Australian National University

CENTRE FOR APPLIED MACROECONOMIC ANALYSIS

The Australian National University

CAMA Working Paper Series **February, 2012**

EVIDENCE ON A DSGE BUSINESS CYCLE MODEL SUBJECT TO NEUTRAL AND INVESTMENT-SPECIFIC TECHNOLOGY SHOCKS USING BAYESIAN MODEL AVERAGING

Rodney W. Strachan
Research School of Economics, ANU
Centre for Applied Macroeconomic Analysis (CAMA), ANU

Herman K. van Dijk
Econometric Institute, Erasmus University Rotterdam
Econometrics Department VU University Amsterdam

CAMA Working Paper 3/2012
http://cama.anu.edu.au

Evidence on a DSGE Business Cycle model subject to 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 Institute, Erasmus University Rotterdam and
Econometrics Department VU University Amsterdam
email: hkvandijk@few.eur.nl

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

ABSTRACT

The empirical support for a DSGE type of real business cycle model with two technology shocks is evaluated using a Bayesian model averaging procedure that 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 equilibrium restrictions and restrictions on long-run responses to technology shocks apart from having a range of lag structures and deterministic processes. These model features are weighted as posterior probabilities and computed using MCMC and analytical methods. Uncertainty exists as to the most appropriate model for our data, with five models receiving significant support. The model set used has substantial implications for the results obtained. We do find support for a number of features implied by the real business cycle model. Business cycle volatility seems more due to investment specific technology shocks than neutral technology shocks and this result is robust to model specification. These technology shocks appear to account for all stochastic trends in our system after 1984. We provide evidence on the uncertainty bands associated with these results.

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 Dynamic Stochastic General Equilibrium (DSGE) business cycle model subject to a investment-specific technology shock and a neutral technology shock. We embed the DSGE model structure within a set 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 and requires careful consideration of prior distributions, the approach we propose makes such as exercise feasible and the empirical results suggest the exercise is worthwhile.

The DSGE model investigated in this paper is based upon one described by Fisher (2006). This model has several empirical implications that can be weighted: 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. Model uncertainty derives from uncertainty over the number of stochastic trends present in the system, further on the form of the deterministic trends, on lag length and, finally, on the form of the reduced form equilibrium (cointegrating) relations. By considering the unconditional evidence, where 'unconditional' means that the empirical evidence does not depend upon a single model, it is possible to identify those features that have stronger empirical support. The joint evidence for those features implied by the model will indicate its empirical support.[2]

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

[2]We use the word 'features' rather than 'structures' to avoid confusing our work with structural VAR analyisis, although we later consider 'structural breaks' in the common use sense of this term.

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 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), Hoogerheide, Kleijn, Ravazzolo, van Dijk and Verbeek (2010) 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).

There is clear evidence from the literature that a structural break should be considered around 1984 (see, for example, McConnell and Perez-Quiros (2000) and Stock and Watson (2002)). This literature suggests there is evidence of a break in possibly both the variances and mean equation coefficients. There is little work to date on changes in the overall structure or features of the model, such as changes in lag dynamics or stability of variables. We find the empirical evidence suggests that allowing for structural changes in the models, that is allowing the process to switch from one model to another, rather than just the parameter values, is justified.[3] However, as we discuss, incorporating this model switching is a computationally challenging task. It is a much simpler task to consider what was the most likely model before the break and which was the most likely model after the break, rather than trying to track the switch from one model structure to another and compute the evidence for that change.

This paper makes three contributions. First, we show 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 model structures. Second, this paper treats a structural break as a change in the entire structure of the model, not just a change in parameter values. We find strong evidence that the entire structure, rather than just the parameter values, has changed. This extension implies a very large model set but we demonstrate how to ob-

[3]The introduction of the structral break analysis was on the suggestion of a referee to whom we are very grateful. This extension has significantly altered the results and conclusions.

tain inference using some simple algebra and fast computation. Third, the proposed methodology is demonstrated with an empirical investigation of a DSGE 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 and empirically weighted features of the economic model used by Fisher (2006) are outlined. In Section 3 the basic econometric models of interest in this paper are introduced, including characterizations of the features implied by the economic model. We present priors, likelihood and the sampling scheme used in Section 4 together with the tools for inference in this paper including the posterior predictive probabilities (Geweke (1996) and Geweke and Amisano (2011)) of alternative model features and the Laplace approximation. The posterior and predictive evidence of model features are presented in Section 5 as are estimates of important functions of parameters. In Section 6 we summarize conclusions and discuss possibilities for further research.

2 A DSGE Business Cycle Model

In this section we outline the features of a DSGE model that is based upon the real business cycle model of Fisher (2006), which is in turn is based upon the competitive equilibrium growth model of Greenwood, Hercowitz, and Krusell (1997). We impose 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 papers for the development of the model as we focus upon certain features that imply restrictions upon our reduced form econometric model that we want to weight using Bayesian model averaging.

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 is 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 investment specific and neutral technology 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

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

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

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

where $\gamma \geq 0$ and $(z_{I,t}, z_{N,t})'$ 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 Great Ratio relations $c_t - w_t$ 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 over short periods, and this does appear to be the case over subsamples, it is not possible for hours worked per capita to have a permanent 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 log Great Ratio relations will be $I(0)$ and form valid cointegrating relations. The assumptions of Fisher (2006) preclude the above $I(0)$ relations having deterministic trends and this is not a feature we would expect to find over long samples. By allowing for structural breaks we may find trends in one or more subsamples, perhaps as a trend in the second period off-sets the effect of the trend in the first period, but we would expect that they are inherently temporary features.

Two important final restrictions apply. First, Fisher (2006) assumes that 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, $z_{I,t}$, and the neutral technology shock, $z_{N,t}$.

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

$$\Delta y_t = y_{t-1}\beta\alpha + d_t\mu + \Delta y_{t-1}\Gamma_1 + \ldots + \Delta y_{t-l}\Gamma_l + \varepsilon_t \quad (1)$$

where $\Delta y_t = y_t - y_{t-1}$. The $1 \times n$ vector of errors ε_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.

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 $r = 0, 1, \ldots, n$. 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. This restriction implies the elements of the first row of β are zeros. If $o = 2$ then the

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

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, $c_t - w_t = c_t - a_t - h_t$ and $x_t - w_t = x_t - a_t - h_t$, 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 :

$$H_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ -1 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The restriction $o = 1$ is allowed to hold when $0 < r < 5$ and $o = 2$ when $r = 3$.

We allow for five different lag lengths such that $l \in \{2, 3, 4, 5, 6\}$. 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 Wold representation of the VECM as

$$\Delta y_t' = C \Omega^{1/2} z_t' + C^* (L) \Omega^{1/2} \Delta z_t' \text{ where } C = \beta_\perp (\alpha_\perp \Gamma \beta_\perp)^{-1} \alpha_\perp.$$

The first two elements of $z_t = \Omega^{-1/2}\varepsilon_t$ are the investment-specific technology shock, $z_{I,t}$, and the neutral technology shocks, $z_{N,t}$, respectively. Further, $\Gamma = I_n - \Sigma_{i=1}^{l}\Gamma_i$. The restriction on $C\Omega^{1/2}$ implies the matrix will have the following zero entries:

$$C\Omega^{1/2} = \begin{bmatrix} c_{11} & 0 & 0 & & 0 \\ c_{21} & c_{22} & 0 & \cdots & 0 \\ * & * & * & & * \\ & \vdots & & \ddots & \\ * & * & * & & * \end{bmatrix} \tag{2}$$

where the asterisks (*) imply no restriction is imposed. This restriction is obtained by the appropriate choice of $\Omega^{1/2}$ such that $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$ and $C\Omega^{1/2}$ has the structure shown above. For an explanation of how such restrictions are implemented see the Appendix, and see Chang and Schorfheide (2003) and Del Negro and Schorfheide (2010) for further examples and discussion.

Since C has rank $n-r$, and $\Omega^{1/2}$ must have rank n, the zero restrictions on $C\Omega^{1/2}$ and the assumption of nonzero responses of p_t and a_t stated above imply C must have at least rank two, then this identification scheme 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.

Fisher (2006) assumes a break date around 1982 Q3, although our results suggest a slightly later date. We allow for a range of 16 break dates from the first quarter of 1982 until the last quarter of 1985. We index these dates by $\tau \in \{1, 2, \ldots, 16\}$ where $\tau = 1$ denotes the break at 1982Q1, $\tau = 2$ denotes the break at 1982Q2, and so on.

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 break date (τ). As the model features may change after the break date, we denote a model by the product $M_{ij\tau} = M_i^{0,\tau}M_j^{1,\tau}$ where $M_i^{b,\tau}$ denotes the model $i = (d, l, r, o)$ with break date τ prior to the break date if $b = 0$ and after the break date if $b = 1$. The vectors i and j index the features (d, l, r, o) and may be different to allow the model features to change post the break. The vectors $i, j \in \Xi$, where Ξ is the set of all i and j considered. As an example of some models we will use, suppose prior to a break at 1984Q3, we allow a linear drift and nonzero mean in the cointegrating relations $(d = 3)$, two lags of differences $(l = 2)$, and stationary great ratios and hours worked $(r = 3, o = 2)$.

This model would be denoted as $M^{0,11}_{(3,2,3,2)}$. The features of the VAR used in Fisher (2006) can be represented as the model $M^{0,-9}_{(3,4^*,0,0)} M^{0,3}_{(3,4^*,0,0)}$. However, Fisher uses additional restrictions - such as the equation for p_t has an extra lag of a_t - and Fisher does not use the same set of variables. Further, his specification does not capture all of the features of the economic model (such as the implied cointegration and stability of the Great Ratios). For these reasons we do not consider the econometric model of Fisher further.

In total we average using 255 models before the structural break, and 255 after the break.[5] As we allow the entire structure of the model to change before and after the break, we therefore have $(255)^2$ models for each break date. Further, we allow 12 possible values for the break dates giving a total of over $780,000$ models. Computation of marginal likelihoods for this many multivariate models would be computationally challenging. As our interest is in the date when the break occurs and modelling after the break date, we need only estimate 510 marginal likelihoods for 12 break dates; a total of 6120 marginal likelihoods. This is still a large number of computations, but much more readily achievable. We provide further details on how this is achieved in Section 4.3 and in the Appendix.

4 Priors, Posteriors and Model Averaging.

In this section the priors and resultant posterior are presented. We begin by specifying the prior followed by a discussion of our choice of this prior. We separate the discussion to allow readers more interested in the empirical results to proceed directly to that section. For notational convenience we collect the lag parameters into a $k_i \times n$ matrix $\Phi = [\Gamma_1' \quad \cdots \quad \Gamma_l']'$ and vectorize into $\phi = vec(\Phi)$. Conditional upon β, the model in (1) is linear in the equation parameters μ, $vec(\alpha)$ and ϕ. This fact makes it relatively straightforward to elicit priors on Ω, μ, $vec(\alpha)$ 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.

[5] There are 450 models implied by the restrictions. However, this reduces to 255 models when we exclude *a priori* impossible models, meaningless models and only consider one in a set of observationally equivalent models.

4.1 The Prior.

All models included for the averaging are treated equally likely. The set of models, Ξ, included for the averaging is a subset of the full set of models that result from all combinations of d, l, r, and o and 16 break dates τ. We exclude impossible and meaningless models and, if two or more models are observationally equivalent, we include only one of them. The implications of these choices are discussed below.

To avoid notational burden, and because we use the same prior for models before and after the structural break, in this section we do not distinguish between parameters and models before and after the break dates except where necessary.

For Ω we use a proper inverted Wishart prior with scale matrix $\underline{S} = I_n 10$ and degrees of freedom $\underline{\nu} = n + 1$ as this prior is rather uninformative. The parameters in Φ, the Γ_i, are given a normal prior with zero mean and covariance matrix

$$\underline{V}_\phi = \left(Iu + \underline{V}_0 \left(1 - u\right)\right) \left(I \left(1 - u\right) + \underline{V}_1 u\right)$$

where $u \in \{0, 1\}$ with prior probabilities $\Pr(u = 1) = \Pr(u = 0) = 0.5$.[6] Here $\underline{V}_0 = \Omega \otimes I_{k_i}$ and $\underline{V}_1$ is the Litterman type prior for a VECM similar to that specified in Villani (2001). We use a gamma prior with mean $E(\eta) = 5$ and a relatively large variance $V(\eta) = 16.67$ for η.

We specify a weakly informative normal proper prior for $vec(\mu, \alpha)'$ conditional upon (Ω, β, M_i) (and hyperparameters discussed below) with zero mean and covariance matrix $\frac{1}{\eta}\underline{V}_a$ where $\underline{V}_a = \Omega \otimes I_r$.[7] Further details on the specification of the full prior is given at the end of the next subsection.

This paper uses a semi-orthogonal specification for β, i.e., $\beta'\beta = I_r$, and a Uniform distribution for β. 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).

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

[7] 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).

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 chosen for the prior. 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). As the matrices α and β always occur in a product form as $\beta\alpha$, 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, $V_{n,r}$, (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).

To give this a more formal explanation let the vector of all parameters in the model that appear in the likelihood, i.e., β, α, μ, ϕ, and Ω , be denoted by θ and the unrestricted support is Θ, $\theta \in \Theta = V_{n,r} \times R^{n(k_i+r)} \times R_+^{n(n+1)/2}$ (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 $p(\theta)$.[8]

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

[8] In an earlier version of the paper, without structural breaks, we restricted the support to exclude explosive roots. With the structural breaks we found the stationary region to capture virtually all of the posterior mass and so the restriction became effectively redundant.

prior distribution for the parameters in a given model is then

$$\begin{aligned} p\left(\theta^{*}, \eta, u \mid M_{i}\right) \propto\; & \exp \left\{-\frac{\eta}{2} a^{* \prime} \underline{V}^{-1} a^{*}-\frac{n}{2} b^{* \prime} b^{*}\right\} n^{n r / 2} \\ & \times|\Omega|^{-\left(\underline{\nu}+n+1+r+u k_{i}\right) / 2} \exp \left\{-\frac{1}{2} \operatorname{tr} \Omega^{-1} \underline{S}\right\} \\ & \times \eta^{\frac{n\left(k_{i}+r\right)+1}{2}} \exp \left\{-\frac{5 \eta}{6}\right\} . \end{aligned}$$

For further details and proofs of results, see Strachan and van Dijk (2010).

4.2 Discussion of the Prior.

Ideally all models would be treated as *a priori* equally likely, however this is not a straightforward issue in VECMs.[9] The priors for the individual elements of $i = (d, l, r, o)$ 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.[10] 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, and o. One might expect, for 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 probability $\Pr(r < n|d)$ may decrease as d increases. Similarly a shorter lag length, l, might be associated with a higher prior probability of

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

[10] 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.

finding (more) stochastic trends. We do not pursue this idea further, but note that it might be a worthwhile topic for investigation.

For 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 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 $\frac{1}{\eta}\underline{V}_0$ and the other with the covariance matrix similar to that in Villani (2001), which we will denote by $\frac{1}{\eta}\underline{V}_1$.

The covariance matrix $\frac{1}{\eta}\underline{V}_1$ 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 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 $\Pr(u=1) = \Pr(u=0) = 0.5$. The posterior estimate of u will inform us on the data's preference between the two specifications and in this 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) with estimates of $\Pr(u = 1|y)$ showing some preference for $\underline{V}_0$, but not strong evidence for either covariance matrix. It would seem, therefore, that mixing over the two normals, rather choosing one, is a reasonable approach.

The parameter η determines the overall degree of shrinkage that is applied to the mean equation parameters. Evidence on the influence of this parameter can be found in Strachan and Inder (2004). The settings we use 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 from a preliminary run, 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.

The specification of β is semi-orthogonal, i.e., $\beta'\beta = I_r$, with a Uniform distribution for β. This specification permits estimation with minimal restrictions (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 by imposing such identifying restrictions as these restrictions 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.[11] The relevant point we make here is that any inference that can be achieved with linear identifying restrictions can be achieved with the identifying restrictions in this paper.

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.[12] By using the semiorthogonal restrictions we can produce unique inference (see Lopes and West (2004) for an example where choice of linear restrictions produces difference results), we have no inconsistency between the model and our assumptions (Strachan and Inder

[11] More recently, the topic of invariance to rescaling of the data has been raised in conversations with colleagues. Our prior is not invariant but, unfortunately, no uniform, invariant prior exists in the literature. While it might be worth further investigation, we do not consider invariance further as it has not proven to be an empirically important issue.

[12] 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.

(2004) show how imposing linear restrictions has the unexpected and undesirable result that it makes the assumption supporting the restrictions *a priori* impossible), and ensure that moments exist, the posterior is always proper, no local nonidentification problems arise and the Markov chain methods used are irreducible (Kleibergen and van Dijk (1994 & 1998) and Bauwens and Lubrano (1996)).

4.3 Posterior Analysis.

We conduct the empirical investigation in two stages. In the first, we compute the estimated model probabilities and the timing of the structural break. In the second stage, we estimate the functions of interest such as impulse responses and variance decompositions. We estimate the model probabilities using a Laplace approximation of predictive densities, and the impulse responses and variance decompositions are estimated using a Gibbs sampler.

When investigating the evidence on structural breaks we encounter two issues that need to be addressed: the strength of the prior information; and proliferation of models to compute. As we have a separate prior for the parameters before and after the break, we have effectively doubled the amount of prior information in the posterior and halved the amount of data (on average) used to estimate the parameters. This introduces the problem that the prior information is strong relative to the data. We could make the priors less informative, but then that increases their influence in the computation of the posterior probabilities (see discussion on Bartlett's paradox in Geweke (2005), Section 2.6.2). This situation implies a trade off between prior uncertainty about the parameters and posterior uncertainty about the models. To mitigate this issue we compute predictive probabilities (see, for example, Geweke (1996) and Geweke and Amisano (2011)) from predictive densities.

The predictive densities can be derived for each model via the expression

$$p_i^{b,\tau} = p\left(y_{T_1+1}^{T_2} | y_{T_0}^{T_1}, M_i\right) = \frac{\int_\Theta p\left(y_{T_1+1}^{T_2} | y_{T_0}^{T_1}, M_i, \rho\right) p\left(y_{T_0}^{T_1} | M_i, \rho\right) p\left(\rho | M_i\right) d\rho}{\int_\Theta p\left(y_{T_0}^{T_1} | M_i, \rho\right) p\left(\rho | M_i\right) d\rho} \tag{3}$$

where $\rho = (\theta^*, u, \eta)$ and $y_{t_0}^{t_1}$ is the data from observation t_0 to t_1. For a model prior to the structural break at time τ, then $T_0 = 1$ and $T_2 = \tau$. For a model after the structural break, $T_0 = \tau + 1$ and $T_2 = T$. As $T_2 \to \infty$ or $T_1 \to 0$, then $p\left(y_{T_1+1}^{T_2} | y_{T_0}^{T_1}, M_i\right)$ becomes the standard marginal likelihood. The notation $p_i^{b,\tau}$ denotes the predictive density for model M_i with a break date τ for the time period before the break if $b = 0$ and in the time period after the break if $b = 1$. Thus the full predictive density for a process that switches from model M_i to model M_j at time τ is the product $p_i^\tau = p_i^{0,\tau} p_j^{1,\tau}$.

We estimate the numerator and denominator in (3), and therefore $p_i^{b,\tau}$, using the Laplace approximation. This technique approximates the integral by a second order expansion around the mode of the log of the integrand (see Tanner (1993)). Although often described as a normal approximation, the Laplace approximation has been shown to work very well with very non-normal distributions. For example, Strachan and Inder (2004) apply the approach to the VECM where the integral is over linear subspaces which have bounded supports and non-standard distributions.

At a given break date τ, using the fact that assume all prior model probabilities are equal, we can compute the posterior predictive probability of a model $M_i^{b,\tau}$ holding before ($b = 0$) or after ($b = 1$) the break using the $p_i^{b,\tau}$ as

$$p\left(M_i^{b,\tau} | y_{T_0}^{T_2}\right) = \frac{p_i^{b,\tau}}{\sum\limits_{j\in\Xi} p_j^{b,\tau}}. \tag{4}$$

While using the Laplace approximation greatly speeds up the computation of the model probabilities, as discussed earlier, we allow the entire model structure to change at the break. This implies we have a very large number of models because the process is allowed to switch from any model before the break to any model after the break. To further reduce computation time, we make use of the fact that we are only interested in computing the probability that a break occurs at a point in time, and not which model after the break followed a specific model before the break. This observation greatly reduces the number of necessary computations.

If we have M models and n possible break dates, then the posterior probability of a break at date $\tau = \tau^*$ is obtained by

$$p\left(\tau = \tau^* | y\right) = \frac{\prod\limits_{b=0}^{1} \left(\Sigma_{i=1}^{M} p_i^{b,\tau^*}\right)}{\sum_{\tau=1}^{n} \prod\limits_{b=0}^{1} \left(\Sigma_{j=1}^{M} p_j^{b,\tau}\right)}.$$

This expression is explained with a simple example in the Appendix. The probability of a particular feature (e.g., $d = 2$) holding after the break is then obtained by summing the products of the probabilities of the models with that feature and the probabilities of their break dates.

As we did for the presentation of the priors for the parameters, in presenting the posterior parameter distributions we suppress notation indicating which sample is being discussed. Thus models are denoted by M_i to represent $M_i^{b,\tau}$ and no notation is included on the parameters indicating pre or post break or which break is assumed. This simplification is to reduce the notational burden and little information is lost by doing so.

To estimate the impulse responses we require draws from the posteriors of the models with non-negligible posterior weight. 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,[13]

$$p(\theta^*, \eta, u|M_i, y) \propto p(\theta^*, \eta, u|M_i) L(y|\theta^*, M_i). \tag{5}$$

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 $z_t = (y_{t-1}\beta^*, z_{2,t})$. Next 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 $b^* = vec(\beta^*)$, standard results show that the posterior for $a^* = vec(A)$ conditional on all other parameters will be normal with mean $\overline{a}$ and covariance matrix $\overline{V}$ constructed as

$$\overline{a} = \overline{V}\left(\Omega^{-1} \otimes I_{k_i+r}\right) vec\left(\sum_{t=1}^{T} z_t' \Delta y_t\right)$$

and

$$\overline{V} = \left(\left(\Omega^{-1} \otimes \sum_{t=1}^{T} z_t' z_t\right) + \eta \underline{V}^{-1}\right)^{-1}.$$

Next, the posterior for b^* conditional upon the other parameters will be

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

normal with mean $\overline{b}$ and covariance matrix $\overline{V}_b$ which are constructed as

$$\overline{b} = \overline{V}_b \left(\alpha^* \Omega^{-1} \otimes I_n\right) vec \left(\sum_{t=1}^{T} y'_{t-1} \left(\Delta y_t - z_{2,t}\Phi\right) \right)$$

and

$$\overline{V}_b = \left[\left(\alpha^* \Omega^{-1} \alpha^{*\prime} \otimes \sum_{t=1}^{T} y'_{t-1} y_{t-1} \right) + n I_{nr} \right]^{-1}.$$

The posterior for η will be Gamma with degrees of freedom $\overline{\nu}_\eta = n\left(k_i + r\right) + 3$ 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 $\overline{p} = \Pr\left(u = 1|a^*, \Omega, \beta^*, y\right)$ 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 $(a^*, \Omega, \beta^*, \eta, u)$:

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 $IW\left(\underline{S} + u\eta A'A + \sum_{t=1}^{T} \varepsilon'_t\varepsilon_t, T + uk_i + r\right)$;
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 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.

4.4 Bayesian Model Averaging with MCMC.

In this section we outline how we implement Bayesian model averaging to provide unconditional inference. For this discussion we again abstract from the structural break and specific model features and use M_i to denote a generic model. This might be, for example, a specific model before or after a break $M_i = M_i^{b,\tau}$ with posterior probability $p\left(M_i|y\right) = p\left(M_i^{b,\tau}|y\right)$.

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 J draws of the parameters from the posterior distribution for each of the M models, first obtain estimates of $E(\zeta|y, M_i)$ from each model by

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

These estimates are then averaged as

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

in which $\widehat{p}(M_i|y)$ is an estimate of $p(M_i|y)$ presented in Section 4.3.

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 providing evidence on the timing of the structural break, and the posterior probabilities of the features of the reduced form VECM that are implied by this RBC model. We estimate the posterior probability of a structural break occurring at a range of dates and, in subsequent analysis, focus upon the results for the data after the structural break. We then report the estimates of objects of interest including impulse response functions.

The variables and the data: 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.

Break Dates and Stability of Great Ratios: As discussed in Section 3, for each break date, excluding observationally equivalent, impossible or meaningless models leaves 255 models to estimate. Three break dates receive measurable support, 1984 Q3, 1984 Q4 and 1985 Q1 with probabilities of 0.01%, 57.78% and 42.21% respectively. Of the 765 models over these three dates, there were 62 post break models with measurable support and 12 pre-break models with measurable support. From the pre-break models, the model $M^{0,12}_{(4,2,1,0)}$ has 57.8% of the probability mass and $M^{0,13}_{(4,2,1,0)}$ has 41.2% posterior probability. The models prior to the break provide strong evidence against the features we would expect under the economic model. While the results prior to 1984 have important information in them, there is evidence of several breaks in the 1970. As the aim of this paper is not a comprehensive treatment of all breaks, the remainder of the discussion focuses on the post break results as this period is well represented by a single, stable model.

Among the post break models, one model accounted for half the posterior mass and twenty models accounted for 99.99% of the posterior probability mass. The posterior probabilities of the top six models are presented in Table 1. Post the break, the models $M^{1,12}_{(2,2,3,2)}$ with a break at 1984 Q4 and $M^{1,13}_{(4,3,2,1)}$ with a break at 1985 Q1 capture 79% of the probability mass. Although relatively few models get any support, it is clear that support is fairly strong for the top five models.

According to the model of Fisher, it might be reasonable to expect few deterministic processes in the system and so to see $d \leq 3$ as the technologies are commonly described as random walks possibly with drifts, but the economic model does not suggest we would expect trends in the cointegrating relations ($d = 2$) or quadratic trends in the variables ($d = 1$). Table 2 presents the marginal probabilities of the various features of the VECM. There is a 76% probability that there is a drift in the levels and a trend in the cointegrating relations ($d = 2$) and 24% probability of a drift in levels but no trend in the equilibria ($d = 3$). With a 76% posterior probability that the cointegrating relations are the Great Ratios of consumption to income and investment to income, the trend may be picking up the decline in the savings that occurred since the mid 1990s.

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

d	l	r	o	*Break Date*	$P(M_i\|y)$	Cumulative probabilities
2	2	3	2	1984.4	0.5552	0.5552
4	3	2	1	1985.1	0.2343	0.7895
2	2	3	2	1985.1	0.0968	0.8863
2	3	3	2	1985.1	0.0884	0.9747
2	3	3	2	1984.4	0.0226	0.9973

Table 2: Posterior probabilities of model features after the break.

$d=2$	$d=4$	$d=5$	$l=2$	$r=3$
0.7636	0.2347	0.0019	0.6546	0.3454
$r=2$	$r=3$	$o=1$	$o=1$	$o=2$
0.2364	0.7638	0.0019	0.2348	0.7635

The posterior probability of having only two stochastic trends is 76%, although there is evidence of a third if the break is delayed until 1985 Q1. As the primary aim of this paper is to investigate the degree and role of model uncertainty in the empirical evidence on technology shocks and important features of RBC models, we discuss the results with reference to a number of important papers in the literature. 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 a unique technology is (in the three variable model) the only stochastic trend that enters the system. The economic model of Fisher implies there are only two common 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. We conclude the evidence on an extra unit root is an empirical issue. It is possible the extra stochastic trend could be entering from the hours worked variable, h_t. While h_t does not display a trend, it does display a very large and slow cyclical component with peak-to-peak cycles of around 10 years duration and a sudden drop in 2009. Such large cycles and outliers are not easily captured in linear models (such as the VECM) and may be manifesting as evidence of a third unit root. An alternative explanation is provided in Chang, Doh and Schorfheide (2007), that hours may appear nonstationary if labour adjustment costs are not explicitly incorporated into the model.

The data are positively informative about the form of the cointegrating relations. That is, there is a 76% probability that the Great Ratios are stable and that the investment price has a unit root and does not cointegrate with the other variables. The assumptions that the price of an investment good is nonstationary and does not cointegrate with any other variable in the system, with 99% probability mass on $o = 1$ and $o = 2$, have very strong support. Since the posterior probability that $(r = 3, o = 2)$ is 76%, and the posterior probability of $(r = 3, o = 0$ or $o = 1)$ is zero, the evidence that the Great Ratios are stationary is positive, although not compelling.

Overall the evidence in the estimated probabilities for the features of the econometric model suggested by the RBC of Fisher (2006) positively to strongly supported. But, with 25% of the mass on features not supported by the economic model, the evidence is not decisive and there remains considerable model uncertainty.

Business Cycle Volatility due to Investment and Neutral Technology Shocks: An important area of interest in RBC models is the dynamics of w_t, c_t, and x_t, including the role of the investment specific and neutral technology shocks in the business cycle. By decomposing the variance into the components due to these technology 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, 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 models used in other studies. KPSW and CC did not consider the two technology shocks *per se*, rather the role of permanent (possibly technology) shocks and transitory shocks. As the model in this paper has two additional variables (p_t and h_t), the results will differ from those if we had used exactly the models in KPSW and CC unless (p_t, h_t) is strongly exogenous to (c_t, x_t, w_t) . 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 a 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 our study suggest there is some uncertainty associated with this aspect of the theory as the evidence suggests there may be more than two stochastic trends entering the system. However, the equilibrium relations appear to be well described by the Great Ratios. Notwithstanding this ambiguity, it is not evident that the excess of stochastic trends affects estimates of other outputs such as proportions of the variance over the business cycle that can be attributed to the technology shocks.

Our interest is in the proportion of business cycle fluctuations due to the technology shocks in total, and the investment specific shocks, $z_{I,t}$, and neutral technology shocks, $z_{N,t}$, specifically. Therefore we adapt the approach of CC who consider the variance decomposition within the frequency domain.

Figure 1 presents the posterior distribution averaged over all models, of the proportion of the variance of c_t, x_t and w_t over the business cycle due to investment specific technology shocks and Figure 2 presents the posterior distribution of the proportion due to neutral technology shocks constructed by averaging over all models. These estimates are obtained from 30,000 draws from the posterior of each model. This plot shows an amount of mass at zero for all three variables suggesting a slightly larger role for non-technology shocks. Relatively, the investment specific technology shocks are far more important the neutral technology shocks and the proportions have considerable mass away from zero. The proportion of variation due to neutral technology shocks has little mass above 15%. We computed the same posterior densities as those reported in Figures 1 and 2 but using the CC, KPSW and the best models, as well as for all models in which the Great Ratios are the cointegrating relations. These estimates all share similar forms and pairwise comparisons showed no stochastic dominance of any distribution over any other.

Table 3 reports estimates of the mean proportions of the variance over the business cycle due to $z_{I,t}$ and $z_{N,t}$ for a range of models and model sets. The single models considered are those used by CC, KPSW and the best model in our model set. The model sets we consider are, first, the set of all models and, second, the set of all models in which the Great Ratios are stable. Comparing across different models and model sets, the proportions appear relatively robust to specification. The CC model is the only one that does not include the Great Ratios as stable cointegrating relations and we see that this results in slightly higher proportions of the variance due to investment specific technology shocks for output and consumption, although the difference is not large. Imposing the Great Ratios (as in KPSW, the best model and the set of all models with stable Great Ratios) slightly reduced the role of $z_{I,t}$ in business cycle variation for consumption and output, but not for investment. Overall, the technology shocks explain between 26% and 38% of the variation over the business cycle for these variables.

Table 3: Estimated proportion of variance over the business cycle due to the investment-specific technology shock, $z_{I,t}$, and the neutral technology shocks, $z_{N,t}$

		CC model	KPSW model	Best model	All models	Balanced Growth models
c_t	$z_{I,t}$	0.299	0.179	0.212	0.223	0.212
	$z_{N,t}$	0.050	0.018	0.041	0.038	0.040
x_t	$z_{I,t}$	0.254	0.320	0.233	0.227	0.234
	$z_{N,t}$	0.051	0.01	0.035	0.033	0.035
w_t	$z_{I,t}$	0.281	0.160	0.202	0.216	0.202
	$z_{N,t}$	0.097	0.045	0.082	0.076	0.081

Relatively speaking, the investment specific technology shocks seem far more important than the neutral technology shocks as they explain between 70% and 97% of the total variation due to technology shocks.

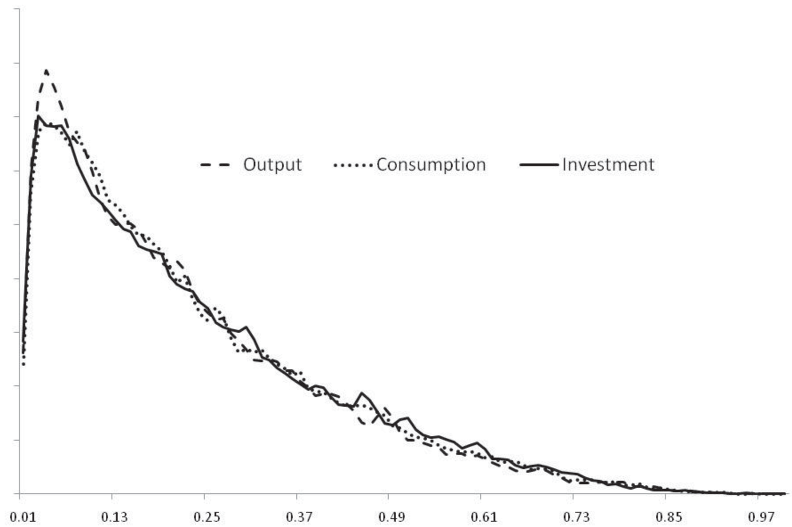

Figure 1: Posterior densities of the proportion of the variance over the business cycle of output, consumption and investment that is due to investment specific technology shocks, $z_{I,t}$.

Propagation of Shocks: We conclude by discussing the responses in investment price, hours, output, consumption and investment to technology shocks. The percentiles of the posterior distribution of the response of investment price p_t to an investment specific shock are shown in Figure 3. This path is very much as would be expected from the model of Fisher and agrees with the form of his estimates. The plot of the percentiles of the posterior response of p_t (not reported but see Figure 7 discussed below) to a neutral technology shock, $z_{N,t}$, shows little response, again agreeing with the results of Fisher. Figures 4, 5 and 6 show the posterior mean impulse responses to

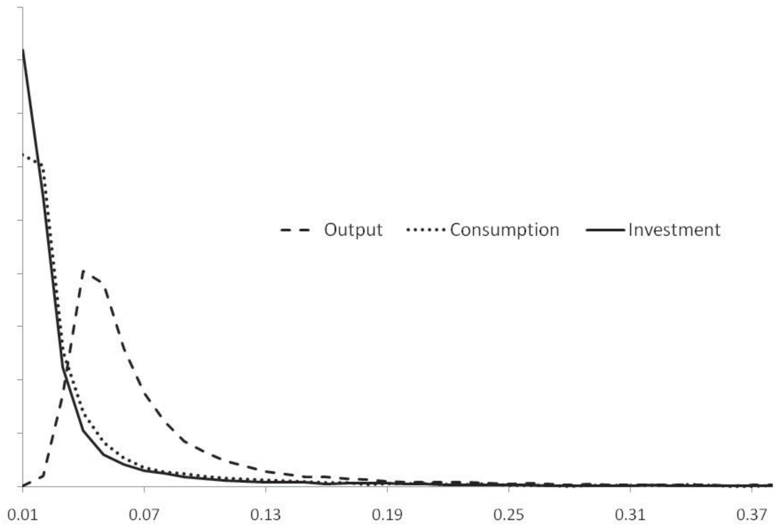

Figure 2: Posterior densities of the proportion of the variance over the business cycle of output, consumption and investment that is due to neutral technology shocks, $z_{N,t}$.

the technology shocks constructed from 30,000 draws of the parameters in each included model. Output has a positive response to both shocks while consumption does not seem to react. Investment has an initially negative and then positive response to investment specific technology shocks, but declines in response to neutral technology shocks. The response of hours worked shows a clear difference depending upon the type of technology shock. Investment specific shocks lead to a gradual fall in hours worked while the neutral shocks have an initially positive effect that quickly dies away. The mean responses of hours worked to the two technology responses is presented in Figure 6. This figure tells a similar story to the results of Fisher that the response of hours to neutral technology shocks is short lived. However, they suggest quite different responses to investment specific shocks those reported in Fisher (2006). Both our results and those of Fisher suggest persistence in the response, however ours show a negative response while Fisher's results suggest an increase in hours worked.

While many of these figures agree in form with those of Fisher (2006), they differ in magnitude by a factor of 10. With the exception of the response of investment price to $z_{I,t}$, the mean paths are not far from zero if we take the full posterior as the metric for distance. The posterior credible intervals for many of the responses encompass the origin. The densities for the responses of the investment price to a neutral technology shock in Figure 7 at three horizons give an impression of the general form and evolution of the response densities for other variables. The leptokurtic form and mass generally around zero is representative of the responses of a_t, w_t, h_t, c_t and x_t to investment specific shocks, and of h_t to neutral technology shocks.

The responses of a_t, w_t, c_t and x_t to neutral technology shocks show a very different response. As an example, Figure 8 presents the percentiles of the posterior distribution of the response of output w_t to an neutral technology shock at three different horizons. These densities all have platykurtic forms at short horizons, responses after one or two periods. However, they become increasingly bimodal at the time since the shock increases. This bimodality is not due to different models producing different paths as the responses for the best model in Table 1 produced very similar responses. A number of possible explanations exist, and one is that the identifying restrictions, at least for $z_{N,t}$, disagree with the data. However, as we are using just identifying restrictions it is not possible for us to report evidence at this point on this question.

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 features implied by this model we find reasonable support and some, such as stability of the Great Ratios, receive quite strong support, although not before 1984. Further, the impulse responses demonstrate that the predictions of the model are quite plausible although there is considerable uncertainty around these estimates.

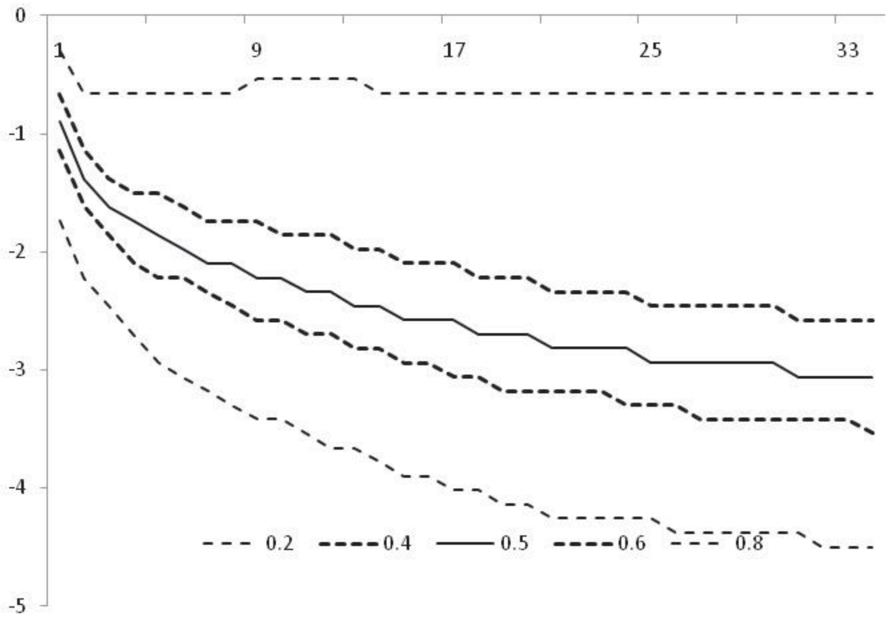

Figure 3: The 20%, 40%, 50%, 60% and 80% percentiles of the posterior distribution of the response of investment price to an investment specific technology shock. The x-axis is periods since the shock and the y-axis is in percent.

The methodology is an important contribution of this paper. The approach 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. To account for structural breaks and model switching, methods are presented to make inference computationally feasible in a very large set of multidimensional models.

The methods presented in this paper, without the structural breaks, have already found applications in several other areas. Koop, Potter and Strachan (2005) investigate the support for the hypothesis that variability in US wealth

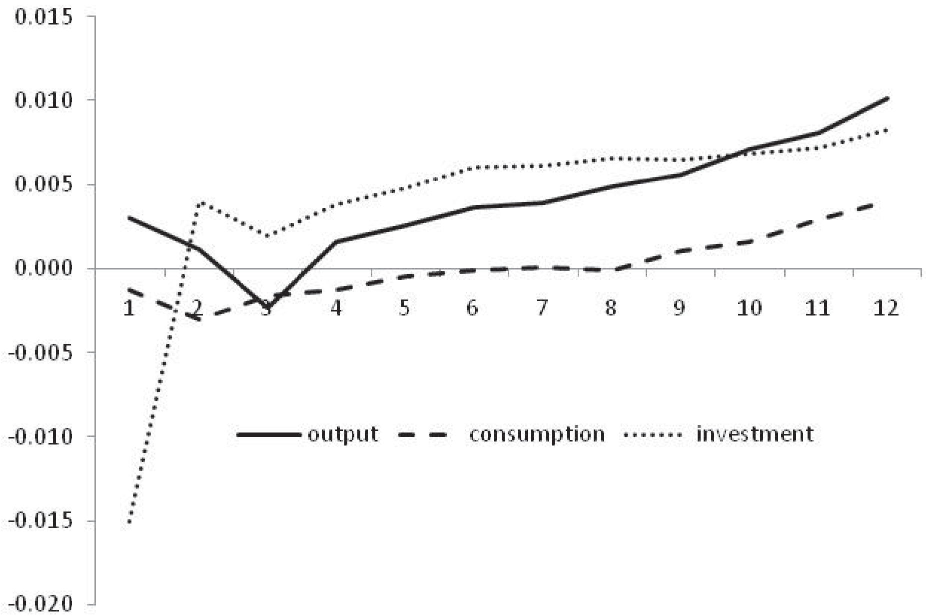

Figure 4: Mean impulse responses of output, consumption and investment to an investment specific technology shock, $z_{I,t}$. The x-axis is time since the shock and the y-axis is percent.

is largely due to transitory shocks. They demonstrate the sensitivity of this conclusion to model uncertainty. Koop, León-González and Strachan (2008) develop methods of Bayesian inference in a flexible form of cointegrating VECM panel data model. These methods are applied to a monetary model of the exchange rate commonly employed in international finance. Other current work includes investigating the impact of oil prices on the probability of encountering the liquidity trap in the UK and stability of the money demand relation for Australia.

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 time variation in the model itself as Cogley and Sargent (2001, 2005) and Primiceri (2005) do for the VAR. These papers suggest

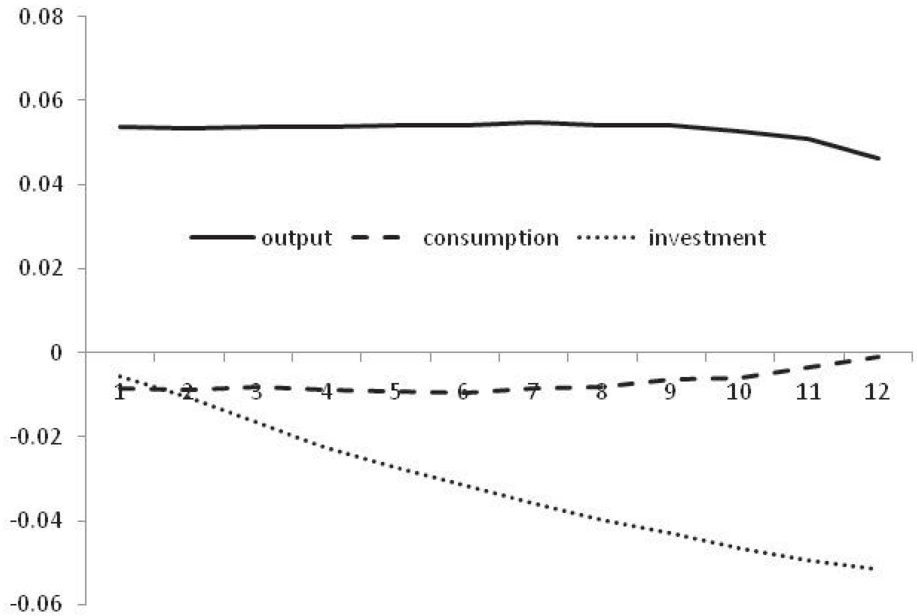

Figure 5: Mean impulse responses of output, consumption and investment to an neutral technology shock, $z_{N,t}$. The x-axis is time since the shock and the y-axis is percent.

there is considerable variation in the parameters, particularly over the 1970s. However, the reduced rank restrictions due to cointegration introduce further conceptual and computational issues for time varying models and a first example of how to implement such a time varying VECM is presented in Koop, Leon-Gonzales and Strachan (2011). Alternatively, in using a SVAR for business cycle analysis one may use prior information on the length and amplitude of the period of oscillation (see Harvey, Trimbur and van Dijk (2007)). Unresolved issues in this work include systematic use of inequality conditions which imply a more intense, or better, use of MCMC algorithms. Extending this approach to large model sets is yet to be considered and, given the computational issues, remains a challenge. 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.

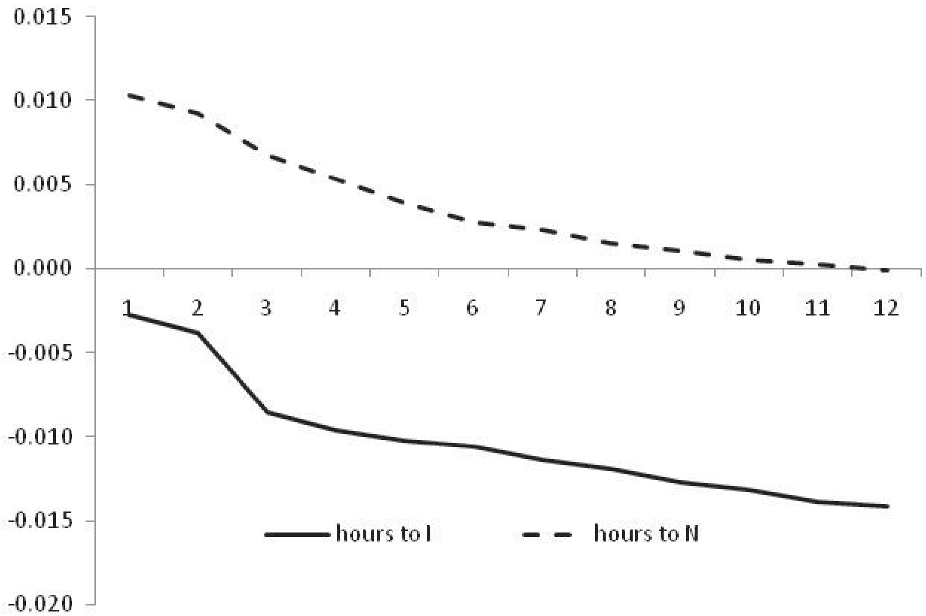

Figure 6: Mean impulse responses of hours to an investment specific shock, $z_{I,t}$, (denoted by hours to I) neutral technology shock, $z_{N,t}$, (denoted by hours to N), The x-axis is time since the shock and the y-axis is percent.

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 which led to a substantial revision and extension of the results presented in the original working paper. We also thank Luc Bauwens, Geert Dhaene, John Geweke, 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.

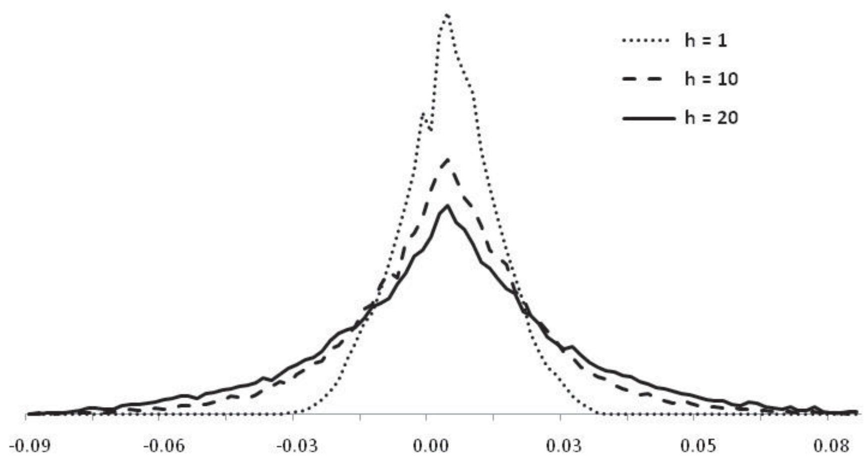

Figure 7: Posterior densities of the impulse responses of investment price, p_t, to a neutral technology shock, $z_{N,t}$, at horizons h = 1, h = 10 and h = 20. The x-axis is in percent.

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.

Chang, Y., and F. Schorfheide (2003): "Labor-Supply Shifts and Economic Fluctuations," *Journal of Monetary Economics* **50(8)**, 1751-1768.

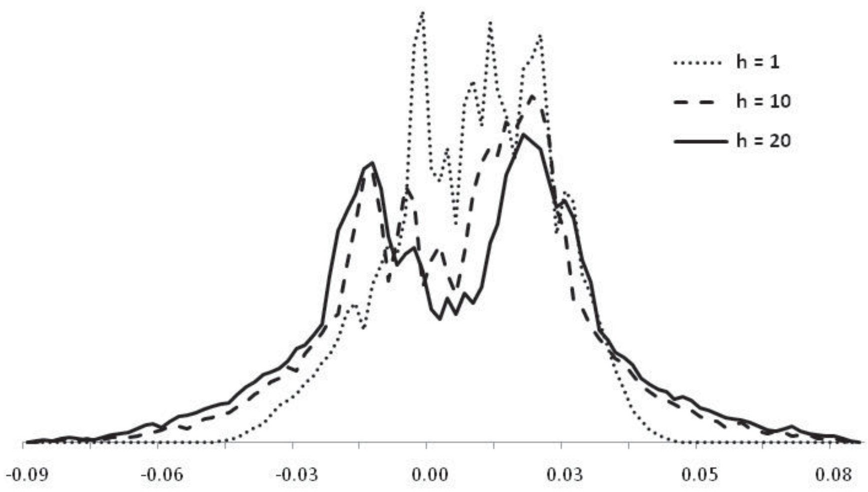

Figure 8: Posterior densities of the impulse responses of output, w_t, to a neutral technology shock, $z_{N,t}$, at horizons h = 1, h = 10 and h = 20. The x-axis is in percent.

Chang, Y., Doh, T., and F. Schorfheide (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit, and Banking*, 39(6), 1357-1373.

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.

Del Negro, M., and F. Schorfheide (2010): "Bayesian Macroeconometrics" to appear in: J. Geweke, G. Koop, and H.K. van Dijk, ed., *Handbook of Bayesian Econometrics.*

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.

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. (2005): "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks," Manuscript, Fed. Reserve Bank Chicago.

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

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

Geweke J. (2005): *Contemporary Bayesian Econometrics and Statistics.* Wiley: Hoboken, NJ.

Geweke, J., Amisano, G. (2011): "Hierarchical Markov Normal Mixture Models with Applications to Financial Asset Returns," *Journal of Applied Econometrics* 26, 1-29.

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.

Hoogerheide, L.F., R. Kleijn, F. Ravazzolo, H.K. Van Dijk, and M. Verbeek (2010): "Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights," *Journal of Forecasting* **29**, 251-269.

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

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

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

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. 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 (2011): "Bayesian Inference in the Time Varying Cointegration Model" forthcoming in *The Journal of Econometrics*, on line publication September, 6, 2011: DOI 10.1016/j.jeconom.2011.07.007.

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

Koop G., 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.

McConnell, M. and G. Perez-Quiros (2000): "Output Fluctuations in the United-States: What Has Changed Since the Early 1980's," *American Economic Review* 90, 1464–1476.

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.

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 regression models," *Journal of the American Statistical Association* **92**, 179–191.

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

Stock, J. H. and M. W. Watson (2002): "Has the Business Cycle Changed and Why?," *NBER Macroeconomics Annual*, Volume 17, MIT Press.

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

Strachan, R. W. and H. K. van Dijk (2010): "Evidence on a Real Business Cycle Model with Neutral and Investment-Specific Technology Shocks using Bayesian Model Averaging," *Tinbergen Institute Working Paper* 10-050/4.

Tanner, M. A. (1993): *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions.* (2nd Edition). (Springer Series in Statistics) New York: Springer.

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

Villani, M. (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.

8 Appendix

Identification of the technology shocks: In this appendix we detail how the two technology shocks are identified by the zero restrictions on $C\Omega^{1/2}$ shown in (2). Let $\widetilde{\Omega}^{1/2}$ denote any unique decomposition of Ω such that $\Omega = \widetilde{\Omega}^{1/2}\widetilde{\Omega}^{1/2\prime}$. This could be the unique Cholesky decomposition into a lower triangular matrix or taken from the square root of the singular value decomposition. As we use the singular value decomposition below we will give this definition for $\widetilde{\Omega}^{1/2}$. Take the singular value decomposition of the $n \times n$ matrix A as $A = USV'$ where $U \in O(n)$, $V \in O(n)$ where $O(n)$ is the orthogonal group of $n \times n$ matrices, and $S = diag(s_1, s_2, \ldots, s_n)$ where $s_i \geq 0$. A square root of A may be obtained by the construction $A^{1/2} = US^{1/2}V'$.

To identify z_t we need to choose an orthogonal matrix $\widetilde{U}$ such that $\Omega^{1/2} = \widetilde{\Omega}^{1/2}\widetilde{U}$ such that $\Omega = \Omega^{1/2}\Omega^{1/2\prime} = \widetilde{\Omega}^{1/2}\widetilde{U}\widetilde{U}'\widetilde{\Omega}^{1/2\prime} = \widetilde{\Omega}^{1/2}\widetilde{\Omega}^{1/2\prime}$.

Let c^1 be $(1 \times n)$ vector of the first row of $C\widetilde{\Omega}^{1/2}$ and let c^2 be the $(1 \times n - 1)$ of the second to last elements in the second row of $C\widetilde{\Omega}^{1/2}$. Next let the projection matrix projecting into the space of c^1 be $P^1 = c^{1\prime}\left(c^1 c^{1\prime}\right)^{-1} c^1$. Similarly, let the projection matrix projecting into the space of c^2 be P^2.

Then $\widetilde{U} = U^1 U^2$ where U^1 is the orthogonal matrix spanning the column space of the singular value decomposition of $P^1 = U^1 SV'$. Further,

$$U^2 = \begin{bmatrix} 1 & 0 \\ 0 & U^{22} \end{bmatrix}$$

where U^{22} is the orthogonal matrix spanning the column space of the singular value decomposition of $P^2 = U^{22}SV'$. We can then define z_t as $z_t = \Omega^{-1/2}\varepsilon_t$.

Computation of break date probabilities: To demonstrate the approach we use to computing the probabilities of the structural break occurring at different dates, first assume we have only three models before the break date and three models after the break date. Index the models by $i \in \{1, 2, 3\}$. Next, we assume we are considering only two break dates and these are indexed by $\tau \in \{0, 1\}$ and denote the models for the pre-break sample by $b = 0$ and for the post-break sample by $b = 1$. Thus the predictive density for model i at date τ for sample b is $p_i^{b,\tau}$. The predictive densities play the role of the marginal likelihoods for the models and the predictive density for a process that switches from model M_i to model M_j at time τ is the product $p_i^\tau = p_i^{0,\tau} p_j^{1,\tau}$.

The probability of a break at time $\tau = 0$ is therefore

$$\frac{\Sigma_{i=1}^3 \Sigma_{j=1}^3 p_i^{0,0} p_j^{1,0}}{\Sigma_{i=1}^3 \Sigma_{j=1}^3 p_i^{0,0} p_j^{1,0} + \Sigma_{i=1}^3 \Sigma_{j=1}^3 p_i^{0,1} p_j^{1,1}} = \frac{\prod_{b=0}^{1} \left(\Sigma_{i=1}^3 p_i^{b,0}\right)}{\sum_{\tau=0}^{1} \prod_{b=0}^{1} \left(\Sigma_{i=1}^3 p_i^{b,\tau}\right)}.$$

Writing the probability this way reveals a simplification that can be used to compute the probabilities of breaks by ignoring the path that the process takes over models. That is, in the above example, we need not estimate each of the eight combinations $p_i^{0,\tau} p_j^{1,\tau}$. Instead we can simply compute the sums $\Sigma_{i=1}^{3} p_i^{b,0}$ which in this simple example reduces the number of computations from 18 to 12. As the size of the model set increases, the gains become more significant. If we have n break periods and M models in each period, the computations reduce from nM^2 to $2nM$. In our application with $\tau = 12$ and $M = 255$, this is a reduction from 780300 to 6120 computations.