STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

STOCHASTIC SEARCH VARIABLE SELECTION IN VECTOR ERROR CORRECTION MODELS WITH AN APPLICATION TO A MODEL OF THE UK MACROECONOMY

BY

MARKUS JOCHMANN, GARY KOOP, ROBERTO LEON-GONZALEZ AND RODNEY W. STRACHAN

NO. 09-19

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

Stochastic Search Variable Selection in Vector Error Correction Models with an Application to a Model of the UK Macroeconomy[*]

Markus Jochmann
University of Strathclyde

Gary Koop[†]
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute
for Policy Studies

Rodney W. Strachan
University of Queensland

September 2009

Abstract

This paper develops methods for Stochastic Search Variable Selection (currently popular with regression and Vector Autoregressive models) for Vector Error Correction models where there are many possible restrictions on the cointegration space. We show how this allows the researcher to begin with a single unrestricted model and either do model selection or model averaging in an automatic and computationally efficient manner. We apply our methods to a large UK macroeconomic model.

Keywords: Bayesian, cointegration, model averaging, model selection, Markov chain Monte Carlo

JEL Codes: C11, C32, C52

---

1   Introduction

Empirical macroeconomic modelling is typically a complicated affair, involving many time series variables. For instance, the UK macroeconomic models considered in this paper are based on the influential model of Garratt, Lee, Pesaran and Shin (2003, 2006), hereafter GLPS. This involves nine variables. Even if one remains in a conventional multivariate time series framework involving Vector autoregressive (VAR) and Vector error correction models (VECMs), there can be hundreds or more modelling choices. Some of these are driven by econometric considerations (e.g. choice of cointegrating rank, lag length, etc.) while others can be driven by economic theory considerations (e.g. GLPS develop a theoretical model of the UK macroeconomy which implies certain restrictions on the cointegrating relationships). In macroeconomic modelling exercises involving structural breaks, time variation in parameters or regime switching such modelling choices proliferate (see, among many others, Primiceri, 2005 and Sims and Zha, 2006).

The challenge facing the empirical macroeconomist is how to navigate through this morass of choices. One common strategy, used by GLPS, is to select a single model. This can be done through sequential hypothesis testing procedures or by choosing the model with the highest value for an information criteria or marginal likelihood. Alternatively, economic theory can be used to guide model choice by giving special consideration to models consistent with a particular economic theory.

A second strategy is to do model averaging and present empirical results which are a weighted average over all models. Common choices for weights are marginal likelihoods or functions of information criteria. Empirical papers adopting this approach include Sala-i-Martin, Doppelhoffer and Miller (2004), Fernandez, Ley and Steel (2001) and Koop, Potter and Strachan (2008). The literature is replete with extensive discussion of the comparative advantages and disadvantages of model averaging and model selection (e.g. Draper, 1995, Leamer, 1978, Raftery, Madigan and Hoeting, 1997). A problem with model averaging is that it can be extremely computationally demanding especially if, as in the current application, Markov Chain Monte Carlo (MCMC) methods are required to carry out inference in each model and the number of models is high.

In this paper, using the modelling framework of GLPS, we use a third strategy which shares similarities with a type of approach which has become increasingly popular in macroeconomics. This strategy can informally be de-

scribed as working with a single very flexible model. The problems with such models are that they risk over-fitting and typically involve a large number of parameters which can be difficult to estimate with any precision. The solution to such problems is prior information. This prior information can be purely subjective or involve some data information (as in the classic Minnesota prior for VAR models, see Doan, Litterman and Sims, 1984). Alternatively the prior can be hierarchical (i.e. can be expressed in terms of a set of unknown parameters which are estimated from the data). Indeed the state equations from state space models such as those used by Primiceri (2005) can be interpreted as hierarchical priors. The mixture innovation models of, e.g., Giordani and Kohn (2008) or Koop, Leon-Gonzalez and Strachan (2009b) can also be interpreted in this manner. Yet another recent hierarchical approach is explored and adapted in this paper. This is based on the Stochastic Search Variable Selection (SSVS) approach to VAR models developed in George, Ni and Sun (2008). The basic idea of this approach can be explained quite simply. Let λ be a parameter (e.g. a coefficient in a VAR). A conventional Bayesian approach would specify a prior for λ such as $\lambda \sim N\left(0, \underline{V}_\lambda\right)$. By setting $\underline{V}_\lambda$ to a large value, a relatively noninformative prior is obtained. Smaller values of $\underline{V}_\lambda$ will shrink the coefficient towards zero, something which has been found to be useful in many empirical applications. Instead of such an approach, the SSVS prior is a hierarchical one, involving a mixture of two normal distributions:

$$\lambda \sim (1-\delta)\, N\left(0, \underline{V}_{\lambda 0}\right) + \delta N\left(0, \underline{V}_{\lambda 1}\right), \qquad (1)$$

where δ is a dummy variable which equals 0 if λ is drawn from the first normal and equals 1 if it is drawn from the second. The prior is hierarchical since δ is treated as an unknown parameter and estimated in a data-based fashion. The SSVS aspect of this prior arises by choosing the first prior variance, $\underline{V}_{\lambda 0}$, to be extremely "small" (so that the parameter is virtually zero) and the second prior variance, $\underline{V}_{\lambda 1}$, to be "large" (implying a relatively noninformative prior for the parameter). The way in which "small" and "large" prior variances can be chosen is discussed in George, Ni and Sun (2008) and below.[1]

---

[1]Formally, the SSVS prior allows for coefficients to be shrunk to "virtually zero" but not "precisely zero" since $\underline{V}_{\lambda 0}$ is not precisely zero. In the remainder of this paper, we will sometimes omit such qualifying phrases like "virtually zero" and simply refer to SSVS as allowing for parameters to be included (if $\delta = 1$) or excluded (if $\delta = 0$) from the model.

SSVS can be interpreted as working with a single very flexible model and then (in a data-based fashion) choosing a prior which ensures over-parameterization problems do not occur. But SSVS can also be used as a way of implementing the two strategies described above. The researcher can do model selection by selecting a single restricted model based on some metric involving δ (e.g. if $\Pr(\delta = 1|Data) > \frac{1}{2}$ then the explanatory variable corresponding to λ is included in the model, else it is excluded). Bayesian model averaging (BMA) can be done by using an unrestricted model with SSVS prior. This will implicitly average over models which include and exclude the explanatory variable associated with λ. The weights in this average will be $\Pr(\delta = 1|Data)$ and $\Pr(\delta = 0|Data)$, respectively.

The SSVS approach is very attractive for VAR modelling since VARs can have a large number of coefficients and many of them have no explanatory power. For instance, the trivariate VAR(4) used in Jochmann, Koop and Strachan (2009) has 36 VAR coefficients and SSVS picks out only 10 of these as having important explanatory power. In such a situation, the number of hypothesis tests required in a sequential testing procedure could lead the researcher to serious worries about pre-test problems. Alternatively, the researcher who wants to select the single restricted VAR with highest marginal likelihood or highest value for an information criteria would face the serious problem of estimating at least 2^{36} models (i.e. if each restricted model is defined according to whether each of the 36 coefficients is included or excluded, the result is 2^{36} models). Regardless of whether SSVS is used for model selection or model averaging, it seems an attractive approach to VAR modelling which will shrink unimportant coefficients to zero and, thus, yield a much more parsimonious model. In essence, the SSVS prior can pick out restrictions automatically so that the researcher can begin with a flexible unrestricted model and let SSVS pick out the appropriate restrictions.

The major theoretical econometric contribution of this paper is to extend the SSVS approach, which is developed for multivariate normal linear models such as the VAR, to the nonlinear VECM. As we shall see, some subtle econometric issues arise in this regard. A related theoretical contribution is to extend the SSVS approach to restrictions other than ones like (1) where each coefficient is either included or excluded. In particular, macroeconomists such as GLPS are often interested in overidentifying restrictions involving several coefficients. We develop an SSVS approach which can pick out such restrictions automatically.

The major empirical contribution is to use our methods in an important

macroeconomic application involving a high dimensional model. The GLPS model fits the bill perfectly. It involves nine dependent variables, uncertainty over many modelling choices (e.g. cointegrating rank, lag length and treatment of deterministic terms) and many theoretical restrictions motivated by macroeconomic theory. We find SSVS to be computationally efficient and to lead to quite parsimonious modelling. We find support for interest rate parity and Fisher inflation parity conditions which are imposed by GLPS. However, we find less support for purchasing power parity and none at all for a restriction motivated by neoclassical growth models. These latter two restrictions are imposed by GLPS. Nevertheless, when we turn to impulse response analysis we find economically-sensible results which are not too different than those of GLPS.

2 Models

A Bayesian model is defined by a likelihood function and a prior. We deal with each of these in turn in this section.

2.1 Likelihoods

The VECM for an n-dimensional vector, y_t, is written as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \mu d_t + \varepsilon_t \tag{2}$$

for $t = 1, .., T$ where the $n \times n$ matrix Π is of rank $r \leq n$ and d_t denotes deterministic terms. ε_t is i.i.d. $N(0, \Sigma)$. The framework described in (2) defines a set of models which differ in the number of cointegrating vectors (r), lag length (p) and the specification of deterministic terms. In addition, the researcher may wish to entertain various over-identifying restrictions on the cointegrating vectors.

The unrestricted VECM used in this paper, written in matrix form is:

$$Y = X\beta\alpha + W\Gamma + E, \tag{3}$$

where Y is $T \times n$ with t^{th} row given by $\Delta y_t'$, X is $T \times (n+2)$ with t^{th} row given by $(1, t, y_{t-1}')$, β is $(n+2) \times r_{\max}$, α is $r_{\max} \times n$, W is $T \times [n(p_{\max} - 1) + 2]$ with t^{th} row given by $(1, t, \Delta y_{t-1}', .., \Delta y_{t-p+1}')$, Γ is $[n(p_{\max} - 1) + 2] \times n$ and E is

$T \times n$ with $vec(E) \sim N(0, \Sigma \otimes I)$. $r_{\max}$ and $p_{\max}$ are the maximum number of cointegrating relationships and lag length, respectively, that the researcher is willing to entertain. Note that, with regards to deterministic terms, a deterministic trend in the cointegrating residuals has very different implications than a deterministic trend in the levels of the series. Accordingly, following Johansen (1995, Section 5.7), we potentially allow for deterministic terms in the cointegrating residual and directly in the conditional mean of Δy_t.

GLPS work with additional restrictions motivated by macroeconomic theory. A detailed explanation of this theory is provided in GLPS. Here we only sketch the details. The nine variables GLPS work with are the log of the oil price (p_t^o), the log of the effective exchange rate (e_t), the domestic and foreign nominal interest rates $(i_t$ and $i_t^*)$, the inflation rate (Δp_t), the logs of domestic and foreign real per capita GDP $(q_t$ and $q_t^*)$, the log of the ratio of domestic to foreign prices $(p_t - p_t^*)$ and the log of the ratio of real high powered money to real GDP per capita $(h_t - q)$. Thus, $y_t = (p_t^o, e_t, i_t^*, i_t, \Delta p_t, q_t, p_t - p_t^*, h_t - q_t, q_t^*)'$. Exact definitions and sources are given in the Data Appendix.

The model GLPS call their core model assumes five cointegrating relationships of which the following four place restrictions on β. These are:

$$
\begin{aligned}
p_t - p_t^* - e_t &= b_1 + e_{1t}, \\
i_t - i_t^* &= b_2 + e_{2t}, \\
q_t - q_t^* &= b_3 + e_{3t}, \\
i_t - \Delta p_t &= b_4 + e_{5t},
\end{aligned}
\tag{4}
$$

where b_i are intercepts in the cointegrating relationships and e_{it} denote mean zero stationary errors.[2]

GLPS derive and explain these cointegrating relationships in detail. Briefly, the first restriction is motivated by the purchasing power parity relation-

---

[2]GLPS use a long run solvency requirement to motivate a fifth cointegrating relationship. However, it implies a cointegrating vector dependent on unknown parameters (i.e. it only implies the exclusion restriction that a cointegrating relationship must exist between two specific dependent variables and does not specify the exact cointegrating vector). When multiple cointegrating relationships exist, imposing such exclusion restrictions does not restrict the cointegrating space because by linearly combining cointegrating vectors one can come up with a new cointegrating vector which incorporates the long run solvency restriction but does not restrict the coefficients on the other dependent variables. Dealing with such issues is possible using an extended version of our approach, but substantially complicates the analysis and, accordingly, we do not consider this fifth restriction in this paper.

ship. The second from an interest rate parity condition. The third by neo-classical growth models for the UK and the rest of the world. The fourth is based on Fisher inflation parity arguments. These equations define over-identifying restrictions and we consider variants of our models which impose some or all of them.

In sum, we have so far developed a very general unrestricted model. However, there are many restricted models of interest which differ in choice of $p \leq p_{\max}, r \leq r_{\max}$ and the number of over-identifying restrictions imposed (i.e. restrictions on the cointegration space). Furthermore, setting individual coefficients in α and/or Γ (if empirically warranted) would lead to the desirable goal of a more parsimonious model. Note, in particular, that (3) allows for deterministic terms (both intercept and time trend) to enter both the cointegrating residual and Δy_t. Such a specification is typically too flexible in macroeconomic empirical work and, thus, restricted models which zero out some of these deterministic terms are potentially of interest.

In most applications, this will lead to a huge model space. For instance, even if we ignore restrictions on α and/or Γ, if we set $r_{\max} = 5$, $p_{\max} = 2$ and consider eighteen restrictions on β,[3] we would have hundreds of models to work with. This is the approximate number of models used in the conventional Bayesian cointegration analysis of Koop, Potter and Strachan (2008). Given that the methodology of Koop, Potter and Strachan (2008) required the running of an MCMC algorithm in each model, the computational requirements of that paper were quite substantial. In the present application, α is a 5×9 matrix and Γ an 11×9 matrix and, thus, in total they contain 144 parameters. If we are to extend our model space by considering models which allow for each individual element of α and/or Γ to be zero, then the number of models is further multiplied by 2^{144}. Clearly, doing a conventional Bayesian model selection or model averaging exercise which involves estimating each of more than 100×2^{144} models is computationally impossible. This consideration helps motivate our SSVS approach which, as we shall see, allows the researcher to work with the one model given in (3) and select appropriate restrictions in an automatic manner.

---

[3]These restrictions will be explained below. Basically, they arise from imposing the restrictions in (4) singly or in all possible combinations and restricting the deterministic terms in the cointegrating residuals.

2.2 The SSVS Prior

In this paper, we will extend the SSVS prior for VARs of George, Ni and Sun (2008) to VECM models. The new econometric issues involve β and α. However, before describing our treatment of these parameters we will describe the prior for Γ and Σ. We do this since our treatment of Γ and Σ is exactly the same as George, Ni and Sun and, thus, we can introduce basic SSVS ideas in a familiar framework.

2.2.1 Prior for Γ and Σ

Note that, conditional on β and α, the VECM can be written as:

$$\widetilde{Y} = W\Gamma + E, \tag{5}$$

where $\widetilde{Y} = Y - X\beta\alpha$. This takes the form of a VAR and the SSVS methods for VARs of George, Ni and Sun (2008) can be applied directly in the context of an MCMC algorithm (except using $\widetilde{Y}$ as the vector of dependent variables instead of Y).

SSVS can be interpreted as defining a hierarchical prior for Γ and Σ. Each element in the hierarchy is a mixture of two normals, one with a small variance (implying the coefficient is not in the model) and one with a large variance (implying the coefficient is included in the model). The SSVS prior for $\gamma = vec(\Gamma)$ can be written as:

$$\gamma|\delta \sim N(0, D), \tag{6}$$

where δ is a vector of unknown parameters with typical element $\delta_j \in \{0, 1\}$, and D is a diagonal matrix with the jth element given by κ_j^2 where

$$\kappa_j^2 = \begin{cases} \underline{\kappa}_{0j}^2 & \text{if } \delta_j = 0, \\ \underline{\kappa}_{1j}^2 & \text{if } \delta_j = 1. \end{cases} \tag{7}$$

Note that this prior implies a mixture of two normals:

$$\gamma_j|\delta_j \sim (1 - \delta_j) N\left(0, \underline{\kappa}_{0j}^2\right) + \delta_j N\left(0, \underline{\kappa}_{1j}^2\right), \tag{8}$$

where γ_j is the jth element of γ. δ is treated as a vector of unknown parameters and is estimated in a data based fashion. If $\delta_i = 0$, then the prior for γ_j has such a small variance that the posterior for γ_j is concentrated near

zero. In this sense, we can say that, if $\delta_j = 0$, then the jth explanatory variable is excluded from the model (formally its coefficient is shrunk to be very near zero). If $\delta_j = 1$, then the jth explanatory variable is included and γ_j is estimated using a relatively noninformative prior. As discussed in the introduction, priors of this form can be used for model selection by choosing a single model which includes only those explanatory variables for which $\Pr(\delta_i > 1|Data) > c$ for some threshold c (e.g. $c = \frac{1}{2}$). Alternatively, Bayesian model averaging can be done by running the MCMC algorithm for the unrestricted model and simply averaging any feature of interest (e.g. an impulse response function or a predictive moment) over all MCMC draws. Since the MCMC algorithm provides draws of δ (as well as all other parameters), this strategy amounts to averaging over models (i.e. in the sense that different drawn values for δ define different models). Crucially, though, doing model averaging in this way requires only that the researcher run a single MCMC algorithm for the unrestricted model (as opposed running an MCMC algorithm in each restricted model).

We use what George, Sun and Ni (2008) call the "default semi-automatic approach" to selecting the prior hyperparameters $\underline{\kappa}_{0j}^2$ and $\underline{\kappa}_{1j}^2$ and the reader is referred to their paper for additional justification for this approach. Basically, $\underline{\kappa}_{0j}^2$ should be selected so that γ_j is essentially zero and $\underline{\kappa}_{1j}^2$ should be selected so that γ_j is empirically substantive. The default semi-automatic approach involves choosing $\underline{\kappa}_{0j}^2 = c_0 \widehat{var}(\gamma_j)$ and $\underline{\kappa}_{1j}^2 = c_1 \widehat{var}(\gamma_j)$ where $\widehat{var}(\gamma_j)$ is an estimate of the posterior variance of the coefficient. In our empirical work, this is based on a preliminary MCMC run using a non-informative prior, although other estimates (e.g. based on maximum likelihood estimation) are possible. The pre-selected constants c_0 and c_1 must have $c_0 << c_1$ and we set $c_0 = \frac{1}{10}$ and $c_1 = 10$.[4]

For δ, the SSVS prior posits that each element has a Bernoulli form (independent of the other elements of δ) and, hence we have

$$\begin{aligned} \Pr(\delta_j = 1) &= \underline{q}_j, \\ \Pr(\delta_j = 0) &= 1 - \underline{q}_j. \end{aligned} \tag{9}$$

We set $\underline{q}_j = \frac{1}{2}$ for all j. This is a natural default choice, implying each

---

[4]Note that some related approaches set the first element in the mixture to be a spike at zero. We do not adopt this approach since, as sample size increases, it leads to the inclusion of all coefficients. See Section 3.4 of Chipman, George and McCulloch (2001) for further discussion.

coefficient is *a priori* equally likely to be included as excluded.

For the error covariance matrix, we begin by decomposing it as:

$$\Sigma^{-1} = \Psi\Psi', \tag{10}$$

where Ψ is upper-triangular. The SSVS prior involves using a standard Gamma prior for the square of each of the diagonal elements of Ψ and the SSVS mixture of normals prior for each element above the diagonal. Note that this implies that the diagonal elements of Ψ are always included in the model, ensuring a positive definite error covariance matrix. This form for the prior also greatly simplifies posterior computation. Precise details are provided in the next paragraph.

Let the non-zero elements of Ψ be labelled as ψ_{ij} and define $\psi = (\psi_{11}, .., \psi_{nn})'$, $\eta_j = (\psi_{1j}, .., \psi_{j-1,j})'$ and $\eta = (\eta_2', .., \eta_n')'$. For the diagonal elements, we assume prior independence with

$$\psi_{jj}^2 \sim G\left(\underline{a}_j, \underline{b}_j\right), \tag{11}$$

where $G\left(\underline{a}_j, \underline{b}_j\right)$ denotes the Gamma distribution with mean $\frac{\underline{a}_j}{\underline{b}_j}$ and variance $\frac{\underline{a}_j}{\underline{b}_j^2}$. We specify $\underline{a}_j = 0.1$ and $\underline{b}_j = 0.001$ which are relatively noninformative choices, but center the prior at 100.0. The latter is a sensible value given the scale of the variables and the fact that ψ_{jj}^2 is the inverse of the error variance in equation j.

The hierarchical prior for η takes the same mixture of normals form as γ. In particular, the SSVS prior has

$$\eta_j | \omega_j \sim N\left(0, F_j\right), \tag{12}$$

where $\omega_j = (\omega_{1j}, .., \omega_{j-1,j})'$ is a vector of unknown parameters with typical element $\omega_{ij} \in \{0, 1\}$, and $F_j = diag\left(\xi_{1j}^2, .., \xi_{j-1,j}^2\right)$ where

$$\xi_{ij}^2 = \begin{cases} \underline{\xi}_{0ij}^2 & \text{if } \omega_{ij} = 0, \\ \underline{\xi}_{1ij}^2 & \text{if } \omega_{ij} = 1, \end{cases} \tag{13}$$

for $j = 2, .., n$ and $i = 1, .., j-1$. Note that this prior implies a mixture of two Normals for each off-diagonal element of Ψ:

$$\psi_{ij} | \omega_{ij} \sim (1 - \omega_{ij}) N\left(0, \underline{\xi}_{0ij}^2\right) + \omega_{ij} N\left(0, \underline{\xi}_{1ij}^2\right). \tag{14}$$

10

As we did with $\underline{\kappa}_{0j}^2$ and $\underline{\kappa}_{1j}^2$ (the hyperparameters in the SSVS prior for Γ), we use a semi-automatic default approach to selecting $\underline{\xi}_{0ij}^2$ and $\underline{\xi}_{1ij}^2$. That is, we set $\underline{\xi}_{0ij}^2 = c_0\widehat{var}\left(\psi_{ij}\right)$ and $\underline{\xi}_{1ij}^2 = c_1\widehat{var}\left(\psi_{ij}\right)$, where $\widehat{var}\left(\psi_{ij}\right)$ is based on an estimate of the variance of the appropriate off-diagonal element of Σ. As with Γ, this estimate is obtained from a preliminary MCMC run using a non-informative prior. The pre-selected constants c_0 and c_1 are set (as before) to be $c_0 = \frac{1}{10}$ and $c_1 = 10$.

For $\omega = (\omega_2', .., \omega_n')'$, the SSVS prior posits that each element has a Bernoulli form (independent of the other elements of ω) and, hence, we have

$$
\begin{aligned}
\Pr\left(\omega_{ij} = 1\right) &= \underline{q}_{ij}, \\
\Pr\left(\omega_{ij} = 0\right) &= 1 - \underline{q}_{ij}.
\end{aligned}
\tag{15}
$$

We make the default choice of $\underline{q}_{ij} = \frac{1}{2}$ for all i and j.

The preceding material describes the basic ideas underlying SSVS and the details of how we implement it for Γ and Σ. SSVS has been found to be a very useful technique in finding parsimonious restricted versions of VARs. Posterior simulation methods are described in George, Ni and Sun (2008) and in Appendix A. Suffice it to note here that we use their algorithm to provide posterior draws of Γ, Σ, δ and ω (ω is the vector containing all the ω_j) conditional on α and β.

2.2.2 Prior for α and β

SSVS has never been extended to VECMs. The fact that $\beta\alpha$ enters in product form and the fact that (without further restrictions) the VECM only identifies the cointegrating space precludes the direct use of SSVS ideas for VARs. Before describing how SSVS can be implemented in a manner which overcomes these problems, we must digress and explain some basic issues which arise in Bayesian analyses of VECMs. An identification problem arises in the VECM since $\Pi = \beta\alpha$ and $\Pi = \beta CC^{-1}\alpha$ are identical for any nonsingular C. In early work, a common practice was to impose a linear normalization such as $\beta = \begin{pmatrix} I_r \\ \beta_0 \end{pmatrix}$. However, a literature has recently emerged which argues that it is only the cointegrating space which is identified (see Strachan, 2003, Strachan and Inder, 2004 and Villani, 2005, 2006) and this should be the focus of interest (rather than a particular identified parameter such as β_0). For instance, Strachan and Inder (2004) show how the use of linear

identifying restrictions places a restriction on the estimable region of the cointegrating space. Furthermore, a flat and apparently "noninformative" prior on β_0 in the linear normalization strongly favors regions of the cointegration space near where the linear normalization is invalid. Hence, the linear normalization is used under the assumption that it is valid while at the same time the prior puts weight near the region where the normalization is likely to be invalid. These considerations suggest that an alternative identification restriction is required to form the basis for prior elicitation on the cointegration space. As shown in Strachan and Inder (2004) an identification restriction which does not suffer from the drawbacks of the linear and similar normalizations is:

$$\beta'\beta = I \tag{16}$$

and we will impose this semi-orthogonality restriction on β throughout the remainder of this paper.

Formally, let $sp(\beta)$ denote the cointegration space (i.e. the space spanned by the columns of β) which is an element of the Grassmann manifold. The semi-orthogonality restriction, (16), restricts the matrix of cointegrating vectors to the Stiefel manifold. These spaces are compact and, hence, a Uniform distribution over them is proper (the integrating constant is given in Strachan and Inder, 2004). Strachan and Inder (2004) show how a sensible noninformative prior for β which does not restrict the cointegrating space is simply equal to this integrating constant with (16) imposed. This can be used as a noninformative prior on the cointegrating space.

Computation is complicated by the fact that β is semi-orthogonal and conventional algorithms such as the Gibbs sampler of Geweke (1996) cannot be used. The parameter-augmented Gibbs sampler developed in Koop, Leon-Gonzalez and Strachan (2009a) overcomes this complication by introducing a non-identified $r \times r$ symmetric positive definite matrix U with the property:

$$\Pi = \beta\alpha = \beta U U^{-1}\alpha \equiv \beta^*\alpha^* \tag{17}$$

where $\alpha^* = U^{-1}\alpha$ and $\beta^* = \beta U$. The introduction of the non-identified U facilitates posterior computation because, under sensible priors (including the SSVS prior we adopt), the posterior conditional distributions of α^* and β^* in the MCMC algorithm are normal. The fact that U is not identified causes no problem and, in fact, introducing non-identified parameters in particular ways is a common strategy for improving the efficiency of MCMC algorithms in many contexts (see, e.g., Liu and Wu, 1999). For proofs of all these

statements, further discussion and exact details of the MCMC algorithm, see Koop, Leon-Gonzalez and Strachan (2009a).

Both for its own sake and since it will form a crucial part of our SSVS prior for VECMs, we also require an informative prior for the cointegrating space. Strachan and Inder (2004) and Koop, Leon-Gonzalez and Strachan (2009a) develop and motivate such a prior. Here we briefly summarize the elements of this strategy and refer the reader to these earlier papers for additional details. Restrictions on the cointegrating space can be always be defined through a semi-orthogonal $(n + 2) \times s$ matrix[5] H with $r \leq s \leq n + 2$ and $H_\perp$ being its orthogonal complement which is also a semi-orthogonal matrix. Koop, Leon-Gonzalez and Strachan (2009a) give some examples of H and in Appendix B we will show how different combinations of the restrictions in (4) can be written in this manner. Consider a prior for the cointegrating space defined through:

$$vec\,(\beta^*) \sim N\left[0, I_r \otimes (HH' + \underline{\tau}\,H_\perp H_\perp')\right]. \tag{18}$$

As shown in Koop, Leon-Gonzalez and Strachan (2009a) this is a sensible informative prior for the cointegrating space in that the prior for the cointegrating space will be centered over $sp\,(H)$. Furthermore, the dispersion of the prior is controlled by the scalar $\underline{\tau} \in [0, 1]$ with $\underline{\tau} = 0$ dogmatically imposing the restrictions expressed by H on the cointegrating space and $\underline{\tau} = 1$ leading to the noninformative prior discussed previously.[6]

To see how this prior can be used to develop an SSVS prior for the cointegrating space, remember that the conventional implementation of SSVS implies a prior which is the mixture of two normals, one with a "small" variance and the other with a "large" variance. Thus, a parametric restriction is either imposed (approximately) or not imposed. For the cointegrating space, an SSVS prior could also be a mixture of two distributions. The first distribution would (approximately) restrict the cointegrating space to lie in the space spanned by H and the second would be a noninformative prior which would allow the cointegrating space to be estimated in an unconstrained manner. A consideration of the prior defined by (18) suggests how this can be done. That is, if we introduce a discrete random variable with two points

---

[5]Remember that, by allowing for an intercept and trend in the cointegrating residual, each column of β and, hence, H will have $n + 2$ elements.

[6]Note that one should not choose $\underline{\tau} = 0$, since this would imply a singular prior covariance matrix in the typical case where $s < n + 2$.

of support, one which sets $\underline{\tau}$ to a small value, the other which sets $\underline{\tau} = 1$, then we obtain a prior which can either (approximately) restrict the cointegration space (if $\underline{\tau}$ is small) or not (if $\underline{\tau} = 1$).

The previous discussion informally motivates how an SSVS prior for the cointegrating space can be developed when there is a single restriction (defined by H) involved. Extensions to multiple restrictions on the cointegrating space (e.g. our empirical work involves four over-identifying restrictions given in equation 4 which are imposed individually and jointly plus we consider restricting the deterministic terms) can be done by mixing over more distributions. Formally, in our empirical work, we use an SSVS prior for the cointegrating space which takes the form of (18) with the additional definition that $H = H_j$ if $\phi = j$ for $j = 1, .., 18$ and $\underline{\tau} = 0.05$. Precise definitions of H_1 through H_{18} are provided in Appendix B. Briefly $H_1 = I$ yields the non-informative prior discussed previously (and setting $\underline{\tau} = 0.05$ is irrelevant). H_2 restricts the deterministic trend in the cointegrating relationship to be zero and H_3 additionally restricts the intercept to be zero. H_4 through H_7 individually impose the four restrictions on the cointegrating space given in (4). H_8 through H_{13} impose all combinations of two of the four restrictions. H_{14} through H_{17} impose all combinations of three of the four restrictions. H_{18} imposes all of the restrictions. $\phi \in \{1, 2, \ldots, 18\}$ is a discrete random variable which selects which restriction applies. If $\phi = 1$ is chosen, then none of the restrictions are selected. Since ϕ is an unknown parameter, it requires a prior:

$$\Pr(\phi = i) = \underline{p}_{\phi i}, \quad \sum \underline{p}_{\phi i} = 1, \quad i = 1, \ldots, 18. \tag{19}$$

In our empirical work, we use the noninformative choice $\underline{p}_{\phi i} = \frac{1}{18}$ for all i.

Finally, we require a prior for α^*. For this we use a standard SSVS prior of the sort used for VARs. In particular, we use

$$a_i^* \sim N\left(0, \nu_i^2\right), \tag{20}$$

where a_i^* is the ith (for $i = 1, .., nr_{\max}$) element of $vec\left(\alpha^*\right)$. For motivation of this prior (in terms of the shrinkage it achieves), see Koop, Leon-Gonzalez and Strachan (2009a). However, for present purposes, the more important issue is that it can be adapted to do SSVS by setting:

$$\nu_i^2 = \begin{cases} \underline{\nu}_{0i}^2 & \text{if } \rho_i = 0, \\ \underline{\nu}_{1i}^2 & \text{if } \rho_i = 1, \end{cases} \tag{21}$$

14

where $\underline{\nu}_{0i}^2$ and $\underline{\nu}_{1i}^2$ are set to big and small values (comparable to $\underline{\kappa}_{0j}^2$ and $\underline{\kappa}_{1j}^2$). ρ_i is a dummy variable (comparable to δ_i) which selects whether a_i^* is (approximately) zero or not. Its prior is given by:

$$\Pr(\rho_i = 1) = 1 - \Pr(\rho_i = 0) = \underline{p}_a. \tag{22}$$

In our empirical work, we set $\underline{p}_a = \frac{1}{2}$, $\underline{\nu}_{0i}^2 = c_0 \widehat{var}(a_i^*)$ and $\underline{\nu}_{1i} = c_1 \widehat{var}(a_i^*)$, where $\widehat{var}(a_i^*)$ is an estimate of the variance of a_i^* obtained from a preliminary MCMC run using a non-informative prior.[7] The pre-selected constants c_0 and c_1 are set (as before) to be $c_0 = \frac{1}{10}$ and $c_1 = 10$.

2.3 Posterior Simulation

The previous sub-sections outlined the likelihood and prior we use in this paper. Complete details of the MCMC algorithm used for posterior simulation are provided in Appendix A. Here we note that the algorithm combines blocks from the MCMC algorithm for VECMs described in Koop, Leon-Gonzalez and Strachan (2009a) and the MCMC algorithm for SSVS in VARs of George, Ni and Sun (2008). That is, conditional on the SSVS vectors of dummy variables $\delta, \omega, \phi, \rho$ (where ρ is a vector containing all the ρ_i defined previously) we have a VECM with a particular prior and the algorithm of Koop, Leon-Gonzalez and Strachan (2009a) can be used. But $\delta, \omega, \phi, \rho$ can be drawn using the methods of George, Ni and Sun (2008), with minor alterations in the case of ϕ. Exact formulae are given in Appendix A.

3 Empirical Work

Our data set is an updated version of the one constructed by GLPS and runs from 1965Q1 through 2008Q1. Remember that $y_t = (p_t^o, e_t, i_t^*, i_t, \Delta p_t, q_t, p_t - p_t^*, h_t - q_t, q_t^*)'$ and these variables have been described in Section 2 (with precise definitions and data sources given in Appendix C). Motivated by GLPS and preliminary experimentation with the data we select $r_{\max} = 5$ and $p_{\max} = 2$ as being reasonable values for the maximum possible number of cointegrating relationships and maximum lag length. All other modelling details, including prior hyperparameter values are described in the previous section.

---

[7]To be precise, $\widehat{var}(a_i^*)$ is the average of the variances from all elements of α^* that belong to the same equation as a_i^*.

We divide our empirical work into two parts. First, using our SSVS approach, we present evidence on which models are supported by the data. Second, we present impulse responses using our SSVS approach (which we label "SSVS for Everything") and two important special cases (which we label "Best Model" and "GLPS Model", respectively). The Best Model approach uses an initial SSVS run to select the single best model and then we estimate that model. We choose this best model as follows: For Γ and η we include all coefficients with posterior inclusion probabilities greater than $\frac{1}{2}$ (all other coefficients are set to zero). By "posterior inclusion probability" we mean $\Pr(\delta_i|Data)$ (when referring to the ith element of Γ). For the cointegrating space, we choose the value of ϕ which has highest posterior inclusion probability. For α^* we retain the SSVS prior.[8] The second special case involves using our SSVS approach, but imposing the economic theory derived in GLPS on the model. We do this by using a dogmatic prior which attaches all prior weight to H_{18} [which imposes all four of the restrictions in (4)]. Formally, this amounts to setting $\Pr(\phi = 18) = 1$.

3.1   Which Models are Supported by the Data?

Using SSVS for Everything we can calculate the probability associated with the various restrictions on the cointegration space (i.e. $\phi = 1, .., 18$), the probability associated with each cointegrating rank and the posterior inclusion probability of any coefficient.

Table 1 presents the posterior distribution of ϕ. It can be seen that there is no one restriction that is predominant, but rather the posterior probability is spread over many different values for ϕ. Imposing some restrictions clearly is preferred by the data since $\Pr(\phi = 1|Data) = 0$ which indicates no support for the noninformative prior. However, there is also little support for all of the restrictions holding since $\Pr(\phi = 18|Data)$ is also near zero. The most support is for $\phi = 12$, which imposes interest rate parity and Fisher inflation parity. However, there is also substantial support for $\phi = 7$ which imposes only Fisher inflation parity. The next most popular choice is $\phi = 15$

---

[8]We do this since our initial run of SSVS averages over different values of ϕ, not just a single best choice for ϕ. Given the fact that α^* and β^* enter in product form and ϕ determines β^*, SSVS treatment of α^* for fixed ϕ and SSVS treatment of α^* for random ϕ are conceptually and empirically very different. Hence, it would be questionable to choose restrictions on α^* using an initial SSVS run treating ϕ as random and then use these restrictions in a model where ϕ is fixed.

which adds to $\phi = 12$ the additional restriction that purchasing power parity holds. There is never support for any value of ϕ which involves imposing the restriction inspired by neoclassical growth models for the UK and the rest of the world.

Much of the support for the various restrictions occurs jointly, e.g., the joint probability of purchasing power parity and uncovered interest rate parity is 0.062. To obtain the marginal probability of an individual restriction, therefore, we need to sum over all joint spaces. Doing this we obtain: a marginal probability of purchasing power parity of 33.5%; a marginal probability of uncovered interest rate parity of 59.6%; a marginal probability of the neoclassical growth restriction of 4.4%; and a marginal probability of Fisher inflation parity of 86.4%.

When using these models for economic policy (e.g. to do impulse response analysis or forecasting), SSVS for Everything will average over these different choices for ϕ with weights given in Table 1.

Table 1: Posterior probabilities of ϕ

ϕ	Probability	ϕ	Probability
1	0.000	10	0.114
2	0.000	11	0.003
3	0.000	12	0.320
4	0.001	13	0.005
5	0.061	14	0.008
6	0.000	15	0.126
7	0.271	16	0.012
8	0.062	17	0.004
9	0.000	18	0.012

In terms of the cointegrating rank, note that $\Pi = \beta^* \alpha^*$ and our posterior simulation algorithm provides draws of β^* and α^*. The cointegrating rank is the rank of Π. Accordingly, we can shed light on the posterior of the cointegrating rank using these draws of Π. Formally, our posterior for the cointegrating rank is the posterior of the number of singular values of Π which are greater than 0.05. This is presented in Table 2.

Table 2 shows that there is strong support for a cointegrating rank of either four or five. This is consistent with the non-Bayesian analysis of GLPS who (with a shorter data set) find a cointegrating rank of five.

17

Table 2: Posterior of cointegration rank

Rank	Probability
1	0.000
2	0.000
3	0.001
4	0.368
5	0.631

Next, we present evidence that SSVS is automatically leading to a very parsimonious model. First we present evidence relating to the 5×9 matrix α. Let ρ be the vector of dummy variables with typical element ρ_i (see equation 21). As one measure of parsimony we can calculate the posterior of ρ or the posterior of all the elements of ρ corresponding to a single equation. The posterior mean and variance of such features is what we include in Table 3 and label the "Number of included α". Each row of α could potentially have up to 5 non-zero coefficients. Table 3 shows that, in every equation, SSVS sets most of these to zero. The posterior mean number of included α in every equation is less than one (although the standard deviations are moderately large). The final row of Table 3 shows how, of the 45 possible non-zero coefficients in α, SSVS sets approximately 40 to zero.

Table 3: Number of included α

Equation	Mean	St..dev.
p^o	0.524	0.689
e	0.473	0.658
i^*	0.493	0.666
i	0.622	0.744
Δp_t	0.887	0.879
q	0.476	0.653
$p - p^*$	0.458	0.641
$h - q$	0.472	0.658
q^*	0.457	0.640
All equations	4.862	2.097

Table 4 presents posterior inclusion probabilities for each coefficient in the 11×9 matrix Γ. Of these 99 coefficients, only 13 have posterior inclusion probabilities of greater than $\frac{1}{2}$.[9] Often it is the case that the own lag of the dependent variable is the only explanatory variable selected, but there are a few exceptions to this pattern. This reinforces our story that SSVS is an effective method of achieving parsimony in multivariate time series models such as the VECM.

Table 4: Posterior inclusion probabilities for Γ

	p^o	e	i^*	i	Δp_t	q	$p - p^*$	$h - q$	q^*
$const$	0.125	0.096	0.102	0.223	0.263	0.097	0.099	0.105	0.097
$time$	0.114	0.100	0.098	0.127	0.182	0.092	0.093	0.090	0.095
p^o	0.958	0.255	0.273	0.335	0.608	0.223	0.273	0.218	0.225
e	0.263	0.945	0.298	0.248	0.340	0.282	0.602	0.260	0.227
i^*	0.348	0.285	1.000	0.847	0.229	0.266	0.246	0.228	0.334
i	0.427	0.400	0.381	0.731	0.266	0.278	0.223	0.223	0.396
Δp_t	0.268	0.243	0.425	0.239	0.355	0.540	0.246	0.375	0.247
q	0.337	0.291	0.323	0.559	0.324	0.330	0.231	0.226	0.228
$p - p^*$	0.450	0.339	0.596	0.224	0.967	0.394	0.971	0.321	0.227
$h - q$	0.241	0.499	0.234	0.354	0.334	0.264	0.412	0.951	0.288
q^*	0.442	0.225	0.871	0.284	0.250	0.216	0.443	0.301	0.744

---

[9]The reason that the inclusion probabilities for the intercept are so low is that we are also including an intercept in the cointegrating residual.

Table 5 also reveals a similar pattern of parsimony. It presents posterior inclusion probabilities for Ψ (which contains the off-diagonal elements of the error covariance matrix). For most of these error covariances the posterior inclusion probability is much less than $\frac{1}{2}$.

Table 5: Posterior inclusion probabilities for Ψ

	p^o	e	i^*	i	Δp_t	q	$p-p^*$	$h-q$	q^*
p^o		0.268	0.738	0.454	0.236	0.483	0.998	0.428	0.250
e			0.250	0.997	0.250	0.263	0.295	0.333	0.334
i^*				0.967	0.342	0.284	0.567	0.244	0.505
i					0.315	0.295	0.261	0.223	0.307
Δp_t						0.476	0.998	0.688	0.242
q							0.738	1.000	0.994
$p-p^*$								0.629	0.296
$h-q$									0.262
q^*									

3.2   Impulse Response Analysis

We have seen how SSVS does successfully allow us to work with a very flexible unrestricted model and allow for a more parsimonious restricted model to be picked out automatically. We stress that the alternative strategy of evaluating every restricted model with a goal to either selecting a single one or doing model averaging is computationally infeasible when facing the number of models that we are dealing with here. Now we present evidence that SSVS is picking out economically-sensible empirical models through a consideration of impulse responses.

GLPS use a particular set of restrictions which allows them to identify a monetary policy shock and an oil shock. We adopt the same identifying scheme and refer the reader to GLPS for precise details. Point estimates (posterior means) of the impulse responses to a monetary shock and oil price shock for our three cases (SSVS for Everything, Best Model and GLPS model) and plotted in Figures 1 and 2, respectively. It can be seen that the three approaches are giving very similar impulse responses. Furthermore, these impulse responses are all of reasonable shape and magnitude. Qualitatively, they are similar to those given in Figures 2 and 4 of Garratt et al (2003) despite the fact that we are using a substantially longer data span. The un-

restricted VECM produced quite different impulse responses. We find this reassuring in the sense that it is providing evidence that SSVS is successfully choosing a parsimonious and economically-sensible model, but in an automatic and computationally efficient manner. We are not carrying out a long series of sequential hypothesis tests nor are we doing Bayesian model averaging in the conventional and computationally demanding manner of Koop, Potter and Strachan (2008).

Figure 1: Responses to a Monetary Shock. Solid line = SSVS for
Everything, Dots = Best Model, Dashes = GLPS model
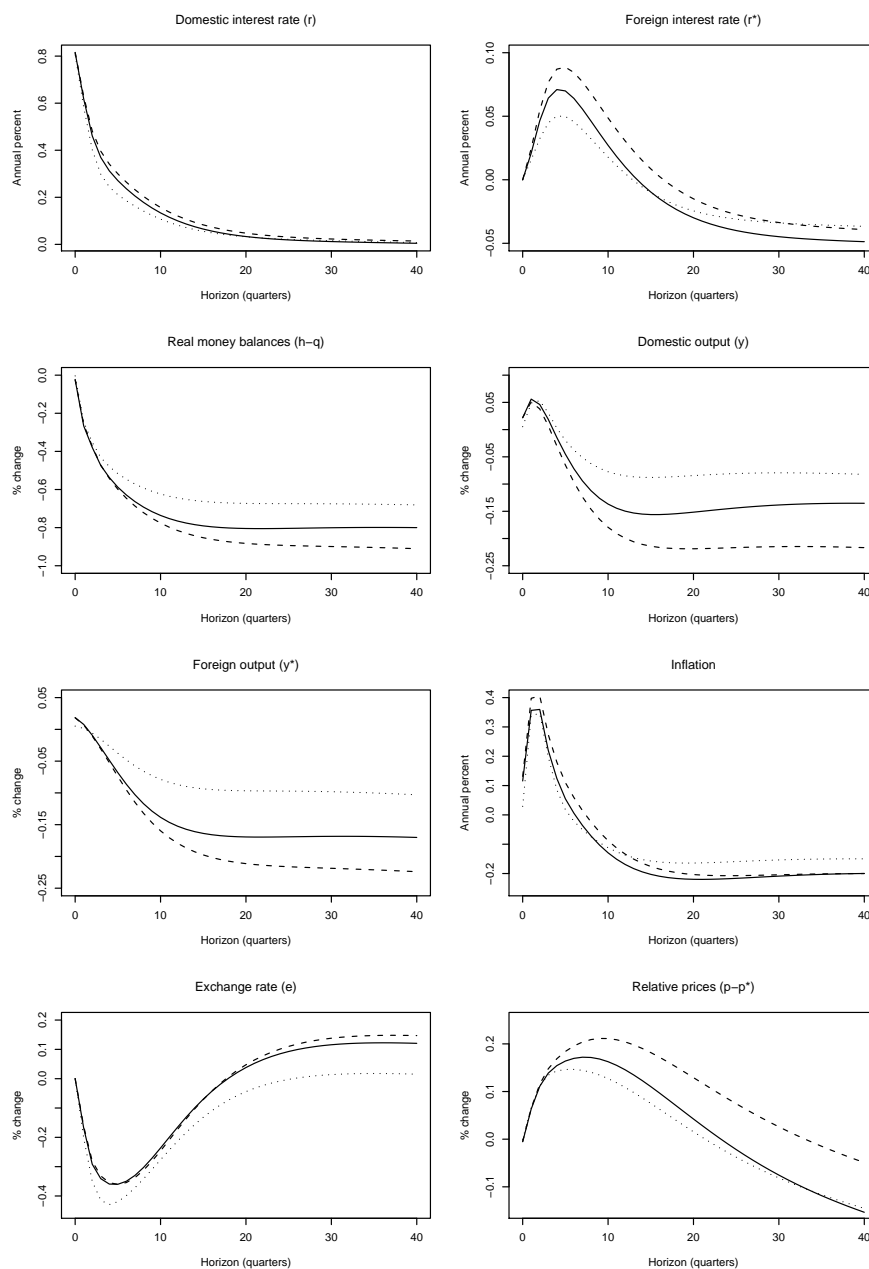
Figure 2: Responses to an Oil Price Shock. Solid line = SSVS for Everything, Dots = Best Model, Dashes = GLPS model
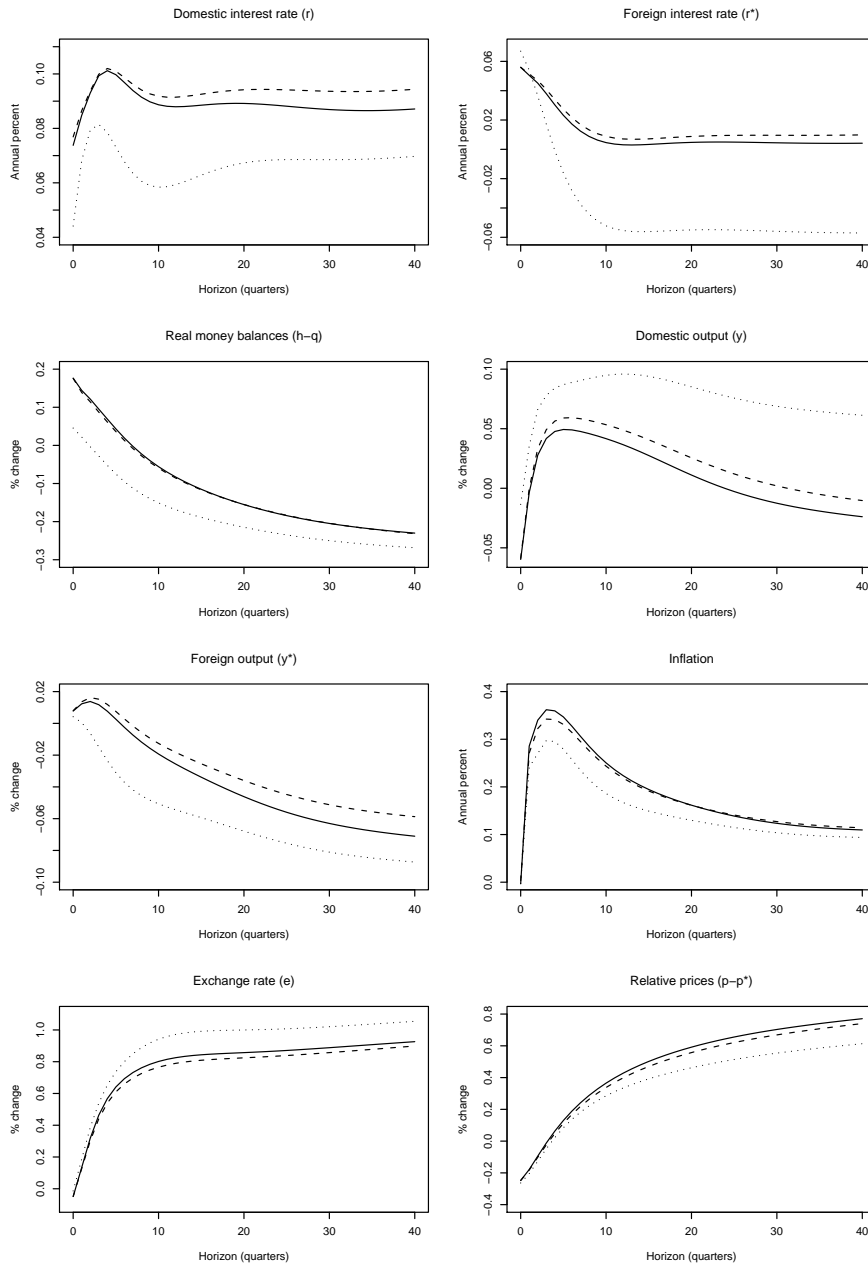
A more detailed set of impulse response functions, including 95% credible intervals, is provided in Appendix D. These credible intervals (similar to the 95% bootstrapped confidence intervals using in Garratt et al, 2003) tend to be fairly wide. Nevertheless, the same pattern of similarity across approaches (and to Garratt et al, 2003) is retained.

4 Conclusions

SSVS methods typically involve, for each parameter, choosing between a tight prior and a loose prior in a data-based fashion. The tight prior is constructed so as to shrink the parameter so that it is effectively zero. In this paper, we have extended these ideas to VECMs and to the case where the researcher is interested in restrictions involving several parameters. These extensions required us to define carefully what a "tight prior" and a "loose prior" was in the case of the cointegration space. An MCMC algorithm was developed for carrying out empirical inference in the resulting model.

In an empirical exercise involving a nine-variate macroeconomic model of the UK economy, we find our SSVS methods for VECMs to work well and stress that conventional BMA methods would simply be computationally infeasible in this case. Despite beginning with an over-parameterized unrestricted model and facing a huge model space, SSVS quickly moved towards much more parsimonious models. Regardless of whether MCMC output was used for model averaging or model selection, the resulting impulse responses were sensible and similar to those found using the model of GLPS.

References

Chipman, H., George, E. and McCulloch, R. (2001). "The practical implementation of Bayesian model selection," pages 65-134 in Institute of Mathematical Statistics Lecture Notes - Monograph Series, Volume 38, edited by P. Lahiri.

Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 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., Ley, E. and Steel, M. (2001). "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16, 563-576.

Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2003). "A Long-run structural macroeconometric model of the UK economy," *Economic Journal,* 113, 412-455

Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2006). *Global and National Macroeconometric Modelling: A Long-Run Structural Approach,* Oxford: Oxford University Press.

George, E., Sun, D. and Ni, S. (2008). "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553-580.

Giordani, P. and Kohn, R. (2008). "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business and Economic Statistics*, 26, 66-77.

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

Jochmann, M., Koop, G. and Strachan, R. (2009). "Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks," forthcoming in *International Journal of Forecasting.*

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

Koop, Leon-Gonzalez, R. and Strachan, R. (2009a). "Efficient posterior simulation for cointegrated models with priors on the cointegration space," *Econometric Reviews*, forthcoming.

Koop, Leon-Gonzalez, R. and Strachan, R. (2009b). "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control,* 33, 997–1017.

Koop, G., Leon-Gonzalez, R. and Strachan, R. (2009c). "Bayesian inference in a cointegrating panel data model," forthcoming in *Advances in Econometrics*, volume 23

Koop, G., Potter, S. and Strachan, R. (2008). "Re-examining the consumption-wealth relationship: The role of model uncertainty," *Journal of Money, Credit and Banking*, 40, 341-367.

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

Liu, J. and Wu, Y. (1999). "Parameter expansion for data augmentation," *Journal of the American Statistical Association*, 94, 1264-1274.

Primiceri. G. (2005). "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821-852.

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

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

Sims, C. and Zha, T. (2006). "Were there regime switches in macroeconomic policy?" *American Economic Review*, 96, 54-81.

Strachan, R. (2003). "Valid Bayesian estimation of the cointegrating error correction model," *Journal of Business and Economic Statistics*, 21, 185-195.

Strachan, R. and Inder, B. (2004). "Bayesian analysis of the error correction model," *Journal of Econometrics*, 123, 307-325.

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

Villani, M. (2006). "Bayesian point estimation of the cointegration space," *Journal of Econometrics*, 134, 645-664.

Appendix A: Markov Chain Monte Carlo Algorithm

In order to derive the Gibbs sampling algorithm we note that the model can be written as:

$$Y = X\beta^*\alpha^* + W\Gamma + E = Z\theta + E, \quad E \sim \text{MN}(0, \Sigma, I_T), \qquad (A1)$$

where $Z \equiv (X\beta^*, W)$ and $\theta \equiv (\alpha^{*\prime}, \Gamma')'$. Vectorizing (A1), we obtain:

$$y = (\alpha^{*\prime} \otimes X)b^* + (I_n \otimes W)\gamma + e = (I_n \otimes Z)d + e, \quad e \sim \text{N}(0, \Sigma \otimes I_T), \;\; (A2)$$

where $y = \text{vec}(Y)$, $a^* \equiv \text{vec}(\alpha^*)$, $b^* \equiv \text{vec}(\beta^*)$, $\gamma \equiv \text{vec}(\Gamma)$, $d \equiv \text{vec}(\theta)$, $e \equiv \text{vec}(E)$ and $\Sigma = \Psi\Psi'$. We label the nonzero elements of Ψ as ψ_{ij} and define $\psi \equiv (\psi_{11}, \ldots, \psi_{nn})'$, $\eta_j = (\psi_{1j}, \ldots, \psi_{j-1,j})'$ and $\eta = (\eta_2', \ldots, \eta_n')'$. Recall the prior distributions:

$$a^*|\rho \sim \text{N}(0, A) \text{ with } A = \text{diag}(\nu_1^2, \ldots, \nu_{nr_{\max}}^2), \nu_j^2 = (1-\rho_j)\underline{\nu}_{0j}^2 + \rho_j\underline{\nu}_{1j}^2, \;\; (A3)$$

$$b^*|\phi \sim \text{N}(0, I_{r_{\max}} \otimes P), \qquad (A4)$$

where $P = (HH' + \underline{\tau} H_\perp H_\perp')$ and $\phi \in \{1, 2, .., 18\}$ denotes the different choices for P. Next

$$\gamma|\delta \sim \text{N}(0, D) \text{ with } D = \text{diag}(\kappa_1^2, \ldots, \kappa_{nk}^2), \kappa_j^2 = (1-\delta_j)\underline{\kappa}_{0j}^2 + \delta_j\underline{\kappa}_{1j}^2, \quad (A5)$$

$$\psi_{jj}^2 \sim \text{Gamma}(\underline{a}_j, \underline{b}_j), \quad j = 1, \ldots, n, \qquad (A6)$$

$$\eta_j|\omega \sim \text{N}(0, F_j) \text{ with } F_j = \text{diag}(\xi_{1j}^2, \ldots, \xi_{j-1,j}^2), \quad j = 2, \ldots, n,$$
$$\xi_{ij}^2 = (1-\omega_{ij})\underline{\xi}_{0j}^2 + \omega_{ij}\underline{\xi}_{1j}^2, \quad j = 2, \ldots, n, \; i = 1, \ldots, j-1, \qquad (A7)$$

$$\rho_j \sim \text{Bernoulli}\left(\underline{p}_{\alpha j}\right), \qquad (A8)$$

$$\delta_j \sim \text{Bernoulli}\left(\underline{p}_{\gamma j}\right), \qquad (A9)$$

$$\omega_{ij} \sim \text{Bernoulli}\left(\underline{p}_{\psi ij}\right). \qquad (A10)$$

From (A9) and (A10) we can derive the prior distribution for d as

$$d|\rho, \delta \sim \text{N}(0, G), \qquad (A11)$$

where G is constructed from A and D.

We now can derive the following conditional posterior distributions:

1. $b^* \sim \mathrm{N}(\bar{b}, \bar{B})$ with

$$\bar{B} = \left[(\alpha^* \Sigma^{-1} \alpha^{*\prime}) \otimes (X'X) + I_{r_{\max}} \otimes P^{-1} \right]^{-1} \tag{A12}$$

and

$$\bar{b} = \bar{B}(\alpha^* \Sigma^{-1} \otimes X')[y - (I_n \otimes W)\gamma]. \tag{A13}$$

2. $d \sim \mathrm{N}(\bar{d}, \bar{D})$ with

$$\bar{D} = \left(\Sigma^{-1} \otimes Z'Z + G^{-1} \right)^{-1} \tag{A14}$$

and

$$\bar{d} = \bar{D}(\Sigma^{-1} \otimes Z')y. \tag{A15}$$

3. ψ_{jj}^2 are independent of one another $\psi_{jj}^2 \sim \mathrm{Gamma}(\underline{a}_j + 0.5\,T, \bar{b}_j)$ with

$$\bar{b}_j = \begin{cases} \underline{b}_1 + 0.5\,v_{11}, & \text{for } j = 1, \\ \underline{b}_j + 0.5\,[v_{jj} - v_j'(V_{j-1} + F_j^{-1})^{-1}v_j], & \text{for } j = 2, \dots, n \end{cases} \tag{A16}$$

and

$$V \equiv (Y - X\beta^*\alpha^* - W\Gamma)'(Y - X\beta^*\alpha^* - W\Gamma) \tag{A17}$$

has elements v_{ij}, $v_j \equiv (v_{1j}, \dots, v_{j-1,j})'$ and V_j is the upper left $j \times j$ block of V.

4. η_j are independent of one another $\eta_j \sim \mathrm{N}(\bar{\eta}_j, \overline{V}_j)$ with

$$\overline{V}_j = (V_{j-1} + F_j^{-1}) \tag{A18}$$

and

$$\bar{\eta}_j = -\psi_{jj}\overline{V}_j v_j. \tag{A19}$$

5. $\mathrm{P}(\phi = \phi_j) \propto p_j$ with $p_j = f_N(b^*; 0, I_r \otimes P_j)$ where $f_N(.)$ is the Normal p.d.f.

6. ρ_j are independent of one another $\rho_j \sim \mathrm{Bernoulli}(\bar{p}_{\alpha j})$ with

$$\bar{p}_{\alpha j} = \frac{\dfrac{1}{\underline{\nu}_{1j}} \exp\left(-\dfrac{\alpha_j^2}{2\underline{\nu}_{1j}^2} \right) \underline{p}_{\alpha j}}{\dfrac{1}{\underline{\nu}_{1j}} \exp\left(-\dfrac{\alpha_j^2}{2\underline{\nu}_{1j}^2} \right) \underline{p}_{\alpha j} + \dfrac{1}{\underline{\nu}_{0j}} \exp\left(-\dfrac{\alpha_j^2}{2\underline{\nu}_{0j}^2} \right) \left(1 - \underline{p}_{\alpha j} \right)}. \tag{A20}$$

28

7. δ_j are independent of one another $\delta_j \sim \text{Bernoulli}(\overline{p}_{\gamma j})$ with

$$\overline{p}_{\gamma j} = \frac{\dfrac{1}{\underline{\kappa}_{1j}} \exp\left(-\dfrac{\gamma_j^2}{2\underline{\kappa}_{1j}^2}\right) \underline{p}_{\gamma j}}{\dfrac{1}{\underline{\kappa}_{1j}} \exp\left(-\dfrac{\gamma_j^2}{2\underline{\kappa}_{1j}^2}\right) \underline{p}_{\gamma j} + \dfrac{1}{\underline{\kappa}_{0j}} \exp\left(-\dfrac{\gamma_j^2}{2\underline{\kappa}_{0j}^2}\right)\left(1 - \underline{p}_{\gamma j}\right)}. \quad \text{(A21)}$$

8. ω_{ij} are independent of one another $\omega_{ij} \sim \text{Bernoulli}(\overline{p}_{\psi ij})$ with

$$\overline{p}_{\psi ij} = \frac{\dfrac{1}{\underline{\xi}_{1ij}} \exp\left(-\dfrac{\psi_{ij}^2}{2\underline{\xi}_{1ij}^2}\right) \underline{p}_{\psi ij}}{\dfrac{1}{\underline{\xi}_{1ij}} \exp\left(-\dfrac{\psi_{ij}^2}{2\underline{\xi}_{1ij}^2}\right) \underline{p}_{\psi ij} + \dfrac{1}{\underline{\xi}_{0ij}} \exp\left(-\dfrac{\psi_{ij}^2}{2\underline{\xi}_{0ij}^2}\right)\left(1 - \underline{p}_{\psi ij}\right)}.$$

$$\text{(A22)}$$

Appendix B: Further Details on Prior Elicitation

Our SSVS prior for VECMs requires the construction of H_j for $j = 1, .., 18$ which impose the over-identifying restrictions or restrict the manner in which deterministic trends enters the cointegrating residuals. The manner in which such matrices are constructed is described in Strachan and Inder (2004) and the reader is referred to that paper for details. The matrices below are not semi-orthogonal, but since $sp(H) = sp(H\kappa)$ for any full rank square κ, after constructing H, we may innocuously make it semi-orthogonal by the transformation $H \rightarrow H(H'H)^{-1/2}$.

Noting that the tth row of X is ordered as $(1, t, p_t^o, e_t, i_t^*, i_t, \Delta p_t, q_t, p_t - p_t^*, h_t - q_t, q_t^*)$, the following restriction matrices can be constructed. The matrix which imposes no restrictions at all and simply leads to the noninformative prior for the cointegrating space defined in the body of the text is:

$$H_1 = I.$$

The matrix which deletes the deterministic trend term, but otherwise leaves the cointegrating space unrestricted is:

$$H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The matrix which deletes the intercept and deterministic trend term, but

otherwise leaves the cointegrating space unrestricted is:

$$
H_3 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

The matrix which imposes the purchasing power parity relationship in (4), but otherwise leaves the cointegrating space unrestricted is:

$$
H_4 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

The matrix which imposes the interest rate parity condition in (4), but oth-

erwise leaves the cointegrating space unrestricted is:

$$
H_5 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

The matrix which imposes the restriction implied by neoclassical growth models for the UK and the rest of the world in (4), but otherwise leaves the cointegrating space unrestricted is:

$$
H_6 = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

The matrix which imposes the restriction implied by Fisher inflation parity arguments in (4), but otherwise leaves the cointegrating space unrestricted

is:

$$H_7 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The matrix which imposes the first and second of the restrictions in (4) is:

$$H_8 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The matrix which imposes the first and third of the restrictions in (4) is:

$$H_9 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

33

The matrix which imposes the first and fourth of the restrictions in (4) is:

$$H_{10} = \begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

The matrix which imposes the second and third of the restrictions in (4) is:

$$H_{11} = \begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

The matrix which imposes the second and fourth of the restrictions in (4) is:

$$H_{12} = \begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

The matrix which imposes the third and fourth of the restrictions in (4) is:

$$H_{13} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix which imposes the first, second and third of the restrictions in (4) is:

$$H_{14} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix which imposes the first, second and fourth of the restrictions in

(4) is:

$$
H_{15} = \begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

The matrix which imposes the first, third and fourth of the restrictions in (4) is:

$$
H_{16} = \begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & -1 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

The matrix which imposes the second, third and fourth of the restrictions in

(4) is:

$$H_{17} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix which imposes all of the restrictions in (4) is:

$$H_{18} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Appendix C: Data Appendix

The data set is an updated version of the data set constructed by Garratt et al. (2003) and runs from 1965Q1 through 2008Q1. The variables are:

1. q_t: the natural logarithm of UK real per capita domestic output, computed as:
$$\ln(GDP_t/POP_t).$$

   GDP_t is real gross domestic product (at 1995 prices; 1995=100; seasonally adjusted; source: Office of National Statistics (ONS), Economic Trends, code: YBEZ). POP_t is total UK population (in thousands; source: ONS, Monthly Digest of Statistics, code: DYAY, annual series), which was available up to 2007. The 2008 number is obtained by extrapolating the 2007 number using the average growth rate for the period 2003-2007. The population series is converted into a quarterly series through geometric interpolation and rebased 1995=1.

2. p_t: the natural logarithm of the domestic price level, computed as:
$$\ln(P_t),$$

   where P_t is the UK producer price index (manufactured products; 1995=100; not seasonally adjusted; source: ONS, Economic Trends, code: PLLU).

3. Δp_t: the UK inflation rate, computed as:
$$\ln(\tilde{P}_t) - \ln(\tilde{P}_{t-1}),$$

   where $\tilde{P}_t$ is the UK retail price index (all items; 1995=100; not seasonally adjusted; source: ONS, Economic Trends, code: CHAW).

4. i_t: the quarterly domestic nominal interest rate, computed as:
$$\frac{1}{4} \ln \left(1 + \frac{R_t}{100} \right) \times 100,$$

   where R_t is the 90 day treasury bill average discount rate (annualized rate; source: ONS, Financial Statistics, code AJNB).

5. $h_t - q_t$: the natural logarithm of real per capita money stock as a proportion of real per capita income, computed as:

$$\ln\left(\frac{H_t}{Y_t}\right),$$

where H_t is the M0 money stock (end period; in Million Pound Sterling; seasonally adjusted; source: ONS, Financial Statistics, code AVAE). The AVAE series was available for the period 1969Q2-2006Q1. Outside this period the series is projected using the percentage change of estimated circulation of notes and coins with the public. Nominal income Y_t is measured using gross domestic product at market prices (in Million Pound Sterling; seasonally adjusted, source: ONS, Economic Trends, code: YBHA).

6. e_t: the natural logarithm of the UK nominal effective exchange rate, computed as:

$$- \ln(E_t),$$

where E_t is the Sterling effective exchange rate (1995=100; source: ONS, Financial Statistics, code: AGBG (old code: AJHX); ONS defines E_t as the foreign price of domestic currency, hence the minus sign). The AGBG series was available until 2006Q1. Afterwards, we project the series using the percentage change of a broader exchange rate index (source: ONS, Financial Statistics, code: BK82).

7. q_t^*: the natural logarithm of real per capita foreign output, computed as:

$$\ln\left(\frac{GDP_t^*}{POP_t^*}\right),$$

where GDP_t^* is the gross domestic product index for the original OECD countries (at 2000 prices; 1995=100; seasonally adjusted; source: OECD, Main Economic Indicators (MEI), code: OTF.VPVOBARSA.2000.S1). POP_t is the population of these countries (annual series; source: OECD, Labor Force Statistics), which was available up to 2003. The later figures are obtained by extrapolating the 2003 number using the average growth rate for the period 1999-2003. The population series is converted into a quarterly series through geometric interpolation and rebased 1995=1.

8. p_t^*: the natural logarithm of the foreign price index, computed as:

$$\ln(P_t^*),$$

where P_t^* is the OECD producer price index (1995=100; source: OECD, MEI, code: OTO.PPI). Data was available on this series from 1982Q1. The data prior to 1982Q1 was constructed by backwardly imposing the percentage changes of a separately constructed weighted average index of OECD consumer and producer prices on the 1982Q1 figure.

9. i_t^*: the quarterly foreign nominal interest rate:

$$\frac{1}{4} \ln\left(1 + \frac{R_t^*}{100}\right) \times 100,$$

with R_t^* being a weighted average of foreign annualized interest rates computed as:

$$R_t^* = 0.4382\, R_t^{\text{US}} + 0.236\, R_t^{\text{Germany}} + 0.2022\, R_t^{\text{Japan}} + 0.1236\, R_t^{\text{France}}.$$

The weights are chosen according to Garratt et al. (2003). The annualized interest rates are all from the IMF's International Financial Statistics (IFS). For the US we use the three-month treasury bill rate (code: 60C.ZF), for Germany the money market rate (code: 60B.ZF), for Japan the money market rate (code: 60B.ZF) and for France the three-month treasury bill rate (code: 60C.ZF).

10. p_t^o: the natural logarithm of the oil price, computed as:

$$\ln(P_t^o),$$

where P_t^o is the average price of crude petroleum (in US dollar per barrel; source: IMF, IFS, code: 176AAZZF; 1995=100).

Appendix D: Additional Empirical Results

In the text, point estimates of the impulse responses were provided (these remain the solid lines in the figures below). Here we also add 95% credible intervals (the dashed lines in the figures below) to give the reader some idea of the dispersion associated with the point estimates.

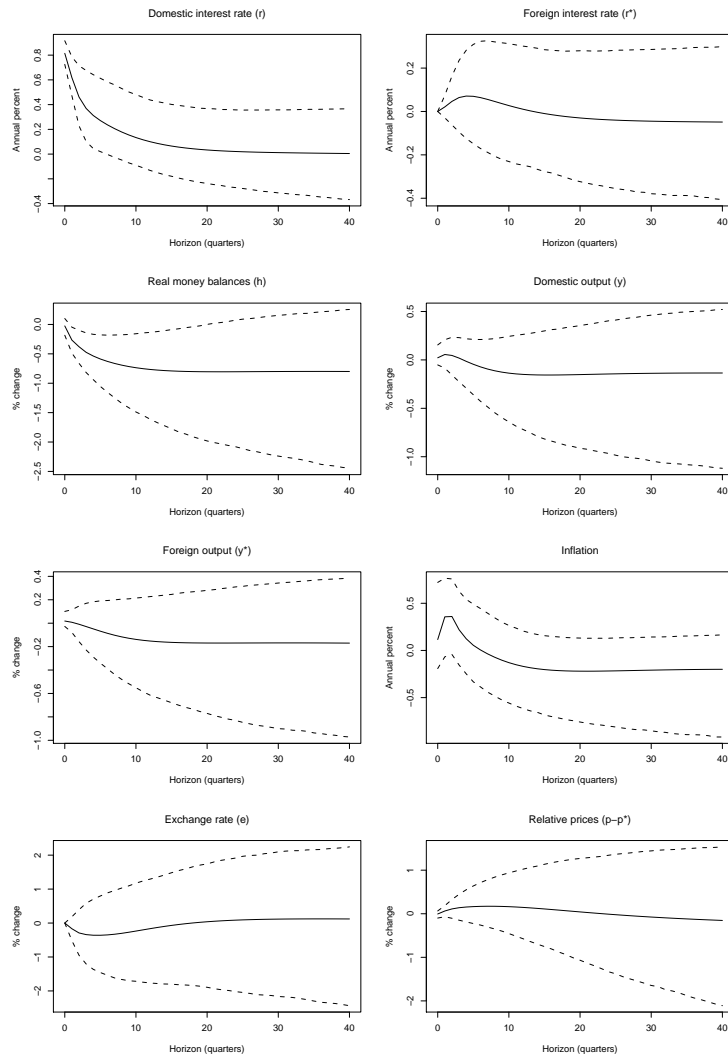Figure D1: Response to a Monetary Shock (SSVS for Everything)
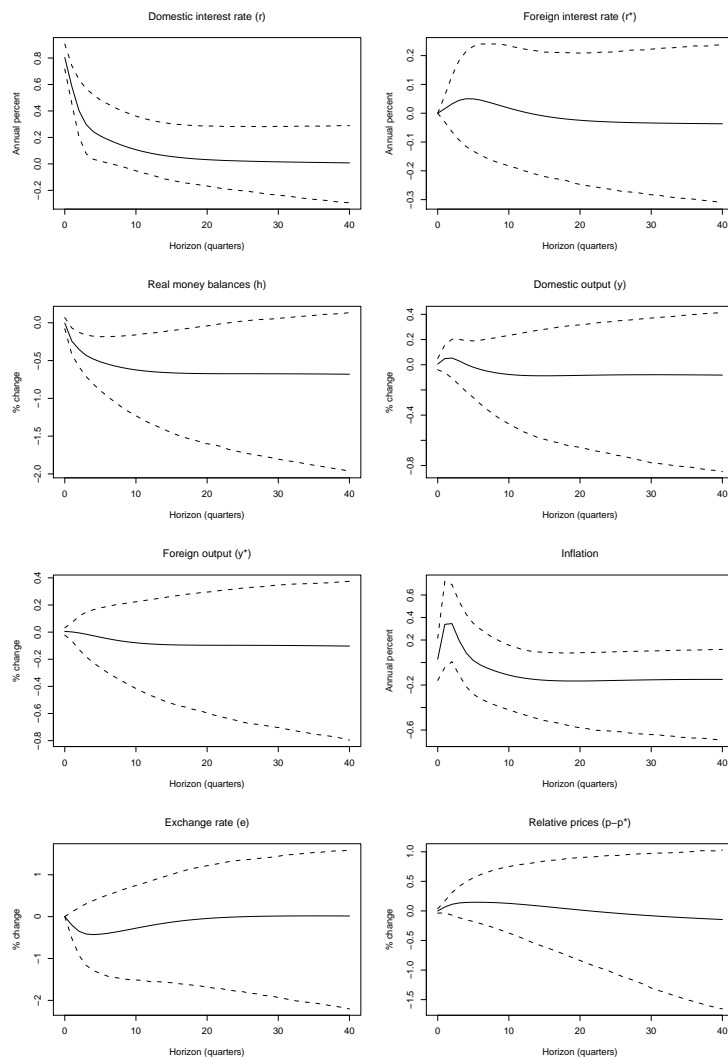
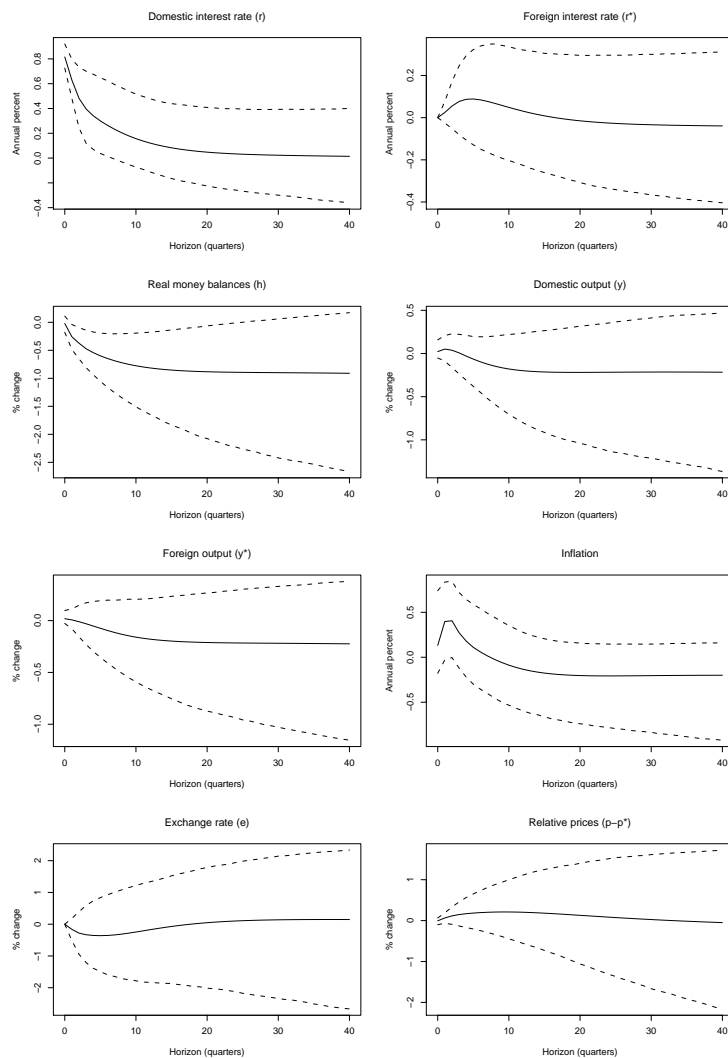Figure D2: Response to a Monetary Shock (Best Model)

Figure D3: Response to a Monetary Shock (GLPS Model)

43

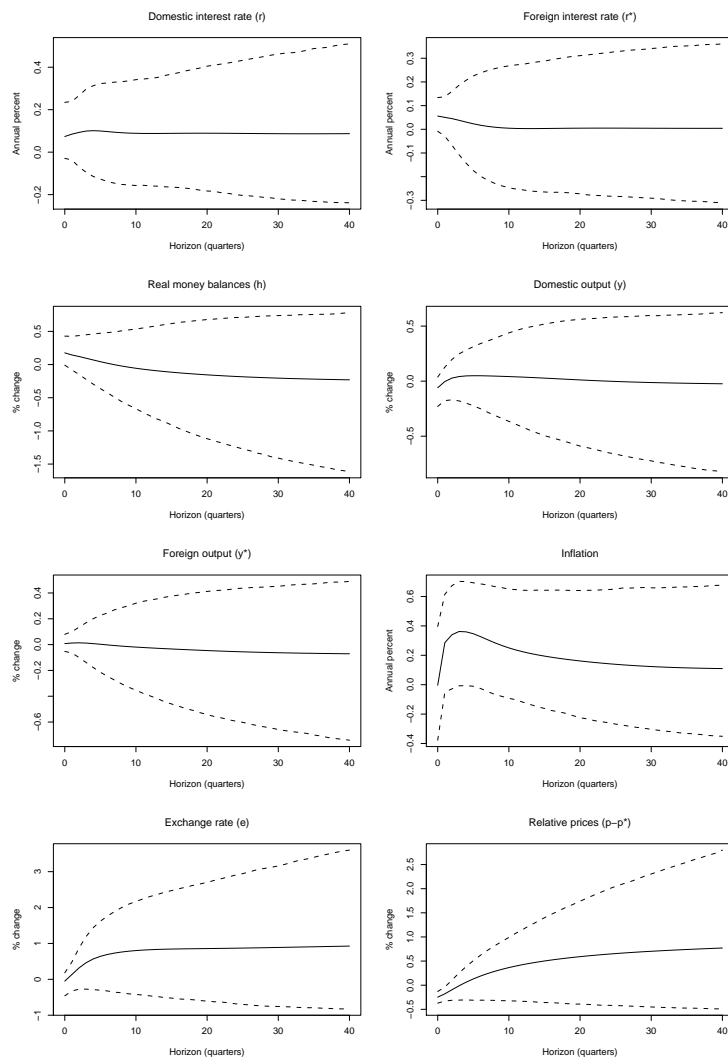Figure D4: Response to an Oil Price Shock (SSVS for Everything)
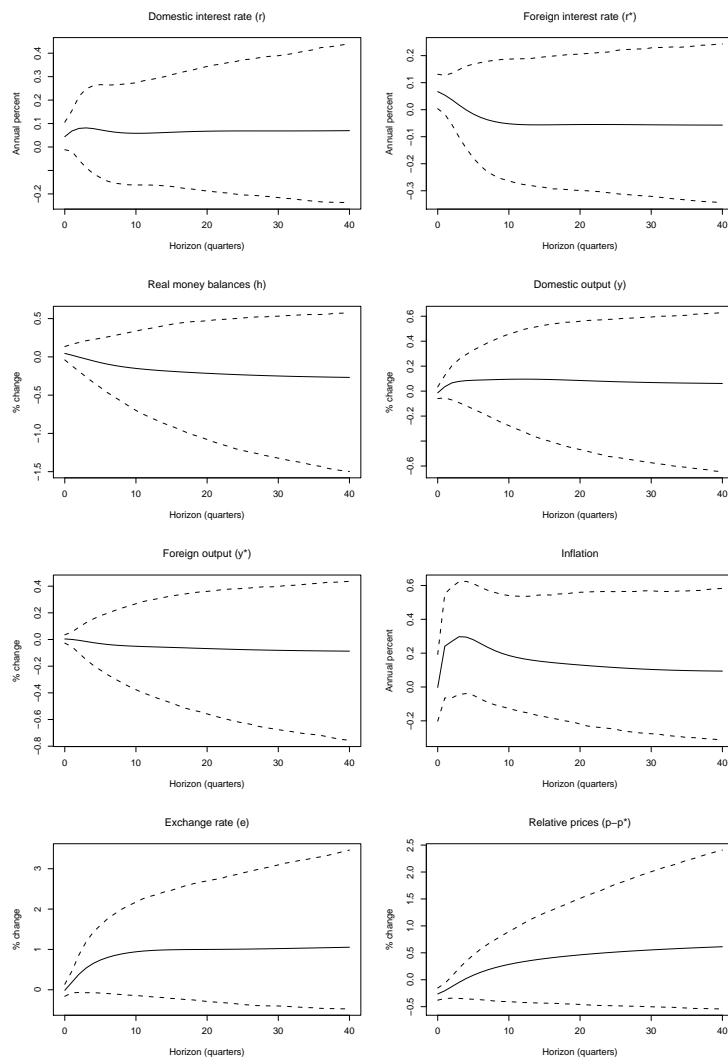
Figure D5: Response to an Oil Price Shock (Best Model)

Figure D6: Response to an Oil Price Shock (GLPS Model)