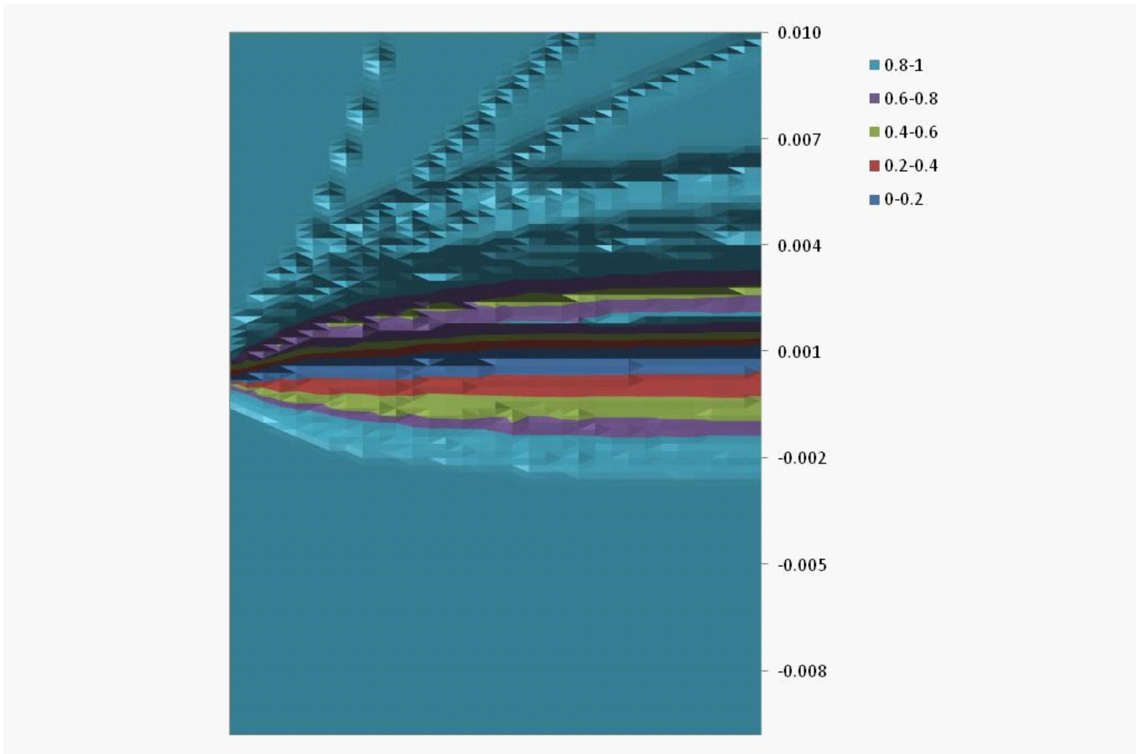

Figure 9: HPD regions for impulse responses of investment to an investment specific technology shock.

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.

We end with mentioning two topics for further research. First, although our mixing over priors partially addresses this issue, there remains the issue of the robustness of the results with respect to wider prior and model specifications. Very natural extensions of the approach in this paper are to consider forms of nonlinearity and time variation in the model itself as Cogley and Sargent (2001, 2005) and Primiceri (2005) do for the VAR. For instance, in using a SVAR for business cycle analysis one may use prior information on

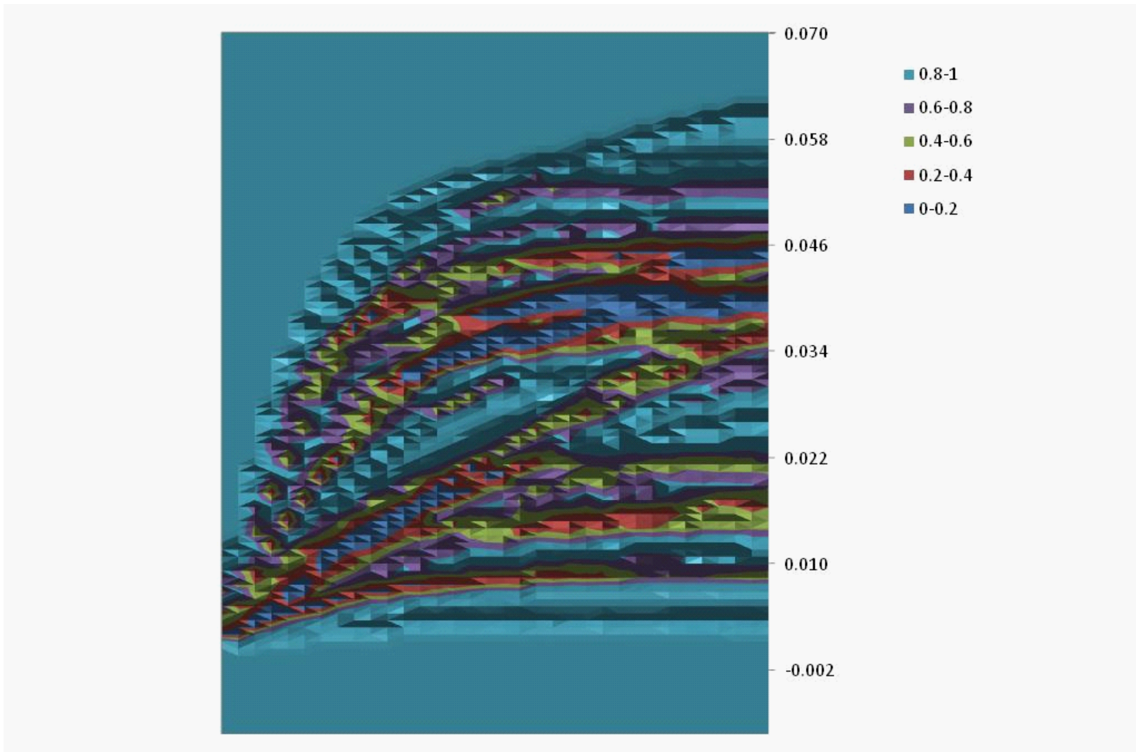

Figure 10: HPD regions for impulse responses of consumption to a neutral technology shock.

the length and amplitude of the period of oscillation (see Harvey, Trimbur and van Dijk (2007)). An example of a possible nonlinear time varying model 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 we are grateful to three referees and Frank Schorfheide for extensive comments

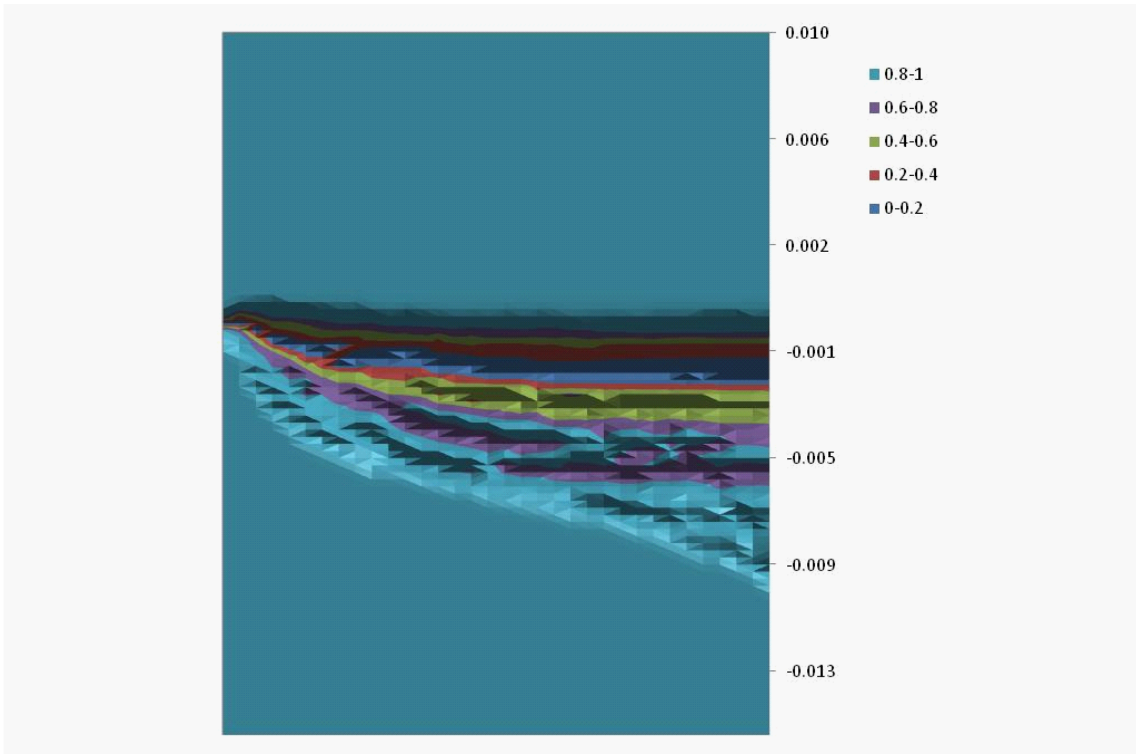

Figure 11: HPD regions for impulse responses of investment to a neutral technology shock.

which led to a substantial revision and extension of the results presented in the original working paper. We also 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 Australian National University and van Dijk acknowledges financial support from the Netherlands Organization of Scientific Research.

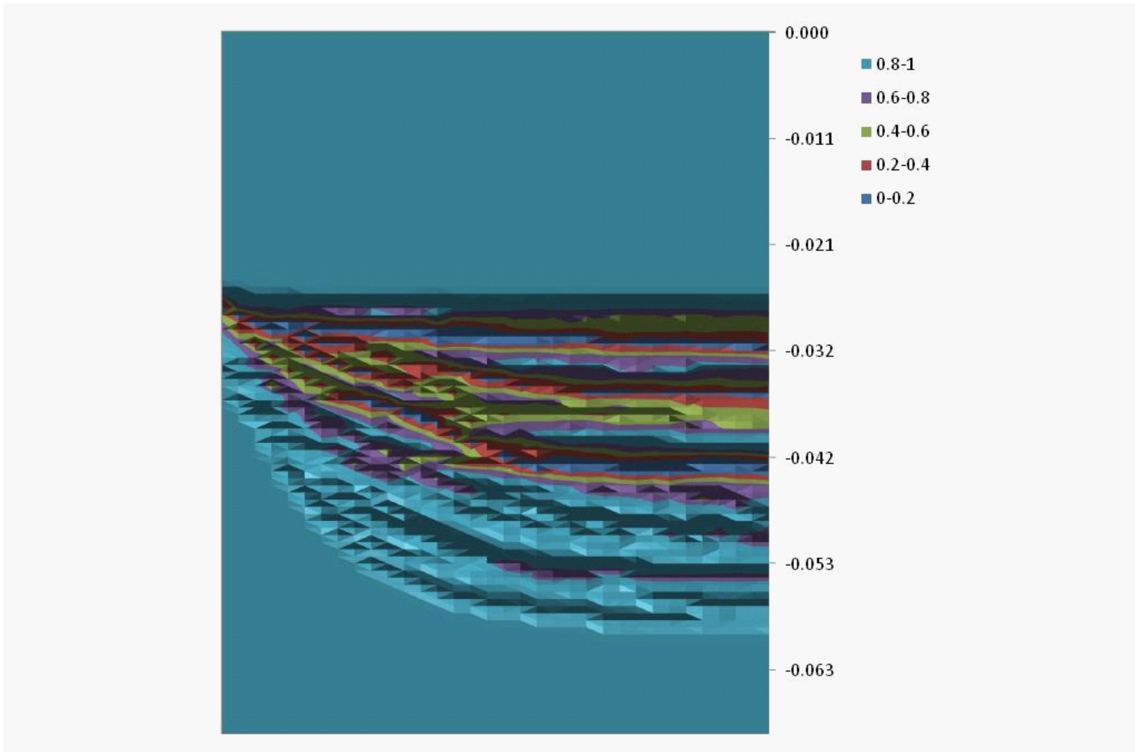

Figure 12: HPD regions for impulse responses of $x_t - w_t$ to a neutral technology shock.

7 References.

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

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.

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

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.

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.

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.

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

Fisher, Jonas D. M. (2006): "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks," *Journal of Political Economy* **114**, 413-451.

Fisher, Jonas D. M. (2005): "The Dyanamic Effects of Neutral and Investment-Specific Technology Shocks," Manuscript, Fed. Reserve Bank Chicago.

Fuentes-Albero C., M. Kryshko, J-V. Ríos-Rull, R. Santaeulàlia-Llopis and F. Schorfheide (2009): "Methods versus Substance: Measuring the Effects of Technology Shocks on Hours," NBER Working Paper 15375.

Geweke, J. (1996): "Bayesian reduced rank regression in econometrics," *Journal of Econometrics* **75**, 121-146.

Greenwood, J., Z. Hercowitz, and P. Krusell (1997): "Long-Run Implications of Investment-Specific Technological Change," *The American Economic Review* **87**, 342-62.

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.

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

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

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

King, R., C.I. Plosser, and S. Rebelo (1988): "Production, growth, and business cycles I: The basic neoclassical model," *Journal of Monetary Economics* **21**, 195-232.

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.

Klugkist, I., Hoijtink, H. (2007): "The Bayes factor for inequality and about equality constrained models," *Computational Statistics and Data Analysis* **51**, 6367–6379.

Koop, G. (2003): *Bayesian Econometrics* (Wiley, Chichester).

Koop G., R. Léon-Gonzalez and R. W. Strachan (2010): "Efficient posterior simulation for cointegrated models with priors on the cointegration space" *Econometric Reviews* **29**, Issue 2, 224-242.

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

Koop G., R., S. Potter (2003): "Bayesian Analysis of Endogenous Delay Threshold Models," *Journal of Business and Economic Statistics* **21**, 93-103.

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.

Kydland, F. and E. Prescott (1982): "Time to build and aggregate fluctuations," *Econometrica* **50**, 173-208.

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

Litterman, R. B. (1980): "A Bayesian procedure for forecasting with vector autoregressions," Massachusetts Institute of Technology, Mimeo.

Litterman, R. B. (1986): "Forecasting with Bayesian vector autoregressions — Five years of experience," *Journal of Business and Economic Statistics* **4**, 25–38.

Lopes, H. F., and M. West (2004): "Bayesian Assessment in factor analysis," *Statistica Sinica* **14**, 41-67.

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

Magnus, J.R. and H. Neudecker. (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

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.

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.

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.

Ravazzolo, F., H.K. van Dijk, & M.J.C.M. Verbeek (2007): "Predictive gains from forecast combinations using time-varying model weights," Econometric Institute Report EI 2007-26, Erasmus University Rotterdam.

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.

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.

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. (2001): "Bayesian prediction with cointegrated vector autoregressions," *International Journal of Forecasting* **17**, 585-605.

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

Wright, Jonathon H. (2008): "Bayesian Model Averaging and exchange rate forecasts," *Journal of Econometrics* **146**, 329-341.

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

8 Appendix I: Impropriety of priors defined on B almost everywhere.

This appendix provides a proof that that at any point $R'\alpha = a$ for fixed R and a, the posterior for B, for a wide class of priors, is improper. This result is demonstrated for the restriction $\alpha_2 = 0$ where α_2 is the last $n_2 \leq n - r$ rows of $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$. The claim is for a more general restriction than this, $R'\alpha = a$, but the proof carries through since we can rewrite α as

$$\begin{aligned} \alpha &= R\left(R'R\right)^{-1} R'\alpha + R_\perp \left(R_\perp' R_\perp\right)^{-1} R_\perp'\alpha \\ &= Ra_1 + R_\perp a_0 \\ &= \begin{bmatrix} R & R_\perp \end{bmatrix} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \overline{R}\overline{a} \end{aligned}$$

where $\overline{R} = [R \quad R_\perp]$ and $\overline{a} = \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$. $R'\alpha = a$ implies $a_1 = a$ and this result shows that a proof for $\alpha_2 = 0$ will imply the same results for $a_0 = 0$.

To focus the proof on the relevant features of the models, consider the model for the $n \times 1$ vector y_t given by

$$\begin{aligned} \Delta y_t &= \alpha\beta' y_{t-1} + u_t \text{ where } u_t \sim iidN\left(0, \Sigma\right) \\ \text{and } t &= 1, ..., T. \end{aligned}$$

Collecting the terms into matrices and transposing gives

$$Y = X\beta\alpha' + u$$

where the t^{th} rows of Y, X and u are $\Delta y_t'$, y_{t-1}' and u_t' respectively. The above assumptions give the likelihood as

$$L\left(Y|\beta, \alpha, \Sigma\right) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}u'u\right\}.$$

The proof is built upon the model specification of Villani (2005) as that paper is probably the most important in that it has generated much applied work and made an important contribution to the theory of Bayesian cointegration analysis. The form of the posterior in Villani (2005) is also very general as it can be used to represent posteriors from the flat prior and inform on results we will obtain using convex priors such as that in Kleibergen and Paap (2002). The prior of Villani (2005) is given by

$$p(\alpha,\beta,\Sigma) \propto |\Sigma|^{-(n+r+q+1)/2} \exp\left\{-\frac{1}{2}tr\Sigma^{-1}\left(A+\nu\alpha\beta'\beta\alpha'\right)\right\}$$

where A is a known full rank $n\times n$ matrix, and $q,\nu>0$. Combining the likelihood and prior above gives the posterior

$$p(\alpha,\beta,\Sigma|y,r) \propto |\Sigma|^{-(T+n+r+q+1)/2} \exp\left\{-\frac{1}{2}tr\Sigma^{-1}\left(u'u+A+\nu\alpha\beta'\beta\alpha'\right)\right\}.$$

Integrating $p(\alpha,\beta,\Sigma|y,r)$ with respect to Σ results in

$$p(\alpha,\beta|y,r) \propto |u'u+A+\nu\alpha\beta'\beta\alpha'|^{-(T+r+q)/2}.$$

At this point, it is necessary to rewrite the posterior in terms of α and then α_1 and α_2. To do this, use the following result

$$\begin{aligned} u'u+A+\nu\alpha\beta'\beta\alpha' &= \left(Y-X\beta\alpha'\right)'\left(Y-X\beta\alpha'\right)+A+\nu\alpha\beta'\beta\alpha' \\ &= S+\left(\alpha-\widehat{\alpha}\right)V^{-1}\left(\alpha-\widehat{\alpha}\right)' \end{aligned}$$

where $S=Y'Y+A-\widehat{\alpha}V^{-1}\widehat{\alpha}'$, $V=\left(\beta'M\beta\right)^{-1}$, $M=X'X+\nu I_r$, and $\widehat{\alpha}=Y'X\beta V$. This expression gives the posterior in terms of α, which can be rewriten in terms of α_1 and α_2 as

$$\begin{aligned} p(\alpha,\beta|y,r) &\propto \left|S+\left(\alpha-\widehat{\alpha}\right)V^{-1}\left(\alpha-\widehat{\alpha}\right)'\right|^{-(T+r+q)/2} \\ &= \left|V+\left(\alpha_2-\widehat{\alpha}_2\right)'S_{22}^{-1}\left(\alpha_2-\widehat{\alpha}_2\right)+\left(\alpha_1-\widehat{\alpha}_{1.2}\right)'S_{11.2}^{-1}\left(\alpha_1-\widehat{\alpha}_{1.2}\right)\right|^{-(T+r+q)/2} \end{aligned}$$

where $\widehat{\alpha}_{1.2}'=\widehat{\alpha}_1+\left(\alpha_2-\widehat{\alpha}_2\right)'S_{22}^{-1}S_{21}$, $\widehat{\alpha}_2=Y_2'X\beta V$ where Y_2 is the $T\times n_2$ made up of the last n_2 columns of Y,

$$S=\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

and $S_{11.2}=S_{11}-S_{12}S_{22}^{-1}S_{21}$. The marginal distribution of α_2 and β is then

$$p(\alpha_2,\beta|y,r) \propto \left|V+\left(\alpha_2-\widehat{\alpha}_2\right)'S_{22}^{-1}\left(\alpha_2-\widehat{\alpha}_2\right)\right|^{-(T+r+q-n_1)/2}\left|S_{11.2}\right|^{-n_1/2}$$

Next, impose the restriction $\alpha_2 = 0$ to obtain the marginal distribution for β as

$$\begin{aligned}
p(\alpha_2, \beta | y, r) &\propto \left| V + \widehat{\alpha}_2' S_{22}^{-1} \widehat{\alpha}_2 \right|^{-(T+r+q-n_1)/2} |S_{11.2}|^{-n_1/2} \\
&= \left| (\beta' M \beta)^{-1} + (\beta' M \beta)^{-1} \beta' X' Y_2 S_{22}^{-1} Y_2' X \beta (\beta' M \beta)^{-1} \right|^{-(T+r+q-n_1)/2} \\
&\quad \times |S|^{-n_1/2} |S_{22}|^{n_1/2} \\
&= |\beta' M \beta|^{(T+r+q-n_1)/2} \left| \beta' M \beta + \beta' X' Y_2 S_{22}^{-1} Y_2' X \beta \right|^{-(T+r+q-n_1)/2} \\
&\quad \times |S|^{-n_1/2} |S_{22}|^{n_1/2}.
\end{aligned}$$

At this point the aim is to obtain the posterior for β in terms of quadratic forms of β which will give the density for a $l, m-$ poly-t distribution. Then using the results of Drèze (1977) for these distributions, it is possible to determine some of the properties of the distribution, such as whether it exists.

Use the following results to derive the form of the conditional density $p(\beta | \alpha_2, y, r)$:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} Y_1' Y_1 & Y_1' Y_2 \\ Y_2' Y_1 & Y_2' Y_2 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} - \begin{bmatrix} \widehat{\alpha}_1 V^{-1} \widehat{\alpha}_1' & \widehat{\alpha}_1 V^{-1} \widehat{\alpha}_2' \\ \widehat{\alpha}_2 V^{-1} \widehat{\alpha}_1' & \widehat{\alpha}_2 V^{-1} \widehat{\alpha}_2' \end{bmatrix}$$

such that $S_{22} = Y_2' Y_2 + A_{22} - Y_2' X \beta (\beta' M \beta)^{-1} \beta' X' Y_2$ and $S = Y' Y + A - Y' X \beta (\beta' M \beta)^{-1} \beta' X' Y$. Therefore, it is now possible to rewrite the determinants $|S|$ and $|S_{22}|$ as

$$\begin{aligned}
|S| &= |\beta' M_1 \beta| \, |\beta' M \beta|^{-1} |Y' Y + A| \text{ and} \\
|S_{22}| &= |\beta' M_2 \beta| \, |\beta' M \beta|^{-1} |Y_2' Y_2 + A_{22}|
\end{aligned}$$

where $M_1 = M - X' Y (Y' Y + A)^{-1} Y' X$ and $M_2 = M - X' Y_2 (Y_2' Y_2 + A_{22})^{-1} Y_2' X$. The expression $\left| \beta' M \beta + \beta' X' Y_2 S_{22}^{-1} Y_2' X \beta \right|$ can be shown to have a quadratic form by using the matrix

$$\begin{bmatrix} \beta & 0 \\ 0 & I_n \end{bmatrix}$$

to show

$$\begin{aligned}
\left| \beta' M \beta + \beta' X' Y_2 S_{22}^{-1} Y_2' X \beta \right| &= \left| \begin{bmatrix} \beta' & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} M + X' Y_2 S_{22}^{-1} Y_2' X & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \beta & 0 \\ 0 & I_n \end{bmatrix} \right| \\
&= \left| \begin{bmatrix} \beta' & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} \beta & 0 \\ 0 & I_n \end{bmatrix} \right| \left| \begin{bmatrix} M + X' Y_2 S_{22}^{-1} Y_2' X & 0 \\ 0 & I_n \end{bmatrix} \right| \\
&= |\beta' \beta| \left| M + X' Y_2 S_{22}^{-1} Y_2' X \right|.
\end{aligned}$$

Rewrite $\left|M + X'Y_2 S_{22}^{-1} Y_2' X\right|$ as a quadratic form of β as

$$\begin{aligned}
\left|M + X'Y_2 S_{22}^{-1} Y_2' X\right| &= \left|S_{22} - Y_2' X M^{-1} X' Y_2\right| |M| \left|S_{22}\right|^{-1} \\
&= \left|Y_2'Y_2 + A_{22} - Y_2' X M^{-1} X' Y_2 - Y_2' X \beta \left(\beta' M \beta\right)^{-1} \beta' X' Y_2\right| |M| \left|S_{22}\right|^{-1} \\
&= \left|\beta' M_3 \beta\right| \left|\beta' M \beta\right|^{-1} \left|\beta' M_2 \beta\right|^{-1} \left|\beta' M \beta\right| \\
&\quad \times \left|Y_2'Y_2 + A_{22} - Y_2' X M^{-1} X' Y_2\right| |M| \left|Y_2'Y_2 + A_{22}\right|^{-1} \\
&\propto \left|\beta' M_3 \beta\right| \left|\beta' M_2 \beta\right|^{-1}
\end{aligned}$$

where $M_3 = M - X'Y_2 \left(Y_2'Y_2 + A_{22} - Y_2' X M^{-1} X' Y_2\right)^{-1} Y_2' X$, $S_{11} - S_{12} S_{22}^{-1} S_{21} = Y_2'Y_2 + A_{22} - Y_2' X \beta \left(\beta' M \beta\right)^{-1} \beta' X' Y$.

$$\begin{aligned}
p\left(\beta | \alpha_2 = 0, y, r\right) &\propto \left|\beta' M \beta\right|^{(T+r+q-n_1)/2} \\
&\quad \left|\beta' M \beta + \beta' X' Y_2 S_{22}^{-1} Y_2' X \beta\right|^{-(T+r+q-n_1)/2} \\
&\quad |S|^{-n_1/2} \\
&\quad \left|S_{22}\right|^{n_1/2} \\
&\propto \left|\beta' M \beta\right|^{(T+r+q-n_1)/2} \\
&\quad \left|\beta' \beta\right|^{-(T+r+q-n_1)/2} \left|\beta' M_3 \beta\right|^{-(T+r+q-n_1)/2} \left|\beta' M_2 \beta\right|^{(T+r+q-n_1)/2} \\
&\quad \left|\beta' M_1 \beta\right|^{-n_1/2} \left|\beta' M \beta\right|^{n_1/2} \\
&\quad \left|\beta' M_2 \beta\right|^{n_1/2} \left|\beta' M \beta\right|^{-n_1/2} \\
&\propto \left|\beta' \beta\right|^{-(T+r+q-n_1)/2} \left|\beta' M \beta\right|^{(T+r+q-n_1)/2} \\
&\quad \times \left|\beta' M_1 \beta\right|^{-n_1/2} \left|\beta' M_2 \beta\right|^{(T+r+q)/2} \left|\beta' M_3 \beta\right|^{-(T+r+q-n_1)/2}.
\end{aligned}$$

We assume that all of the M_j are positive definite. The results of Drèze (1977) are expressed for vector poly-t variables not matrix form. Using standard approaches we can show they can be used to learn something of the distribution above. Each of the determinants above can be written as quadratic forms in B such as

$$\begin{aligned}
\beta' M_0 \beta &= \begin{bmatrix} I_r & B' \end{bmatrix} \begin{bmatrix} M_{0,11} & M_{0,12} \\ M_{0,21} & M_{0,22} \end{bmatrix} \begin{bmatrix} I_r \\ B \end{bmatrix} \\
&= M_{0,11} + B' M_{0,21} + M_{0,12} B + B' M_{0,22} B \\
&= M_{0,11} + B' M_{0,22} M_{0,22}^{-1} M_{0,21} + M_{0,12} M_{0,22}^{-1} M_{0,22} B + B' M_{0,22} B \\
&= M_{0,11} - B' M_{0,22} \widehat{B} - \widehat{B}' M_{0,22} B + B' M_{0,22} B + \widehat{B}' M_{0,22} \widehat{B} - \widehat{B}' M_{0,22} \widehat{B} \\
&= M_{0,11} - \widehat{B}' M_{0,22} \widehat{B} + \left(B - \widehat{B}\right)' M_{0,22} \left(B - \widehat{B}\right).
\end{aligned}$$

Denote by $p(b_1|B_2, y)$ the conditional distribution of b_1, the $(n-r)\times 1$ vector formed from the first column of B, conditional upon the remaining columns, B_2, is a 3-2 poly-t with the same exponents as the joint distribution above for all elements of B. That is,

$$p(\beta|\alpha_2 = 0, y, r) = p(B|y) = p(b_1|B_2, y)\, p(B_2|y).$$

The results of Zellner (1971) show that the conditional distribution of b_1, $p(b_1|B_2, y)$, is a 3-2 poly-t with the same exponents as the joint distribution above for all elements of B, $p(B|y)$. To obtain the marginal distribution for B_2 we integrate $p(B|y)$ with respect to b_1 :

$$p(B_2|y) = \int p(B|y)\, db_1 = \int p(b_1|B_2, y)\, db_1 p(B_2|y)$$

If the conditional for b_1 is not proper, the whole joint distribution for B is not proper. From Drèze (1977), the condition for propriety is

$$0 < -(T + r + q - n_1)/2 + (T + r + q - n_1)/2 + (T + r + q - n_1)/2 - (T + r + q - n_1)/2 - (n - r).$$

Since this sum is $-(n-r) < 0$, this shows that the posterior is not proper.

In the above discussion we have shown that the integral

$$\int p(\alpha_1, B|\alpha_2 = 0, y, r)\, d(\alpha_1, B) \tag{6}$$

diverges implying the posterior is improper. The prior of Kleibergen and Paap (2002) implies a form similar to the joint posterior for (α_1, B) above multiplied by the convex function

$$|I_r + B'B|^{(n-r)/2} |\alpha_1'\alpha_1|^{(n-r)/2}.$$

That is, the posterior has the form

$$p(\alpha_1, B|\alpha_2 = 0, y, r)\, |I_r + B'B|^{(n-r)/2} |\alpha_1'\alpha_1|^{(n-r)/2}\, d(\alpha_1, B)$$

Clearly if (6) is a divergent integral, then as the expression above involves convex functions of B and α_1, the integral of this density will also diverge.

9 Appendix II: Noninvariance of priors on the cointegrating space.

In this appendix we provide a formal proof that none of the proper priors on cointegrating spaces that currently exist in the literature (e.g., Geweke (1996), Kleibergen and Paap (2002), Strachan and Inder (2004) and Villani (2005)) are invariant to rescaling. We first show this for the prior of Villani (2005) and then use this result to provide the foundation for the proof that the priors of Kleibergen and Paap (2002) and Geweke (1996) are not invariant.

There do exist priors that are invariant and these are presented in Kleibergen and van Dijk (1994) and Strachan (2003). However these are both data dependent and so not actually priors. The uniform prior on B implies an invariant prior on $\mathfrak{p} = sp(\beta)$, but this prior is difficult to justify and has been largely discounted because of the large number of computational and theoretical issues that result from its use. As discussed in Strachan and Inder (2004), this prior implies a very informative and unusual prior on the cointegrating space as it assigns mass away from the normalization chosen by the economist; effectively the assumption made by the economist to justify the normalization is made a priori impossible. The priors of Kleibergen and van Dijk (1994) and Kleibergen and Paap (2002) share this feature, although the latter is not invariant to scale. Geweke (1996) presents a prior that sensibly weights the cointegrating space given an economists choice of normalization, but again this prior is not invariant to scale.

Before we turn to the priors that are not invariant, we discuss the source of the invariance and how it manifests in various priors. The lack of invariance of priors results from the transformation from a spherically symmetric distribution to an elliptically symmetric distribution. To demonstrate this effect, consider rescaling the vector y_t by a $n \times n$ diagonal matrix Λ in which the diagonal elements are not all equal[15]. Let the model for y_t be

$$\Delta y_t = \alpha\beta' y_{t-1} + u_t \tag{7}$$

and after rescaling

$$\begin{aligned} \Lambda \Delta y_t &= \Lambda\alpha\beta'\Lambda^{-1}\Lambda y_{t-1} + \Lambda u_t & (8) \\ \Delta \widetilde{y}_t &= \widetilde{\alpha}\widetilde{\beta}'\widetilde{y}_{t-1} + \widetilde{u}_t & (9) \end{aligned}$$

[15] Rescaling by a common value causes no problems.

where $\widetilde{y}_t = \Lambda y_t$, $\widetilde{\alpha} = \Lambda\alpha$, $\widetilde{\beta} = \Lambda^{-1}\beta$ and $\widetilde{u}_t = \Lambda u_t$. Further, note that rescaling implies the transformation of the covariance matrix $\widetilde{\Sigma} = \Lambda\Sigma\Lambda$. For simplicity we assume the first r variables are not rescaled such that

$$\Lambda = \begin{bmatrix} I_r & 0 \\ 0 & \Lambda_1 \end{bmatrix}.$$

Priors on vectors identified using the linear restrictions $\beta' = [I_r \quad B']$ will usually contain a term such as $|I_r + B'B|^{\upsilon} dB$, where ν may be positive or negative. The term $|I_r + B'B|^{\upsilon} dB$ applies the same mass in all directions of the space of β, $\mathfrak{p}$. In Villani (2005) $\nu = -n/2$ and in Kleibergen and Paap (2002) $\upsilon = (n-r)/2$. After rescaling, the term $|I_r + B'B|^{\upsilon} dB$ will become $\left|I_r + \widetilde{B}'\Lambda_1^2\widetilde{B}\right|^{\upsilon} |\Lambda_1|^r d\widetilde{B}$ (see appendix) and it is the Λ_1^2 that implies that the distribution of the $\widetilde{B}$ are no longer spherically symmetric and the distribution now favors particular directions of $\mathfrak{p}$. Therefore, the implied distribution on the cointegrating space has changed. If the prior were uniform prior to rescaling, after rescaling it no longer is, and if it were not uniform prior to rescaling then the implied distribution on the cointegrating space has changed after rescaling.

The full prior for (α, β, Σ) in equation (3.1) of Villani (2005) for the model in (7) is

$$p(\alpha, \beta, \Sigma)\, d(\alpha, \beta, \Sigma) = c_r |\Sigma|^{-(n+r+1)/2} \exp\left\{-\frac{1}{2} tr\Sigma^{-1}(A + \upsilon\alpha\beta'\beta\alpha')\right\} d(\alpha, \beta, \Sigma)$$

where c_r is an integrating constant. Next, consider the rescaling transformation in (8) and (9). After the transformation, we have the prior for $\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right)$

$$\begin{aligned} p\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right) &= c_r \left|\Lambda^{-1}\widetilde{\Sigma}\Lambda^{-1}\right|^{-(n+r+1)/2} \\ &\quad \times \exp\left\{-\frac{1}{2} tr\Lambda\widetilde{\Sigma}^{-1}\Lambda\left(\Lambda^{-1}\widetilde{A}\Lambda^{-1} + \upsilon\Lambda^{-1}\widetilde{\alpha}\widetilde{\beta}'\Lambda^2\widetilde{\beta}\widetilde{\alpha}'\Lambda^{-1}\right)\right\} |\Lambda|^c d\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right) \\ &= \widetilde{c}_r \left|\widetilde{\Sigma}\right|^{-(n+r+1)/2} \exp\left\{-\frac{1}{2} tr\widetilde{\Sigma}^{-1}\left(\widetilde{A} + \upsilon\widetilde{\alpha}\widetilde{\beta}'\Lambda^2\widetilde{\beta}\widetilde{\alpha}'\right)\right\} d\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right) \end{aligned}$$

where $\widetilde{c}_r$ absorbs the Jacobian for the transformation such that

$$c_r d(\alpha, \beta, \Sigma) = \widetilde{c}_r d\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right).$$

The only material difference between the prior $p\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right)$ and the prior $p(\alpha, \beta, \Sigma)$ is the term $\widetilde{\beta}' \Lambda^2 \widetilde{\beta}$ in the exponent rather than the term $\beta'\beta$. However it is this term that results in the prior for the cointegrating space being non-uniform. We use the results of Villani (2005) to demonstrate this. If we apply the linear normalization $\widetilde{\beta} = \begin{bmatrix} I_r \\ \widetilde{B} \end{bmatrix}$ as in Villani (2005) and integrate $p\left(\widetilde{\alpha}, \widetilde{\beta}, \widetilde{\Sigma}\right)$ with respect to $\left(\widetilde{\alpha}, \widetilde{\Sigma}\right)$ we obtain a prior for $\widetilde{B}$ proportional to

$$\left| I_r + \widetilde{B}' \Lambda_1^2 \widetilde{B} \right|^{-p/2}$$

which does not meet Villani (2005) Lemma 3.4 and is not uniform on the Grassman manifold (see Muirhead, 1982 or James, 1954), the support of the cointegrating space.

The prior in Strachan and Inder (2004) assumes $\beta'\beta = I_r$ such that the support for β is the Stiefel manifold, and the prior for β is uniform on the Stiefel manifold. This prior implies a uniform prior on the Grassman manifold, the support of $\mathfrak{p}$. As it is uniform in all directions and the support is spherical and compact, the prior on β is symmetric. Rescaling of the data retains the uniform distribution on the rescaled parameter, but the support is no longer spherical but ellipsoidal. For this reason, a uniform distribution on $\widetilde{\beta}$ implies an nonuniform prior on $\mathfrak{p}$.