Reducing Dimensions in a Large TVP-VAR

Eric Eisenstat
University of Queensland

Joshua C.C. Chan
University of Technology Sydney

Rodney W. Strachan*
University of Queensland

March 16, 2018

*Corresponding author: School of Economics, University of Queensland, r.strachan@uq.edu.au. Rodney Strachan is a Fellow of the Rimini Centre for Economic Analysis and the Centre for Applied Macroeconomic Analysis.

ABSTRACT

This paper proposes a new approach to estimating high dimensional time varying parameter structural vector autoregressive models (TVP-SVARs) by taking advantage of an empirical feature of TVP-(S)VARs. TVP-(S)VAR models are rarely used with more than 4-5 variables. However recent work has shown the advantages of modelling VARs with large numbers of variables and interest has naturally increased in modelling large dimensional TVP-VARs. A feature that has not yet been utilized is that the covariance matrix for the state equation, when estimated freely, is often near singular. We propose a specification that uses this singularity to develop a factor-like structure to estimate a TVP-SVAR for 15 variables. Using a generalization of the recentering approach, a rank reduced state covariance matrix and judicious parameter expansions, we obtain efficient and simple computation of a high dimensional TVP-SVAR. An advantage of our approach is that we retain a formal inferential framework such that we can propose formal inference on impulse responses, variance decompositions and, important for our model, the rank of the state equation covariance matrix. We show clear empirical evidence in favour of our model and improvements in estimates of impulse responses.

Keywords: Large VAR; time varying parameter; reduced rank covariance matrix.

JEL Classification: C11, C22, E31

1 Introduction

Vector autoregressive models (VARs) have provided many valuable insights in applied macroeconometrics. The past decade has seen considerable interest in VARs with parameters that evolve over time — time varying parameter VARs (TVP-VARs) — particularly with heteroscedasticity, to better capture the evolving dynamics of the underlying variables. More recently researchers have been developing methods to estimate larger systems of variables in VARs to avoid limitations that arise when too few variables are modelled. The problems that motivate using both TVP-VARs and large VARs are compelling, but addressing both problems in one model leads to significant computational challenges. This paper proposes an approach to address these challenges.

Banbura, Giannone and Reichlin (2010) argue for modelling many variables in a large VAR to avoid a number of problems that arise from modelling too few variables. They (and other authors such as Carriero, Kapetanios and Marcellino (2011), Giannone, Lenza, Momferatou and Onorante (2014), Koop (2013) and Koop and Korobilis (2013)), point out that forecasts, policy advice and analysis of structure suffer problems resulting from omitted variable bias from using too few variables in the VAR. Typical sample sizes in the VAR literature, however, are not large and so using large VARs leads to significant parameter proliferation making estimation and more general inference either difficult or infeasible. Banbura *et al.* (2010) address this problem by employing the so-called Litterman prior to impose sufficient shrinkage to permit inference.

The time varying parameter vector autoregressive model (TVP-VAR) allows for the processes generating macroeconomic variables to evolve over time. These models, which are most commonly given a state space representation, have informed us on a range of questions of interest to policymakers with perhaps the most notable area of application being on the transmission of monetary policy (see, for example, Cogley and Sargent (2001, 2005), Primiceri (2005), and Koop *et al.* (2009)). Although the number of variables modelled using TVP-VARs tends not to be very large, the arguments for using large VARs have quite naturally led to efforts to develop large TVP-VARs. As the number of states grows polynomially in the number of variables and time then, as in the large VARs, computational difficulties are encountered in these models when there are many variables. These difficulties tend to limit the number of variables modelled using the TVP-VAR.[1]

An issue that has been bubbling away in the background in the literature on TVP-VARs is the treatment of the state equation covariance matrix. This matrix is often specified as diagonal, although there is good reason to specify this as a full matrix. Primiceri (p. 830, 2005) provides an argument that a full covariance matrix for the vector of all mean equation states and the structural parameters would be most appropriate as the states are, and are expected to be, highly correlated. However, he does not adopt such a specification in order to avoid parameter proliferation and the attendant computational issues. Primiceri does maintain a full covariance matrix for the reduced form mean equation states and more papers are doing so (see for example, Eisenstat, Chan and Strachan (2016)). A full state equation covariance matrix poses significant computational challenges for large TVP-VARs. As the number of variables n grows, the number of mean parameters grows at order n^2 and the number of parameters in the state equation covariance matrix grows at n^4.

Koop and Korobilis (2013) present an approach to approximating large TVP-VARs by a particular treatment of the state equation covariance matrix. Using

[1] A few papers, such as Carriero, Clark and Marcellino (2016a,b) and Chan (2018), have developed large VARs with stochastic volatility. But these papers all restrict the VAR coefficients to be constant.

forgetting factors they replace the state equation covariance matrix with a matrix proportional to a filtered estimate of the posterior covariance matrix. In this paper we present an alternative restriction on the state equation covariance matrix that results in a reduced number of state errors driving the time-varying parameters.

The first contribution of this paper is to present an alternative approach to estimating large TVP-VARs. We increase the number of variables we can model in a TVP-VAR by taking advantage of the strong correlations among the states. We preserve the exact state space model but achieve parsimony by imposing a restriction suggested by the data; that the state equation covariance matrix has reduced rank. An early observation by Cogley and Sargent (2005) shows, using principal component analysis, that the posterior estimate of the covariance matrix for the state equation appears to have a very low rank. We formalise this observation into a model specification. Primiceri (2005) points out that small state equation error variances cause problems for frequentist computation. Our approach, by contrast, uses this feature to improve Bayesian estimation.

A TVP-VAR with a reduced rank covariance matrix for the states permits a significant reduction in the dimension of the states without reducing the dimension of the VAR. For example, we estimate a TVP-VAR for 15 variables, with two lags and stochastic volatility. In this model there are 585 time varying parameters which the data suggest are driven by 8 states. The reduction in the number of sources of states driving the time varying parameters comes by removing states that the data suggest are not needed. The resulting estimates of the time varying parameters are far more precise as a result.

As the dimension of models increases, estimation faces computational challenges. We employ a range of strategies, in addition to the reduced rank structure, to mitigate these issues. Each makes a small contribution on their own, but collectively they allow us to estimate larger models. First, by estimating the structural form of the TVP-VAR directly, we remove one sampling step in the Gibbs sampler. This is particularly important as estimating the reduced form TVP-VAR involves drawing two blocks of parameters that are naturally highly correlated. We collapse these two blocks into one and draw that block in one step. Next, to achieve a readily computable specification we generalize the scalar non-centered specification of the state space model by Frühwirth-Schnatter and Wagner (2010) to the matrix non-centered specification. This removes another step from the sampler as we draw the initial states and the state covariance matrix together in a single step. Further, we avoid the Kalman filter and smoother and, instead, use the precision sampler of Chan and Jeliazkov (2009). This precision sampler uses a lower order of computations to draw from the same posterior as the Kalman smoother.

The specification of the reduced rank model requires semi-orthogonal matrices and ordered positive elements. This specification induces nonstandard supports for the parameters and Bayesian computation on such supports is difficult. Another contribution of this paper, then, is to use a judicious selection of parameter expansions and priors for the expanding parameters to develop a specification that is fast, efficient and easy to compute. This expansion is part of the generalization of the re-centering method of Frühwirth-Schnatter and Wagner (2010) to a multivariate setting mentioned above.

The structure of the paper is as follows. In Section 2 we present the idea with a general state space model. We outline the model specifications that result from different assumptions about the rank of the state equation covariance matrix. This section also contains a technical derivation of the reduced sources of errors model that results from a reduced rank state equation covariance matrix. In Section 3 we outline posterior computation. Section 4 presents an application using a TVP-VAR with 15 variables to demonstrate the proposed methodology. Section 5 concludes and gives some indication of directions for future research.

2 Reducing the Sources of Variation

2.1 Overview

We will apply the reduced sources of error approach to a structural form TVP-VAR (TVP-SVAR). In VAR analysis, the measurement equation is often specified on the reduced form parameters, although we can readily transform between the reduced form and structural form. We prefer the structural form as it reduces the number of blocks of parameters to be estimated and makes the dependence among the structural and reduced form parameters simpler (i.e., linear).

For the $n \times 1$ vector y_t for $t = 1, ..., T$, the TVP-SVAR can be written as

$$B_{0,t} y_t = \mu_t + B_{1,t} y_{t-1} + \cdots + B_{p,t} y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), \tag{1}$$

where $B_{0,t}, \ldots, B_{p,t}$ are $n \times n$ and $\Sigma_t = diag(\exp(h_{1,t}), \ldots, \exp(h_{n,t}))$. The first matrix $B_{0,t}$ is $n \times n$ with ones on the diagonal and is commonly specified as lower triangular.

Given the structure of $B_{0,t}$, we may write $B_{0,t} = I - B_t$ so that the matrix B_t has zeros on the diagonal. The TVP-SVAR can now be written as:

$$\begin{aligned} y_t &= \mu_t + B_t y_t + B_{1,t} y_{t-1} + \cdots + B_{p,t} y_{t-p} + \varepsilon_t \\ &= \mu_t + (y_t' \otimes I_n) D b_t + \left(y_{t-1}' \otimes I_n\right) b_{1,t} + \cdots + \left(y_{t-p}' \otimes I_n\right) b_{p,t} + \varepsilon_t, \end{aligned}$$

where $b_{l,t} = vec\left(B_{l,t}\right)$, $l = 1, ..., p$ and $Db_t = vec\left(B_t\right)$ where b_t contains all the $\frac{n(n-1)}{2}$ non-zero elements of B_t in a vector and D is an appropriately defined $n^2 \times \frac{n(n-1)}{2}$

selection matrix. If we define the $n \times k$ matrix

$$x_t = \begin{bmatrix} I_n & (y_t' \otimes I_n) D & (y_{t-1}' \otimes I_n) & \cdots & (y_{t-p}' \otimes I_n) \end{bmatrix}$$

such that $k = \left(np + 1 + \frac{n-1}{2}\right) n$ and the $(k \times 1)$ vector $\alpha_t = \begin{pmatrix} \mu_t' & b_t' & b_{1,t}' & \cdots & b_{p,t}' \end{pmatrix}'$, we can write the above model using a standard but reasonably general specification of the state space model for an observed $n \times 1$ vector of observations y_t with $n \times k$ matrix of regressors x_t :

$$\begin{aligned} y_t &= x_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), & (2) \\ \alpha_t &= \alpha_{t-1} + \eta_t, \quad \eta_t \sim N(0, Q_\alpha), \quad \alpha_0 = \alpha \sim N(\underline{\alpha}, \underline{V}). & (3) \end{aligned}$$

In the application in Section 4, it will be more convenient to transform from the VAR to the VECM form, but this again can be written in the general form in (2) and (3). We therefore continue with the general form of the model and delay giving specific details on the prior we use until Section 4. We can now present the idea of reducing the sources of errors in a general linear Gaussian state space model.

We have not imposed any restrictions on the above model at this point and all of the parameters in the VAR are able to vary over time. The dimension reduction occurs by applying a rank reduction to the covariance matrix for the state equation, Q_α. If we set the rank of Q_α to $r_\alpha = rank(Q_\alpha) \leq k$, then after applying the appropriate transformations (detailed in the next subsection below) we can write the model in (2) and (3) as follows:

$$\begin{aligned} y_t &= x_t \alpha + x_t A_\alpha f_{\alpha,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), & (4) \\ f_{\alpha t} &= f_{\alpha,t-1} + z_{\alpha,t}, \quad z_{\alpha,t} \sim N(0, I_{r_\alpha}), \quad f_{\alpha,0} = 0, & (5) \end{aligned}$$

where A_α is a $(k \times r_\alpha)$ matrix, $f_{\alpha,t}$ and $z_{\alpha,t}$ are $(r_\alpha \times 1)$ vectors and the errors ε_t and z_t are independent of one another. As r_α is generally much smaller than k, we call the model in (4) and (5) the reduced sources of error model.

The technical details on the link between the general form of the state space model in (2) and (3) and the final form in (4) and (5), including centring and parameter expansions, are presented in the following subsection for the interested reader. There are a number of choices in modelling the state space model and the correlation structure. In this paper we extend the above to reducing the rank of the covariance matrix for the volatility states, Q_h. We present two specifications, the second encompasses the first but there are significant differences in computation between the two specifications.

In the transformation from (2) to (3) we use $\alpha_t = \alpha + A_\alpha f_{\alpha,t}$ where $A_\alpha A'_\alpha = Q_\alpha$. This function implies that the k time varying parameters in α_t are driven by $r_\alpha \leq k$ states, $f_{\alpha,t}$, in a factor-like structure for the states. The elements of A_α and $f_{\alpha,t}$ are not identified and this results from the use of parameter expansions. These expansions relax the form of the model to improve estimation. In fact, we derive the above form starting from identified parameters but then introduce the parameter expansions that take away this identification.

To give an impression of the extent of dimension reduction that is typically achieved, consider our empirical application. We have $n = 15$ variables and $T = 250$ observations for a VAR with 2 lags. The dimension of the states α_t and the covariance matrix Q_α in the unrestricted model in (2) and (3) has dimension 305,235. With rank of Q_a set to $r_\alpha = 4$, which is preferred in this application, then Specification 1 in (4) and (5) has dimension 3,850 representing a 98.7% reduction in model dimension. It is worth noting that the dimension reduction is greater, the larger is n.

Recall that with the full covariance matrix Q_α the dimension of this matrix grows at rate n^4. Instead of using the specification of the state space model with a full covariance matrix Q_α, one might therefore use a diagonal specification of Q_α in the hope of reducing the dimension of the model. However, this does not result in as great a dimension reduction as using a reduced rank Q_α. In the case considered in our application, for example, the states α_t and the diagonal covariance matrix Q_α in the unrestricted model in (2) and (3) would have dimension 143,070. Thus the model (4) and (5) with rank of Q_a of $r_\alpha = 4$ still has a dimension 97.3% smaller than if a diagonal form were chosen for Q_α.

2.2 Mapping to the reduced sources of errors model

In this subsection, we present the details of the transformations from (2) and (3) to (4) and (5). Important features of the transformed model are that there are no unknown parameters in the state equations and that the parameters to be estimated all appear in the mean equation. Further, all of the parameters in α, A_α and $f_{\alpha,t}$ have conditionally normal posteriors.

Frühwirth-Schnatter and Wagner (2010) develop a computationally efficient specification of the state space model that permits the time variation in individual parameters to be 'turned off'. This approach involves two transformations: recentering (or non-centering) and parameter expansion. We leave for a subsequent paper consideration of turning off time variation. Rather we use the non-centered specification to develop a reduced rank model from which it is simpler to obtain draws of the parameters.

In recentering, the initial value is subtracted from all states and this is divided by the standard deviation of the state equation error. This transformation moves the initial state and the standard deviation into the mean equation leaving no unknown parameters in the state equation.

The Frühwirth-Schnatter and Wagner (2010) approach is developed for scalar or independent states. That is, Q_α is assumed to be scalar or a diagonal matrix. In our model the covariance matrix Q_α is a full symmetric matrix allowing correlation among the elements of η_t. We denote the initial state by α. Generalizing to this case, the recentering transforms from α_t to $\widetilde{\alpha}_t$ via

$$\alpha_t = \alpha + Q_\alpha^{1/2}\widetilde{\alpha}_t, \tag{6}$$

and the model subsequently becomes

$$y_t = x_t\alpha + x_t Q_\alpha^{1/2}\widetilde{\alpha}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), \tag{7}$$

$$\widetilde{\alpha}_t = \widetilde{\alpha}_{t-1} + \widetilde{z}_t, \quad \widetilde{z}_t \sim N(0, I_k), \quad \widetilde{\alpha}_0 = 0. \tag{8}$$

This more general specification requires a useful definition for $Q_\alpha^{1/2}$, the square root of the covariance matrix Q_α. There are several ways to define the square root of a full symmetric matrix, but for our purposes the definition must allow for Q_α to have reduced rank. Our preferred definition, which can readily accommodate rank reduction, uses the singular value decomposition.

The singular value decomposition of Q_α can be written as $Q_\alpha = U\Lambda U'$ where $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_k\}$, $\lambda_i \geq \lambda_{i+1} \geq 0$ and $U \in O(k) \equiv \{U(k \times k) : U'U = I_k\}$ is an orthonormal matrix: $U'U = I_k$. Given Q_α, the elements of U are identified up to sign (which is trivially resolved). The matrix $Q_\alpha^{1/2}$ is defined simply as $Q_\alpha^{1/2} = U\Lambda^{1/2}U'$. In this paper we impose parsimony by letting the $k - r_\alpha$ smallest singular values of Λ to be zero. That is, we allow $\lambda_{r_\alpha+1} = \lambda_{r_\alpha+2} = \cdots = \lambda_{k-1} = \lambda_k = 0$ and collect the nonzero singular values into $\Lambda_1 = diag\{\lambda_1, \lambda_2, ..., \lambda_{r_\alpha}\}$. In this case, we can conformably decompose $U = [U_1 \quad U_2]$ such that $U_1 \in V_{r_\alpha,k} \equiv \{U\,(k \times r_\alpha) : U'U = I_{r_\alpha}\}$ and $U_1'U_2 = 0$ an $r_\alpha \times (k-1)$ matrix of zeros. Under this restriction

$$Q_\alpha^{1/2} = U\Lambda^{1/2}U' = U_1\Lambda_1^{1/2}U_1'.$$

We introduce the square root of the reduced rank covariance matrix into the specification (6) to obtain the expression

$$\begin{aligned} \alpha_t &= \alpha + U_1\Lambda_1^{1/2}U_1'\widetilde{\alpha}_t \\ &= \alpha + U_1\Lambda_1^{1/2}\underline{f}_t, \end{aligned} \tag{9}$$

where in the second line we have taken the linear combination $\underline{f}_t = U_1'\widetilde{\alpha}_t$. The rank reduction implies a reduction in the number of states from k (in α_t) to r_α (in $\underline{f}_t$). Taking the linear combination $\underline{f}_t$ in the state equation implies also taking the linear combinations of the $\underline{z}_t = U_1'\widetilde{z}_t$. Here we have used the result that a linear combination of standard normal random variables ($\widetilde{z}_t$) in which the linear combinations are formed using a set of unit vectors (U_1 in our case) results in a vector of standard normal variables ($\underline{z}_t$). Thus the resulting state equation vector of errors, $\underline{z}_t$, is an $r_\alpha-$ vector of standard normal variables. That is, the state equation is now

$$\underline{f}_t = \underline{f}_{t-1} + \underline{z}_t, \quad \underline{z}_t \sim N\left(0, I_{r_\alpha}\right), \quad \underline{f}_0 = 0.$$

The above specification involves the parameters U_1 and Λ_1 which have very nonstandard supports. These nonstandard supports significantly complicate computation and it is difficult to obtain an efficient and simple algorithm. This issue is addressed by mapping to a less restrictive form by introducing unidentified parameters.

The second step in the approach of Frühwirth-Schnatter and Wagner (2010) is to introduce an unidentified parameter via an approach called *parameter expansion*, to map the parameters to more standard forms and supports. Used judiciously, transformation via parameter expansion can make computation much simpler and more efficient. This is achieved by the mapping to standard supports and employing standard distributions thereby simplifying computation and breaking down the dependency in the parameters (see discussion in, for example, Liu, Rubin and Wu (1998) and Liu and Wu (1999)). Importantly, this approach has proven useful in reduced rank models such as cointegrating vector error correction models (see Koop, Léon-González and Strachan (2010)), factor models (Chan, Léon-González and Strachan (2018)), and simultaneous equations models (Koop, Léon-González and Strachan (2012)).

Working in the scalar case, Frühwirth-Schnatter and Wagner (2010) introduce an indicator ι that randomly takes the values -1 or $+1$. The support for ι is therefore a one-dimensional orthogonal group, $O(1)$. Generalizing this, we expand the set of parameters by introducing the orthonormal matrix $C \in O(r_\alpha)$ where $O(r_\alpha)$ is the r_α-dimensional orthogonal group. Define the matrix $A_\alpha = U_1 \Lambda_1^{1/2} C'$. Note that the definition of A_α is just a standard singular value decomposition of a real matrix with singular values on the diagonal of $\Lambda_1^{1/2}$. Introducing this expanding parameter C into the model through (9) we obtain

$$\begin{aligned} \alpha_t &= \alpha + U_1 \Lambda_1^{1/2} C' C \underline{f}_t \\ &= \alpha + A_\alpha f_{\alpha,t}, \\ f_{\alpha,t} &= f_{\alpha,t-1} + z_{\alpha,t}, \quad z_{\alpha,t} \sim N(0, I_{r_\alpha}), \quad f_{\alpha,0} = 0 \end{aligned}$$

in which $f_{\alpha,t} = C \underline{f}_t$ and $z_{\alpha,t} = C \underline{z}_t$. Introducing the above transformation into the measurement equation in (7) and replacing the state equation in (8) by the one above, we obtain the final form of the full state space model as that given in (4) and (5).

2.3 Two Specifications for the Variance

The standard model assumed in the literature specifies α_t and $h_t = (h_{1,t}, \ldots, h_{n,t})'$ as *a priori* independent and that the covariance matrix in the state equation for h_t is full rank. For example, a standard specification is a random walk log-volatility

$$h_t = h_{t-1} + \eta_{h,t} \qquad \eta_{h,t} \sim N(0, Q_h)$$

where $Q_h = diag(\sigma^2_{h1}, , \ldots, \sigma^2_{hn})$ and the random walk is initialized with h_0.

In this section we apply the dimension reduction to the log variances, h_t, in (1). That is, we generalise to permit Q_h to be a full, possibly reduced rank symmetric matrix. Much of the parameter proliferation in the TVP-SVAR occurs in the mean equations but we could just as reasonably wish to reduce the number of states driving the stochastic volatility. The volatility component of the models we propose here resembles that of Carriero *et al.* (2016a). Expanding upon the specification in Section 2.1, we consider two specifications of the log volatility h_t for reducing the dimensions of the TVP-SVAR. The first, Specification 1, assumes the mean equation and volatilities share common states while Specification 2 specifies them to be *a priori* independent. The rationale for the first specification is that structural change in the mean and variance could come from a common source. That is, structural change is driven by a common factor. Specification 2 adopts the more standard assumption that the mean and variance states are independent. Specification 1 of the process for α_t and h_t encompasses Specification 2.

It is not difficult to imagine that shocks can drive changes in the whole structure of the model such that changes in the mean and variance parameters are driven by the same states. To allow for this possibility, we allow for the mean equation and volatility to influence each other in the most general model specification. In this model, the most general form, we allow the mean equation states, α_t, to be correlated with the log volatilities in h_t. To permit this we specify a state equation for α_t and h_t jointly as:

$$\theta_t = \begin{pmatrix} \alpha_t \\ h_t \end{pmatrix}.$$

Specification 1 has state equation

$$\theta_t = \theta_{t-1} + \eta_{\theta,t}, \qquad \eta_{\theta,t} \sim N\left(0, Q_\theta\right), \tag{10}$$

such that the mean and variance states are correlated. After applying the rank reduction to the above specification, the time varying parameters in the model are

$$\begin{aligned} \theta_t &= \theta + A f_{\theta,t}, \qquad A = \begin{pmatrix} A_\alpha \\ A_h \end{pmatrix}, \\ f_{\theta,t} &= f_{\theta,t-1} + z_t, \qquad z_t \sim N\left(0, I_r\right), \qquad f_{\theta,0} = 0, \end{aligned}$$

where $r = r_\alpha + r_h$, A is $(n+k) \times r$ and $f_{\theta,t}$ is $r \times 1$.

It is more common to impose, usually for computational convenience, that the errors in the state equations for α_t and h_t are independent. However, we wish to retain dependence among the volatilities. The second model, Specification 2, assumes that α_t and h_t are independent such that

$$A = \begin{pmatrix} A_\alpha \\ A_h \end{pmatrix} = \begin{pmatrix} A_{\alpha,11} & 0 \\ 0 & A_{h,12} \end{pmatrix}.$$

In this case, we could rewrite the model for h_t as

$$\begin{aligned} h_t &= h + A_{h,11} f_{h,t}, \\ f_{h,t} &= f_{h,t-1} + z_{h,t}, \qquad z_{h,t} \sim N\left(0, I_{r_h}\right), \qquad z_{h,0} = 0, \end{aligned}$$

where $A_{h,11}$ is $n \times r_h$, $f_{h,t}$ is $r_h \times 1$ and, as we might reasonably expect that the volatilities can be modelled with common factors, then $r_h \leq n$.

3 Posterior Estimation

The state space structure specifies the priors for the states — $f_{\alpha,t}$, $f_{h,t}$ and $f_{\theta,t}$ — so we now describe the priors for the initial conditions $\theta = (\alpha', h')'$ and covariance matrices $a = vec(A)$.

Frühwirth-Schnatter and Wagner (2010) provide evidence in support of using the Gamma prior, rather than the inverted Gamma prior, for their scalar state equation variance. In the generalisation presented in this paper, this equates to using a Wishart prior for Q_θ. For the full rank $(r = k)$ case, a zero mean normal prior for A implies a Wishart prior for Q_θ (see, for example, Zellner pp. 389-392 (1971) and Muirhead (1982)). We therefore give the matrix A a normal prior distribution, $a = vec(A) \sim N\left(0, cI_{(n+k)r}\right)$ for all three specifications. Through some experimentation, we find $c = 10^{-3}$ to be reasonable in a wide variety of settings.

For the initial states $\theta = \{\theta_j\}$ (which contains the elements of α and h), we note that in large models the dimension may be substantial, and hence, shrinkage priors may be desirable. This implies a choice of structure on the prior covariance matrix $\underline{V}_\theta = \left\{\underline{V}_{\theta_j}\right\}$. A number of options explored in the large Bayesian VAR literature may be applied here. We consider the *stochastic search variable selection* (SSVS) prior of the form:

$$\theta_j \mid \delta_j \sim \mathcal{N}(\underline{\theta}_j, c_{\delta_j} \underline{V}_{\theta_j}), \qquad (11)$$
$$\delta_j \sim q^{\delta_j} (1-q)^{1-\delta_j},$$

where $\delta_j \in \{0, 1\}$, $c_1 = 1$ and c_0 is some small constant. Of course, this will collapse to a typical normal prior if either $q = 1$ or $c_0 = 1$. Further, we combine SSVS with Minnesota priors as suggested in Korobilis (2013). Having normal conjugate priors for the initial conditions (α, h), the covariances $(a = vec(A))$ and the states (the f_t), the resulting conditional posteriors are normal for Specifications 1 and 2.

For the purposes of this section, we collect the T states into the vectors $f_m = \left(f'_{m,1}, f'_{m,2}, \ldots, f'_{m,T}\right)'$ for $m = \alpha, h$ or θ. Further, let $a_\alpha = \left(vec(\alpha)', vec(A_\alpha)'\right)'$ and $a_h = \left(h'_0, vec(A_h)'\right)$. The description of the priors above implies that the vectors a_α, f_α, a_h and f_h have a normal form such as $N\left(\underline{\mu}_m, \underline{V}_m\right)$ for $\mu = a$ or f. Volatility Specification 2 leads to a straightforward sampler. For Specification 2, MCMC involves five blocks:

1. $(a_\alpha | s_\alpha, f_\alpha, h_0, y) \sim \mathcal{N}\left(\overline{a}_\alpha, \overline{V}_\alpha\right)$;
2. $(f_\alpha | a_\alpha, s_\alpha, h_0, y) \sim \mathcal{N}\left(\overline{f}_\alpha, \overline{V}_{f,\alpha}\right)$;
3. $(s_\alpha | a_\alpha, s_\alpha, h_0, y)$;
4. $(a_h | a_\alpha, f_\alpha, y) \sim \mathcal{N}\left(\overline{a}_h, \overline{V}_h\right)$;
5. $(f_h | a_\alpha, f_\alpha, h_0, y) \sim \mathcal{N}\left(\overline{f}_h, \overline{V}_{f,h}\right)$;

of which steps 1, 2, 4 and 5 involve only analytically tractable conditional distributions, all of which are straightforward to sample from. The states, s_h, drawn in step 3 are the states determining the normal mixture components when drawing the stochastic volatilities using the algorithm of Kim, Shephard and Chib (1998).

For Specification 1 the MCMC consists of sampling recursively from:

1. $(a_\alpha | f_\theta, h_0, A_h, y) \sim \mathcal{N}\left(\overline{a}_\alpha, \overline{V}_\alpha\right)$;

2. $(f_\theta | a_\alpha, h_0, A_h, y)$;

3. $(a_h | a_\alpha, f_\theta, y) \sim \mathcal{N}\left(\overline{a}_h, \overline{V}_\alpha\right)$.

Under this specification, the measurement equation is nonlinear in f_θ (since it enters both the conditional mean and the volatility simultaneously), and therefore, $(f_\theta | \alpha, A_\alpha, h_0, A_h, y)$ is not analytically tractable. We therefore sample it using an accept-reject Metrolpolis-Hastings (ARMH) algorithm as described in Chan and Strachan (2012). Specifically, we use a normal proposal centered on the conditional posterior mode $\widehat{f}_\theta$ with the variance $\widehat{V}_\theta$ set to the negative inverse Hessian evaluated at the mode of $\ln p\left(f_\theta|., y\right)$. The derivation of $\widehat{f}_\theta$ and $\widehat{V}_\theta$ is given in Appendix 1.

Once the mode $\widehat{f}_\theta$ is obtained, the proposal precision $\widehat{V}_\theta^{-1}$ is given by a by-product of the scoring algorithm and a matrix that can be easily evaluated at the mode upon convergence (See Appendix 1). We then generate proposals as $f_\theta^c \sim \mathcal{N}\left(\widehat{f}_\theta, \widehat{V}_\theta\right)$ for the ARMH step as detailed in Chan and Strachan (2012). The use of ARMH as opposed to standard M-H appears to provide substantial gains in terms of acceptance rates (and hence sampling efficiency), particularly for larger models (i.e. as the size of f_θ increases). Intuitively, the normal proposal is symmetric, while $p\left(f_\theta|., y\right)$ will typically be skewed. This mismatch in shape will lead to higher rejection rates for a standard M-H approach as the dimension of f_θ increases. ARMH mitigates this by adjusting the shape of the proposal to better fit the skewness of the target distribution. As a result, acceptance rates are substantially increased. For example, in the macroeconomic application discussed below, the model with $n = 15$ and $r_\alpha = 10$ yields an acceptance rate of about 89.9%.

4 Application

4.1 Implementation

We use a data set containing a total of 15 variable to estimate the time-varying effects of surprise productivity (non-news) and news shocks. To understand the effects of dimension upon the results, we estimate the model with $n = 8$ variables and again with all $n = 15$ variables for contrast. The data consists of quarterly macroeconomic series covering the period 1954Q3–2008Q3, with each variable described in Table 1.[2] Given a subset of these variables, we assume the system admits a structural TVP-VAR representation of the form

$$y_t = B_{0,t}^{-1}\mu_t + \Pi_{1,t}y_{t-1} + \cdots + \Pi_{p,t}y_{t-p} + A_t\widetilde{\varepsilon}_t, \quad \widetilde{\varepsilon}_t \sim \mathcal{N}\left(0, I_n\right), \tag{12}$$

where $A_t = B_{0,t}^{-1}\Sigma_t^{1/2}$,

$$\begin{aligned} \Sigma_t &= diag\left(\exp\left(h_{1,t}\right), \ldots, \exp\left(h_{n,t}\right)\right) \text{ and} \\ \Sigma_t^{1/2} &= diag\left(\exp\left(h_{1,t}/2\right), \ldots, \exp\left(h_{n,t}/2\right)\right). \end{aligned}$$

Following Barsky and Sims (2011), non-news and news shocks in $\widetilde{\varepsilon}_t$ are identified by the restrictions:

1. *non-news* is the only shock affecting TFP on impact;

[2] Following standard practice in the news shock literature, all series are de-meaned.

2. *news* is the shock that, among all of the remaining shocks, explains the maximum fraction of the *forecast error variance* (FEV) of TFP at a long horizon (set to 20 years in our application).

Table 1: Variables used in each estimated model.

Core variables		Additional variables for the $n = 15$ model	
1	Log TFP	9	Log RPI
2	FED funds rate	10	Log real SEP500
3	GDP deflator inflation	11	Unemployment Rate
4	Log hours per capita	12	Vacancy rate
5	Log real GDP per capita	13	TB3MS Spread
6	Log real consumption per capita	14	GS10 Spread
7	Log real investment per capita	15	Log real dividends
8	GS5 Spread		

To implement the methodology outlined in the previous sections in estimating (12), we begin with the structural form in (1)

$$y_t = \mu_t + B_t y_t + B_{1,t} y_{t-1} + \cdots + B_{p,t} y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \Sigma_t\right)$$

were $\widetilde{\varepsilon}_t = \Sigma_t^{-1/2} \varepsilon_t$. To more simply apply a prior that are more useful in large models, we respecify the model in VECM form

$$\Delta y_t = \mu_t + B_t \Delta y_t + \Pi_t y_{t-1} + \Gamma_{1,t} \Delta y_{t-1} + \cdots + \Gamma_{p-1,t} y_{t-p+1} + \varepsilon_t \tag{13}$$

where B_t is the same lower triangular matrix defined in (12). Next, define

$$x_t = \begin{pmatrix} I_n & \left(\Delta y_t' \otimes I_n\right) D & \left(y_{t-1}' \otimes I_n\right) & \left(\Delta y_{t-1}' \otimes I_n\right) & \cdots & \left(\Delta y_{t-p+1}' \otimes I_n\right) \end{pmatrix}$$

such that $k = \left(np + 1 + \frac{n-1}{2}\right) n$ and α_t is the $(k \times 1)$ vector

$$\alpha_t = \begin{pmatrix} \mu_t' & b_t' & \pi_t' & \gamma_{1,t}' & \cdots & \gamma_{p-1,t}' \end{pmatrix}',$$

where μ_t and b_t are defined in Section 2 and $\gamma_{l,t} = vec\left(\Gamma_{l,t}\right)$ $l = 1, ..., p-1$ and $\pi_t = vec\left(\Pi_t\right)$. Consequently, we can now write (13) in the form of (2) and (3) as:

$$\begin{aligned} \Delta y_t &= x_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim N\left(0, \Sigma_t\right), \\ \alpha_t &= \alpha_{t-1} + \eta_t \quad \eta_t \sim N\left(0, Q_\alpha\right) \quad \alpha = \alpha_0 \sim N\left(\underline{\alpha}, \underline{V}\right). \end{aligned}$$

The advantage of this VECM specification is that it facilitates specifying more flexible shrinkage priors for

$$\alpha = \begin{pmatrix} \mu_0' & b_0' & \pi_0' & \gamma_{1,0}' & \cdots & \gamma_{p-1,0}' \end{pmatrix}',$$

which is useful in large dimensional settings.

In addition to the SSVS specification in (11) combined with the Minnesota prior for $\theta = (\alpha', h')' = \{\theta_j\}$, we also implement "inexact differencing" as advocated by Doan *et al.* (1984), Banbura *et al.* (2010), and others. This is done by setting the prior mean to $\underline{\theta}_j = 0$ for all j and the prior variance as

$$\underline{V}_{\theta_j} = \begin{cases} 1 & \text{if} \quad \theta_j \in h_0, \theta_j \in \mu_0, \text{or } \theta_j \in b_0, \\ \frac{10^2}{2n} & \text{if} \quad \theta_j \in \pi_0, \\ \frac{0.3}{2nl^2} & \text{if} \quad \theta_j \in \gamma_{l,0} \text{ for } l = 1, \ldots, p-1. \end{cases}$$

For SSVS, we set $c_0 = 0.01$ and $q = 0.5$. Finally, we scale each Δy_i to have sample standard deviation one before commencing sampling, which facilitates the use of generic prior settings like the ones given above. However, the effect of this scaling is reversed in the post-processing of draws such that all outputs such as impulse response functions are reported on the originally scaled data.

Once draws of B_t, Π_t, $\Gamma_{1,t}, \ldots, \Gamma_{p-1,t}$, and Σ_t are obtained, they are transformed to draws of $\Pi_{1,t}, \ldots, \Pi_{p,t}$ from (12) as

$$
\begin{aligned}
B_{0,t} &= I_n - B_t \\
\Pi_{1,t} &= I_n + B_{0,t}^{-1}\left(\Pi_t + \Gamma_{1,t}\right), \\
\Pi_{l,t} &= B_{0,t}^{-1}\left(\Gamma_{l,t} - \Gamma_{l-1,t}\right), \qquad l = 2, \ldots, p-1, \\
\Pi_{p,t} &= -B_{0,t}^{-1}\Gamma_{p-1,t}.
\end{aligned}
$$

To recover A_t, we begin with $\widetilde{A}_t = B_{0,t}^{-1}\Sigma_t^{-\frac{1}{2}}$. Note that by construction $\widetilde{A}_t$ is lower triangular and therefore the non-news shock is identified in accordance with restriction 1 above. However, the news shock generally does not satisfy restriction 2. Following Barsky and Sims (2011), the desired restriction is implemented by constructing an orthogonal matrix Q_t using a spectral decomposition of impulse response functions.

Specifically, for each period t we compute the impulse responses of log TFP to all shocks excluding non-news for the periods $t, t+1, \ldots, t+80$. Let R_s be the $(n-1)\times 1$ vector of impulse responses at time $t+s$ and take the spectral decomposition

$$
\widetilde{Q}_t D_t' \widetilde{Q}_t = \sum_{s=0}^{80} R_s R_s',
$$

where the eigenvalues in D_t are in *descending* order. Setting

$$
Q_t = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{Q}_t \end{bmatrix}
$$

and $A_t = \widetilde{A}_t Q_t$ achieves the desired identifying restriction, which is sufficient for computing forecast error variance decompositions. To derive impulse response functions, we further identify the sign of the news shock by requiring that the maximum impact of news on log TFP across all horizons is positive.[3]

4.2 Results

We begin by conducting an extensive empirical analysis on the choice of r_α (number of mean equation states) and r_h (number of states driving the volatility) using the Deviance Information Criterion (DIC) as the model comparison criterion. The DIC is based on the integrated likelihood — i.e., the joint density of the data marginal of all the latent states — and is computed using the method in Chan and Eisenstat (2018). The relative DICs are presented in Tables 2-3. A lower DIC value indicates more preference for the given model. For the model with $n = 8$ the DIC results suggest the preferred model is Specification 2 with five states in total: two states driving the mean equation coefficients in α_t and three states driving the volatilities h_t. Specification 1 is strongly rejected although does better with fewer states driving θ_t. Increasing the number of variables in the system to $n = 15$, the DIC select Specification 2 again although now with seven states: four states driving the mean equation coefficients in α_t and, again, three states driving the volatilities h_t.

In Figures 1 to 15 we present a range of impulse response functions and variance decompositions for responses to news and non-news shocks. The impact shocks in Figures 3 and 11 show significant variation over time in the impact of non-news shocks upon log TFP. The Figures 4 and 12 show that the long run impact of a new shock on the real variables $-$ log TFP, log real per capita GDP, log real per capita consumption

[3] In computing $\widetilde{Q}_t$ for $t > T - 80$ we set $\Phi_{l,t+s} = \Phi_{l,T}$ for all $t + s \geq T$.

and log real per capita investment − has declined over time with the density of the response moving towards zero. This effect is particularly clear for the larger model with $n = 15$.

There is a noticeable second order effect upon the estimated posterior impulse responses in both the $n = 8$ and the $n = 15$ models. Specifically, we see that the error bands suggest that there was a very large increase in uncertainty about the immediate effect of news shocks upon the Fed funds rate, the spread and to a lesser extent upon inflation around 1980. It is also in these second order effects upon the posterior that we see the effect of estimating a smaller model. Looking at Figures 4 and 12, the error bands are much tighter for the larger model despite this model having many more parameters to be estimated. We also see that estimating the smaller model we have the impression that the posteriors for a number of impacts, particularly to non-news shocks, are skewed at particular points in time and have higher probability of producing outliers from one tail at these times. These effects largely disappear when we estimate the larger, less restricted model.

5 Conclusion

This paper presents an approach to reducing the dimension of the TVP-SVAR, while preserving the full number of time varying parameters. The aim is to permit more efficient estimation of larger systems while preserving a full probability model and all formal inferential opportunities. The specification we employ is new and has a number of advantages. The dimension reduction is achieved by choosing a reduced rank of the state equation covariance matrix using empirical evidence. We employ DIC to select the rank of the covariance matrix. The specification is an exact one, allowing estimation of outputs, such as impulse responses and variance decompositions, and their full posterior distributions.

Computation remains a challenge in any large dimensional model, including the one presented in this paper. To mitigate this issue in this model we present a number of techniques that improve computation. These include careful specification of the model, judicious choice of computation algorithm, SSVS with a Minnesota prior to reduce the number of parameters, and use of parameter expansions to attain more readily computable forms for the final model. As a result, we present an approach that increases the range of models available to macroeconomists.

The application to a large system of 15 variables in a time varying VAR suggests that the estimates remain precise with sensible error bands. We find evidence of time variation in the impulse responses and differences between smaller and larger models. Subsequent work will consider automated selection of the rank of the state equation covariance matrix and inference on whether specific states vary over time or not (as per Frühwirth-Schnatter and Wagner (2010)).

6 References

Banbura M., D. Giannone, and L. Reichlin (2010) “Large Bayesian vector auto regressions”, *Journal of Applied Econometrics*, 25(1):71–92.

Barsky, Robert B. & Sims, Eric R.(2011) ”News shocks and business cycles”, *Journal of Monetary Economics*, Elsevier, 58(3): 273-289.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011) “Forecasting large datasets with Bayesian reduced rank multivariate models”, *Journal of Applied Econometrics*, 26(5):735–761.

Carriero A., T.E. Clark and M. Marcellino (2016a) “Common Drifting Volatility in Large Bayesian VARs”, *Journal of Business & Economic Statistics*, 34(3): 375-390.

Carriero, A., Clark, T. and Marcellino, M. (2016b). “Large Vector Autoregressions with stochastic volatility and flexible priors”, Working paper 1617, Federal Reserve Bank of Cleveland.

Chan, J. (2018). Large Bayesian VARs: A flexible Kronecker error covariance structure. Journal of Business & Economic Statistics, forthcoming.

Chan, J. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. Journal of Applied Econometrics, forthcoming.

Chan J. C. C. and I. Jeliazkov (2009) “Efficient simulation and integrated likelihood estimation in state space models”, *International Journal of Mathematical modelling and Numerical Optimisation* 1: 101–120.

Chan J. C. C., R. Léon-González and R. W. Strachan (2018) “Invariant inference and efficient computation in the static factor model”, *Journal of the American Statistical Association*, forthcoming, DOI: 10.1080/01621459.2017.1287080.

Chan J. C. C. and R. Strachan (2012) "Efficient estimation in non-linear non-Gaussian state-space models", Working paper, Research School of Economics, Australian National University.

Chib, S. (1996) "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models", *Journal of Econometrics*, 75: 79–97.

Cogley T. and T. Sargent (2001) "Evolving post-World War II inflation dynamics", *NBER Macroeconomic Annual* 16: 331-373.

Cogley, T. and T. Sargent (2005) "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.", *Review of Economic Dynamics* 8, 262-302.

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

Eisenstat, E., Chan, J. C. C. and Strachan, R. W. (2016). "Stochastic Model Specification Search for Time-Varying Parameter VARs" *Econometric Reviews* 35: 1638-1665.

Frühwirth-Schnatter S. and H. Wagner (2010) "Stochastic model specification search for Gaussian and partial non-Gaussian state space models", *Journal of Econometrics* 154 (1), 85-100.

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014) "Short-Term Inflation Projections: a Bayesian Vector Autoregressive Approach", International Journal of Forecasting, vol. 30(3), pages 635-644.

Kim, S., Shephard, N. and Chib, S. (1998) "Stochastic volatility: likelihood inference and comparison with ARCH models", *Review of Economic Studies*, 65, 361-93.

Koop, G. (2013) "Forecasting with Medium and Large Bayesian VARs", *Journal of Applied Econometrics*, 28, 177-203.

Koop, G. and D. Korobilis (2013) "Large Time-varying Parameter VARs", *Journal of Econometrics*, 177, 185-198.

Koop, G., R. Léon-González and R. W. Strachan (2009) "On the Evolution of Monetary Policy", *Journal of Economic Dynamics and Control* 33, 997-1017.

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

Koop, G., R. Léon-González and R. W. Strachan (2012) "Bayesian Model Averaging in the Instrumental Variable Regression Model", *The Journal of Econometrics* 171, 237-250.

Korobilis, D. (2013). "VAR Forecasting Using Bayesian Variable Selection", *Journal of Applied Econometrics*, 28, pp. 204-230.

Liu C., D. B. Rubin, Y. N. Wu (1998) "Parameter Expansion to Accelerate EM: The PX-EM Algorithm", *Biometrika*, Vol. 85, No. 4, pp. 755-770

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

Muirhead, R.J. (1982) Aspects of Multivariate Statistical Theory. New York: Wiley.

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

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics,* John Wiley and Sons Inc., New York.

A Appendix 1: Obtaining the mode and Hessian for the ARMH step

We use a scoring algorithm to find $\widehat{f}_\theta$ numerically. To this end, the gradient and Hessian of the log of the conditional posterior of f_θ, $\ln p\left(f_\theta|., y\right)$, are given by:

$$
\begin{aligned}
d &\equiv \frac{d\ln p\left(f_\theta|., y\right)}{\left(df_\theta\right)'} = -H'Hf_\theta - \frac{1}{2}\left(I_r \otimes A_h'\right)\iota_{rT} \\
&\quad + \frac{1}{2}\left(Z+W\right)'\Sigma^{-1}\left(y - X\alpha - Wf_\theta\right), \\
D &\equiv \frac{d^2\ln p\left(f_\theta|., y\right)}{\left(df_\theta\right)\left(df_\theta\right)'} = D_1 + D_2 \\
D_1 &= -H'H - \frac{1}{2}Z'\Sigma^{-1}Z - \frac{1}{2}W'\Sigma^{-1}W, \\
D_2 &= -\frac{1}{2}\left(Z-W\right)'\Sigma^{-1}W - \frac{1}{2}W'\Sigma^{-1}\left(Z-W\right), \\
Z &= Y\left(I_{rT} \otimes A_h\right) + W,
\end{aligned}
$$

where $Y = diag\left(\left(y_1 - x_1\alpha_1\right)', ..., \left(y_T - x_T\alpha_T\right)'\right)$, $\Sigma = diag\left(h_1', \ldots, h_T'\right)$,

$$
W = \begin{bmatrix} x_1A_\theta & & \\ & \ddots & \\ & & x_TA_\theta \end{bmatrix}, \quad H = \begin{bmatrix} I_r & & & \\ -I_r & I_r & & \\ & \ddots & \ddots & \\ & & -I_r & I_r \end{bmatrix},
$$

Observe that given this, a standard Newton-Raphson algorithm could be constructed by updating

$$
\widehat{f}_\theta^{(j+1)} = \widehat{f}_\theta^{(j)} - D^{-1}d.
$$

However, $-D$ is not guaranteed to be positive definite for all f_θ, and in fact, will only be positive definite in a very small neighborhood around $\widehat{f}_\theta$ in many applications.

Thus, using the standard Newton-Raphson scoring algorithm will not work well in practice. Nevertheless, we can construct a similar algorithm by replacing D with D_1.

The advantage of this approach is that D_1 is guaranteed to be positive definite for all f_θ, and therefore, an update from any f_θ will always be an ascent direction. The disadvantage, of course, is that in the neighborhood around the mode where D is positive definite, the convergence may be theoretically slower than what is achieved by standard Newton-Raphson. However, even this drawback may be small to the extent that $E_y(D_2) = 0$. In fact, D_1 is closely related to the *Fisher information matrix*

$$F = -H'H - \frac{1}{2}\left(I_{rT} \otimes A_h' A_h\right) - W'\Sigma^{-1}W,$$

which is sometimes used to construct scoring algorithms. Using either F or D will guarantee positive ascent for any value of f_θ; we prefer D_1 as it appears to yield faster convergence in practice. Finally, note that D, D_1 and F are all sparse, banded matrices which results in fast computation of updates even in large dimensions.

B Appendix 2: Derivation of posterior terms

In this appendix we define the terms in the conditional posteriors presented in Section 3 for a_α and a_h in both specifications and f_h and f_α in Specification 2. Each of these parameters has a normal prior of the form $a_\alpha \sim \mathcal{N}(0, \underline{V}_\alpha)$, $f_\alpha \sim \mathcal{N}\left(0, \underline{V}_{f,\alpha}\right)$, $a_h \sim \mathcal{N}(0, \underline{V}_h)$, and $f_h \sim \mathcal{N}\left(0, \underline{V}_{h,\alpha}\right)$, and a conditional normal posterior.

For Specification 2, recall the model specification in (2) and (3) reproduced here:

$$y_t = x_t \alpha + x_t A_\alpha f_{\alpha,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t),$$

$$f_{\alpha t} = f_{\alpha,t-1} + z_{\alpha,t}, \quad z_{\alpha,t} \sim N(0, I_{r_\alpha}), \quad f_{\alpha,0} = 0,$$

$$\Sigma_t = diag\left(e^{h_{1,t}}, \ldots, e^{h_{n,t}}\right) \qquad h_t = (h_{1,t}, , \ldots, h_{n,t})'$$

$$h_t = h + A_h f_{h,t},$$

$$f_{h,t} = f_{h,t-1} + z_{h,t} \quad z_{h,t} \sim N(0, I_{r_h}), \quad f_{h,0} = 0,$$

To obtain a simple form for the posterior for $a_\alpha = \left(\alpha', vec(A_\alpha)'\right)'$ we use

$$\begin{aligned} y_t &= x_t \alpha + \left(f'_{\alpha,t} \otimes x_t\right) vec(A_\alpha) + \varepsilon_t \\ &= \left[x_t \quad \left(f'_{\alpha,t} \otimes x_t\right)\right] a_\alpha + \varepsilon_t \end{aligned}$$

Stack y_t over time to form the $Tn \times 1$ vector y, stack the matrices $\left[x_t \quad \left(f'_{\alpha,t} \otimes x_t\right)\right]$ into the $Tn \times kr_\alpha$ matrix X, and similarly stack ε_t into the $Tn \times 1$ vector ε. We can now write the measurement equation as

$$y = X a_\alpha + \varepsilon \text{ where } \varepsilon \sim N(0, \Sigma). \tag{14}$$

Σ is the diagonal matrix in which the $(t+i, t+i)^{th}$ element is the variance of the i^{th}

element of ε_t where $i \in \{1, \ldots, n\}$. With a prior of the form $\mathcal{N}(0, \underline{V}_\alpha)$, the posterior has the form $\mathcal{N}\left(\overline{a}_\alpha, \overline{V}_\alpha\right)$ where $\overline{V}_\alpha = \left[X'\Sigma^{-1}X + \underline{V}_\alpha^{-1}\right]^{-1}$ and $\overline{a}_\alpha = \underline{V}_\alpha X'\Sigma^{-1}y$.

To define the terms in the posterior for the factors $f_{\alpha,t}$ and $f_{h,t}$, we first define the form of the prior covariance matrices $\underline{V}_{f,\alpha}$ and $\underline{V}_{f,h}$. Let the $(r \times 1)$ vector f_t be either $f_{\alpha,t}$ or $f_{h,t}$ such that $r = r_\alpha$ or $r = r_h$ respectively. In generic form then, the state equation for the factors can be written as

$$f_t = f_{t-1} + z_t, \quad z_t \sim N(0, I_r), \quad f_0 = 0.$$

Stack the f_t into the $Tr \times 1$ vector f, similarly stack the z_t into z, and let R be the $(Tr \times Tr)$ differencing matrix,

$$R_r = \begin{bmatrix} I_r & 0 & 0 & 0 \\ -I_r & I_r & 0 & 0 \\ 0 & -I_r & I_r & 0 \\ & & & \\ 0 & 0 & 0 & I_r \end{bmatrix}.$$

We can now write

$$R_r f = z \quad z \sim \mathcal{N}(0, I_{Tr}) \text{ and so } f = R_r^{-1} z \sim \mathcal{N}\left(0, (R_r' R_r)^{-1}\right).$$

This shows that the priors covariance matrices for $f_{\alpha,t}$ and $f_{h,t}$ are $\underline{V}_{f,\alpha} = \left(R_{r_\alpha}' R_{r_\alpha}\right)^{-1}$ and $\underline{V}_{f,h} = \left(R_{r_h}' R_{r_h}\right)^{-1}$ respectively.

To define the terms in the posterior for the $f_{\alpha,t}$, $\mathcal{N}\left(\overline{f}_\alpha, \overline{V}_{f,\alpha}\right)$, we write the measurement equation as

$$\begin{aligned} y_t - x_t\alpha &= x_t A_\alpha f_{\alpha,t} + \varepsilon_t \\ &= \overline{x}_t a_\alpha + \varepsilon_t \text{ where } \overline{x}_t = \begin{bmatrix} x_t & \left(f'_{\alpha,t} \otimes x_t\right) \end{bmatrix}. \end{aligned}$$

Stacking $y_t^f = y_t - x_t\alpha$ over time to form the $Tn \times 1$ vector y^f and defining

$$X^f = \begin{bmatrix} x_1 A_\alpha & 0 & & 0 \\ 0 & x_2 A_\alpha & \cdots & 0 \\ & \vdots & \ddots & \\ 0 & 0 & & x_T A_\alpha \end{bmatrix}$$

we can write $\overline{V}_{f,\alpha} = \left[X'\Sigma^{-1}X + R'_{r_\alpha} R_{r_\alpha}\right]^{-1}$ and $\overline{f}_\alpha = \overline{V}_{f,\alpha} X^{f\prime}\Sigma^{-1} y^f$.

For the vector $a_h = \left(h' \quad vec\left(A_h\right)'\right)'$ with posterior $\mathcal{N}\left(\overline{a}_h, \overline{V}_h\right)$, we apply the transformation from Kim, Shephard and Chib (1998) and condition upon the states s_h to obtain the measurement equation as

$$\begin{aligned} y_t^* &= \ln\left(\varepsilon_t^2 + \overline{c}\right) - m_t = h + A_h f_{h,t} + \varepsilon_t^* \\ &= x_t^* a_h + \varepsilon_t^* \text{ where } x_t^* = \begin{bmatrix} I_n & \left(f'_{h,t} \otimes I_n\right) \end{bmatrix}. \end{aligned}$$

The term $\varepsilon_t^* + m_t$ is normal with mean vector m_t.[4] Let $y^* = \{y_t^*\}$ be the $Tn \times 1$ vector of stacked y_t^*, X^* be the $Tn \times nr_h$ matrix of stacked x_t^*. Finally, let Σ_h be the diagonal matrix in which the $(t+i, t+i)^{th}$ element is the variance of the i^{th} element of ε_t^* where $i \in \{1, \ldots, n\}$. Combining the likelihood with the prior $\mathcal{N}\left(0, \underline{V}_{a_h}\right)$ we can write $\overline{V}_h = \left[X^{*\prime}\Sigma_h^{-1}X^* + \underline{V}_{a_h}^{-1}\right]^{-1}$ and $\overline{a}_h = \underline{V}_h X^{*\prime}\Sigma_h^{-1} y^*$.

[4] The means and variances of the elements of $\varepsilon_t^* + m_t$ depend upon the states in s_h and are presented in Table 4 of Kim, Shephard and Chib (1998).

Finally we define the terms in the conditional posterior for $f_{h,t}$, $\mathcal{N}\left(\overline{f}_h, \overline{V}_{f,h}\right)$. Again conditional upon the states identified in s_h, the measurement equation can be written

$$y_t^{**} = \ln\left(\varepsilon_t^2 + \overline{c}\right) - m_t - h = A_h f_{h,t} + \varepsilon_t^*.$$

Let y^{**} be the $Tn \times 1$ vector of stacked y_t^{**}, X^{**} be the $Tn \times Tr_h$ matrix $\left(I_T \otimes A_h\right)$. Finally, let Σ_h be the diagonal matrix in which the $(t+i, t+i)^{th}$ element is the variance of the i^{th} element of $\varepsilon_t^*; i = 1, \ldots, n$. Combining the likelihood with the prior $\mathcal{N}\left(0, \left(R'_{r_h} R_{r_h}\right)^{-1}\right)$ we can write $\overline{V}_{f,h} = \left[X^{**\prime}\Sigma_h^{-1}X^{**} + R'_{r_h}R_{r_h}\right]^{-1}$ and $\overline{f}_h = \underline{V}_{f,h} X^{**\prime}\Sigma_h^{-1}y^{**}$.

In Specification 1, the conditional posteriors for a_α and a_h have the same form as that in Specification 2 except with $f_{\alpha,t}$ and $f_{h,t}$ replaced by $f_{\theta,t}$.

C Tables and Figures

Table 2: DICs for models specified with $n = 8$ and various combinations of r_α and r_h. All values are relative to the DIC of the constant coefficient model (i.e. $r_\alpha = r_h = 0$).

3 states			5 states			7 states			10 states			12 states		
r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC
3	0	-402	5	0	-422	7	0	-351	10	0	-102	12	0	88
2	1	-443	4	1	-452	6	1	-349	8	2	-269	8	4	-200
1	2	-414	3	2	-486	4	3	-478	6	4	-358	7	5	-333
0	3	-334	2	3	-490	3	4	-475	5	5	-446	6	6	-360
			1	4	-408	1	6	-410	4	6	-483	5	7	-384
			0	5	-338	0	7	-336	2	8	-463	4	8	-442
shared		-263	shared		-68	shared		199	shared		441	shared		769

Table 3: DICs for models specified with $n = 15$ and various combinations of r_α and r_h. All values are relative to the DIC of the constant coefficient model (i.e. $r_\alpha = r_h = 0$).

3 states			5 states			7 states			10 states			12 states		
r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC	r_α	r_h	DIC
3	0	-764	5	0	-766	7	0	-742	10	0	-366	12	0	-140
2	1	-771	4	1	-816	6	1	-688	8	2	-486	8	4	-573
1	2	-711	3	2	-887	4	3	-892	6	4	-697	7	5	-655
0	3	-562	2	3	-851	3	4	-888	5	5	-854	6	6	-800
			1	4	-756	1	6	-698	4	6	-876	5	7	-792
			0	5	-583	0	7	-565	2	8	-800	4	8	-840
									0	10	-545	0	12	-577
shared		-770	shared		-835	shared		-719	shared		-418	shared		199

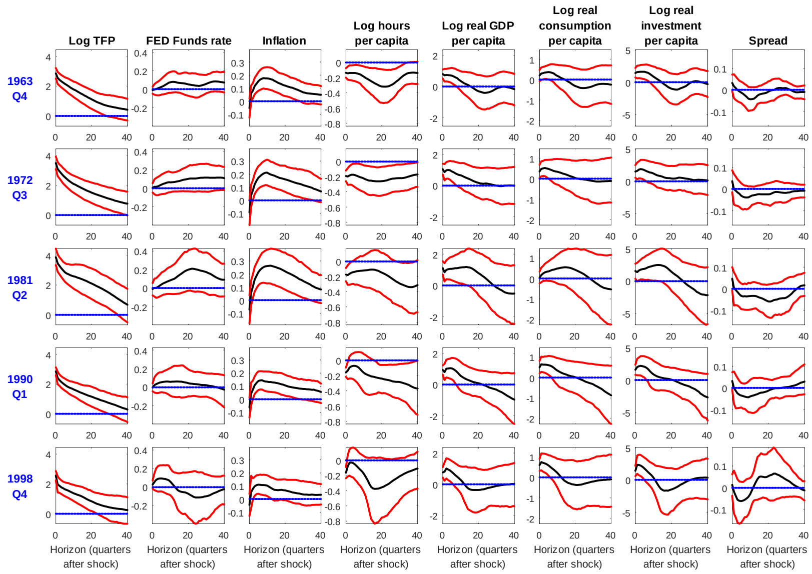

Figure 1: Impulse-response functions to non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

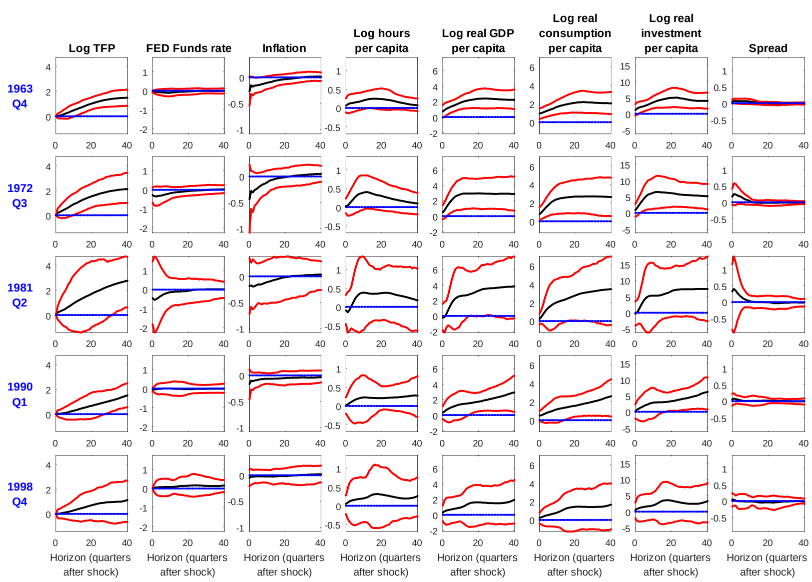

Figure 2: Impulse-response functions to news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

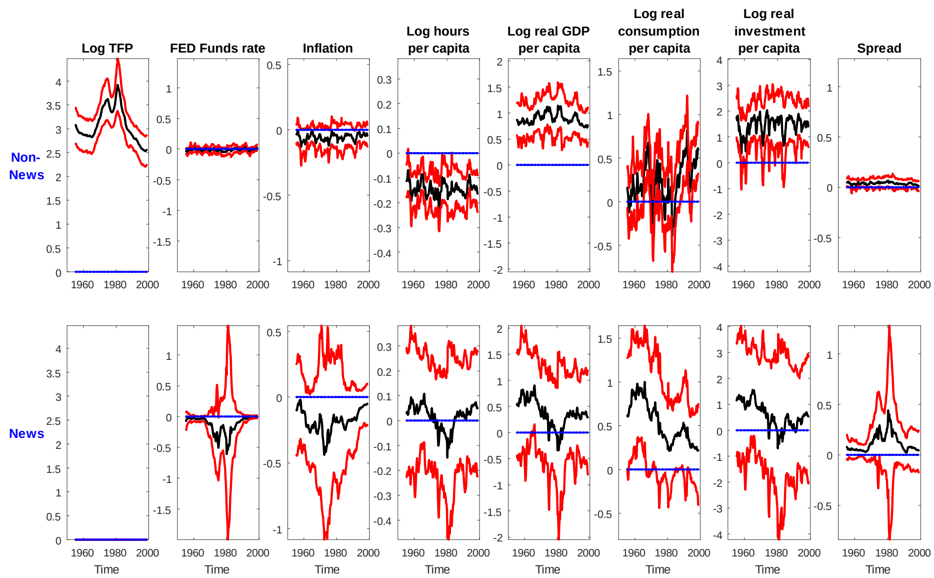

Figure 3: Time-varying responses to non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

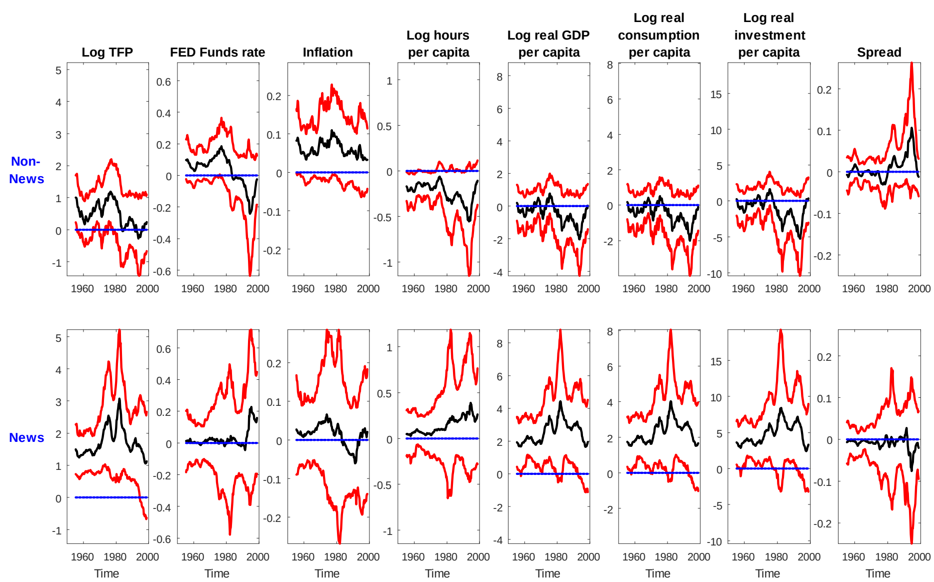

Figure 4: Time-varying responses to non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

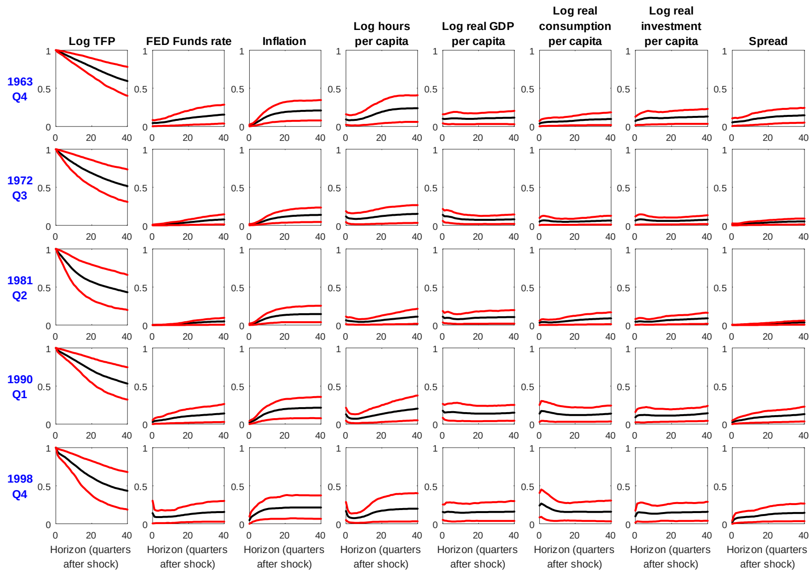

Figure 5: Fractions of forecast error variance explained by non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

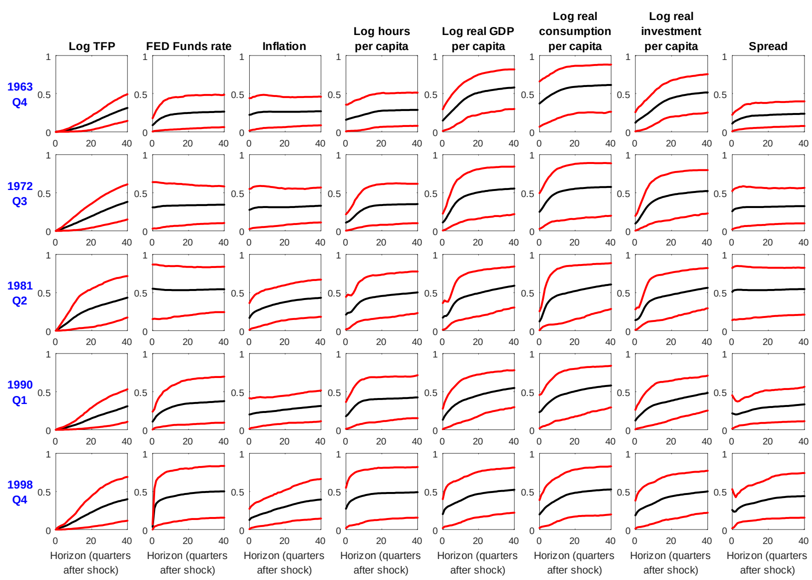

Figure 6: Fractions of forecast error variance explained by news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

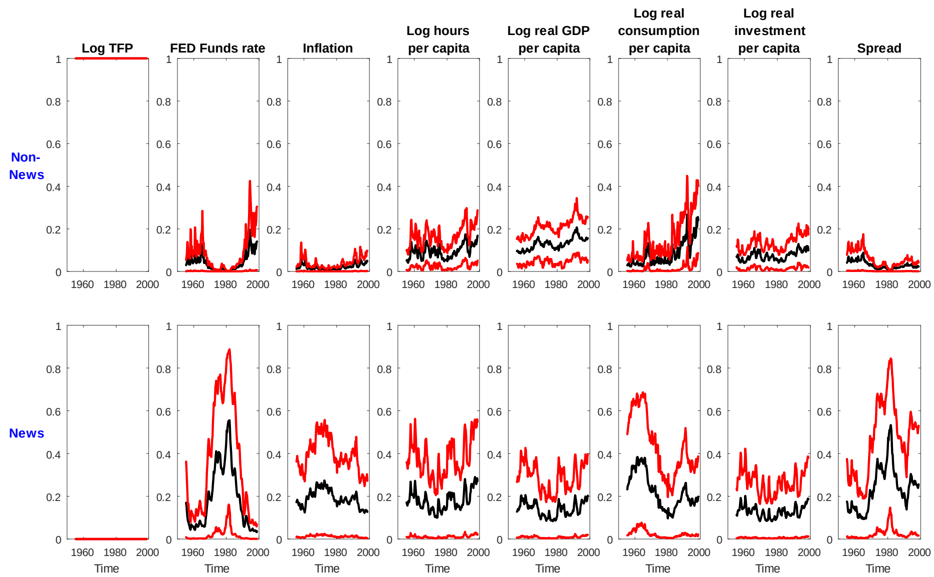

Figure 7: Time-varying fractions of forecast error variance explained by non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

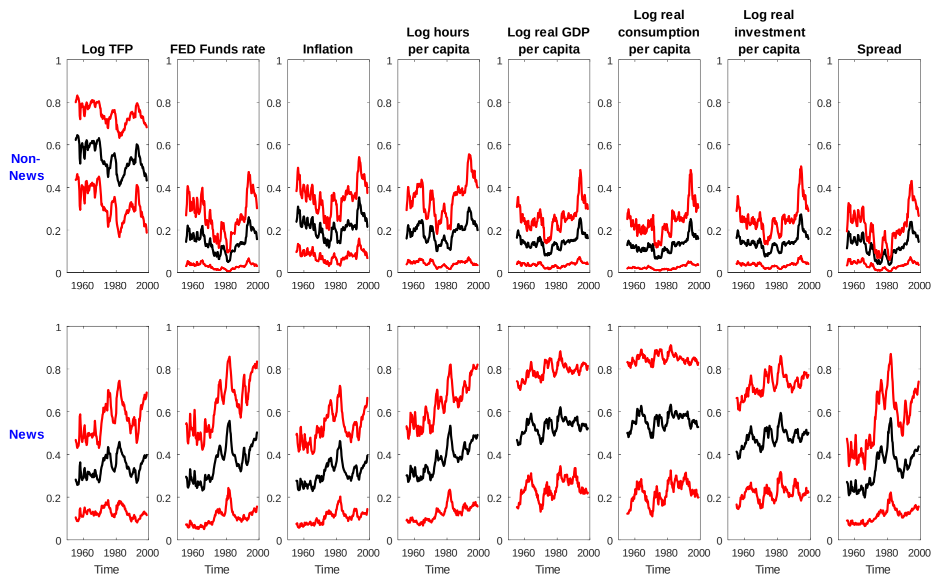

Figure 8: Time-varying fractions of forecast error variance explained by non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 8$ variables model.

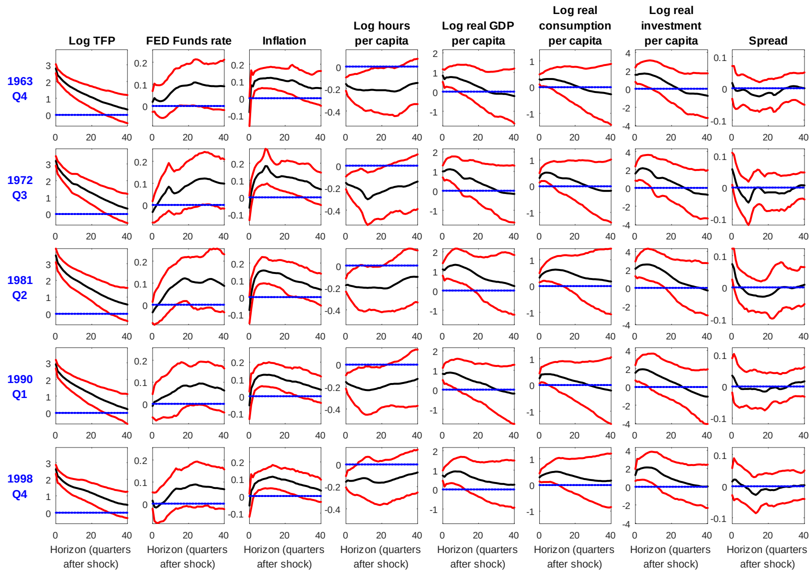

Figure 9: Impulse-response functions to non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

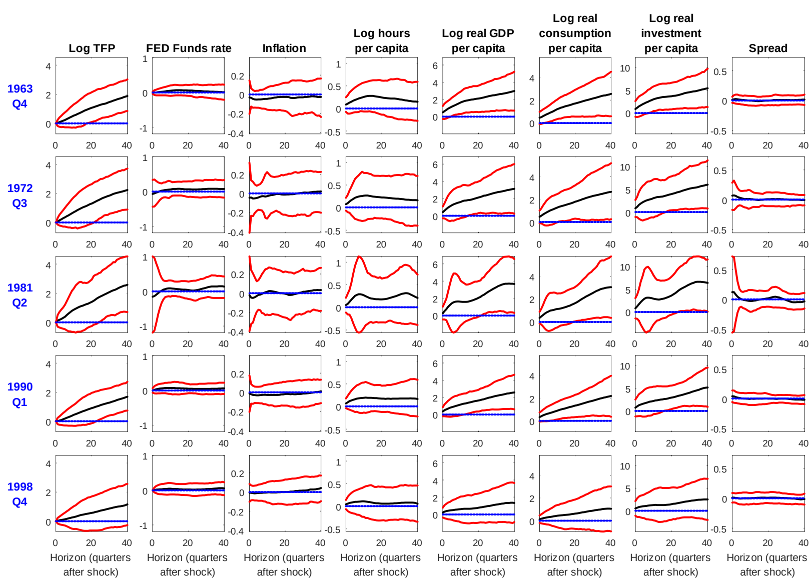

Figure 10: Impulse-response functions to news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

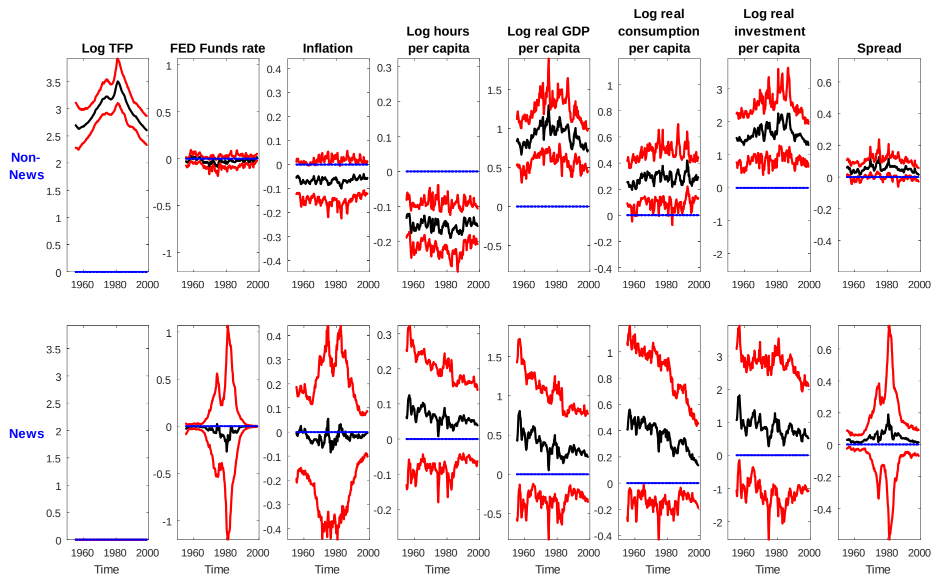

Figure 11: Time-varying responses to non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

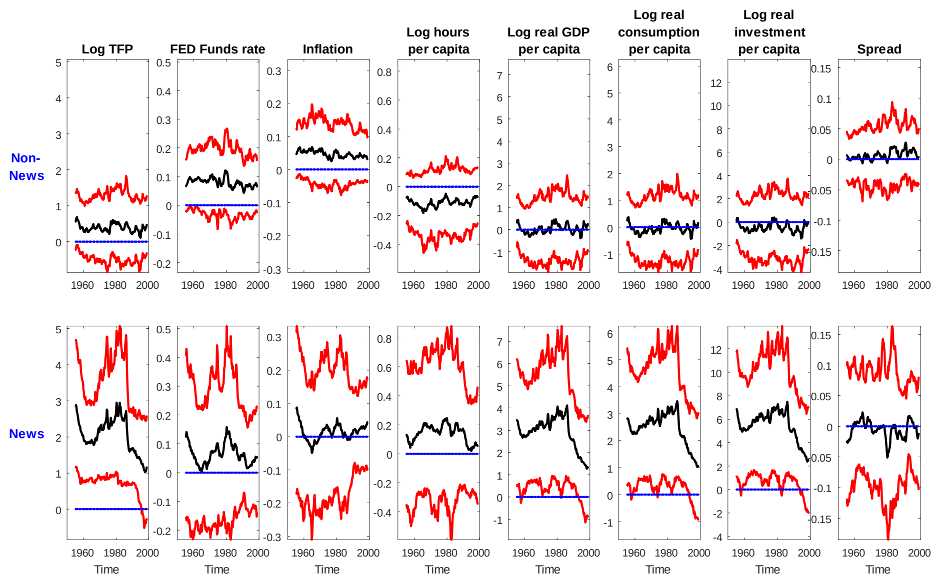

Figure 12: Time-varying responses to non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

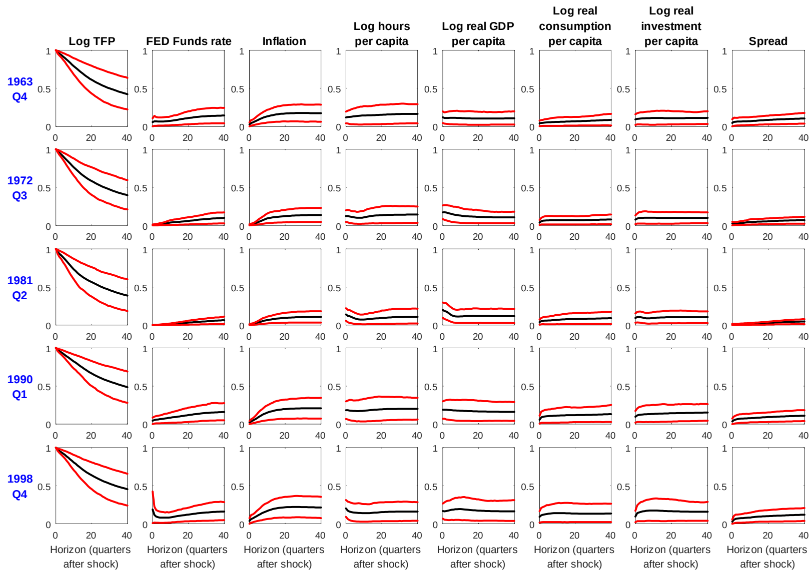

Figure 13: Fractions of forecast error variance explained by non-news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

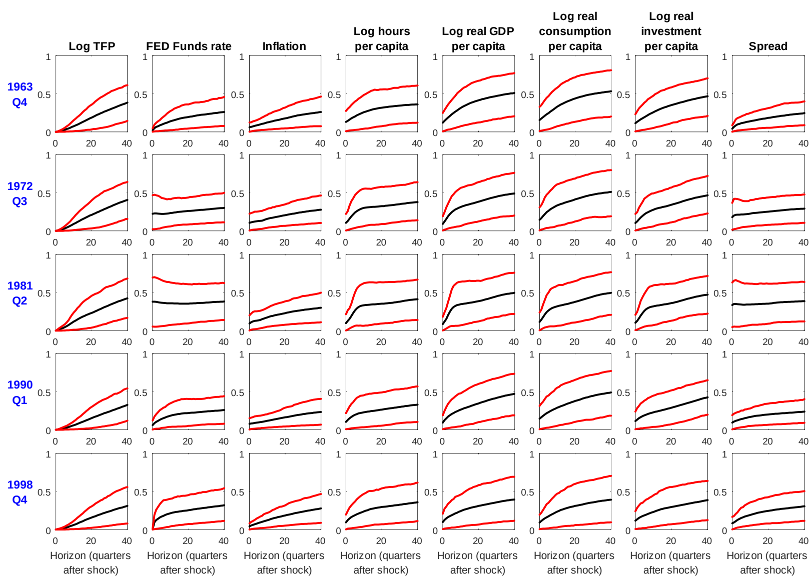

Figure 14: Fractions of forecast error variance explained by news shocks in 1963Q4, 1972Q3, 1981Q2, 1990Q1 and 1998Q4 (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.

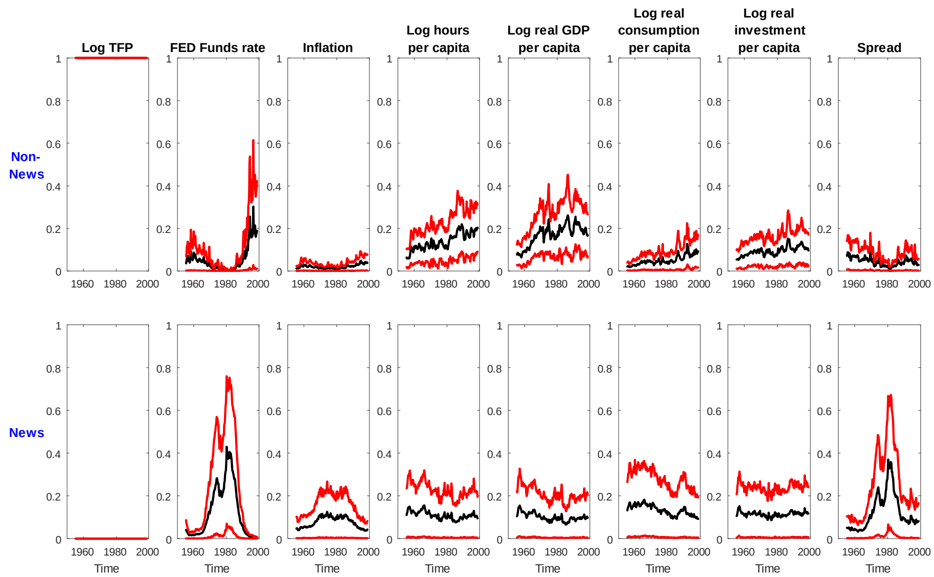

Figure 15: Time-varying fractions of forecast error variance explained by non-news and news shocks on impact (mean, and 16-84 percentiles of the posterior distribution) for the $n=15$ variables model.

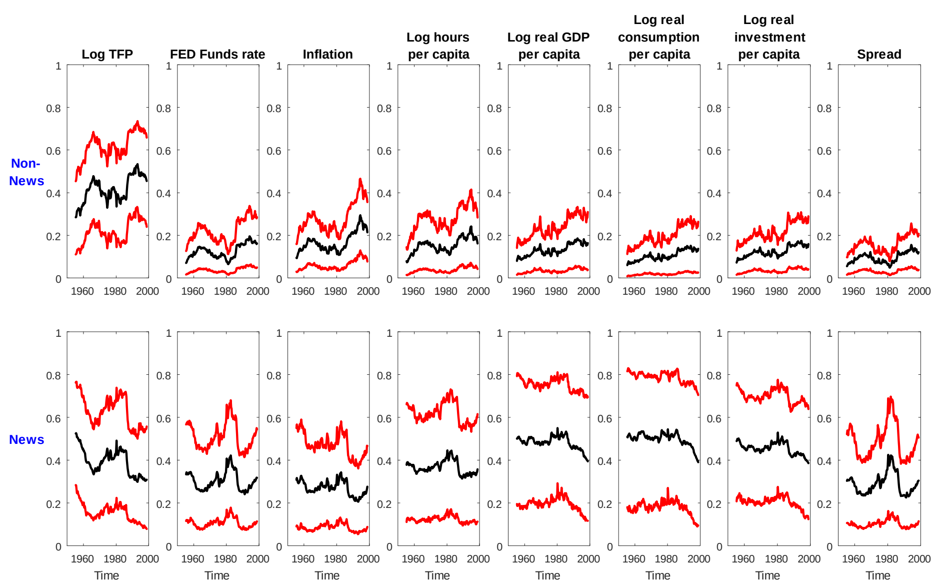

Figure 16: Time-varying fractions of forecast error variance explained by non-news and news shocks at 40 quarters after impact (mean, and 16-84 percentiles of the posterior distribution) for the $n = 15$ variables model.