TVP-VARs: Specification, Estimation, and Model Selection

Guilherme V. Moura[1], Roman Liesenfeld[2]

[1]Departamento de Economia

Universidade Federal de Santa Catarina, Brasil

guilhere.moura@ufsc.br

[2]Institut für Ökonometrie und Statistik

Universität zu Köln, Deutschland

March, 2021

Introduction

- Since Cogley and Sargent (2001, 2005) and Primiceri (2005), TVP-VARs have become standard in analyzing the evolving dynamics of the interdependecies of multiple economic time series.

- However, the usual TVP-VAR speciffication is order-dependent, which makes it not suited for structural analysis, and impacts model selection.

- Additionally, TVP-VARs are heavily parameterized increasing the risk of overfitting.

- We propose the use of shinkage priors in an order-invariant TVP-VAR speciffication and illustrate its performance with an application based on 6 endogenous variables.

Time-Varying Parameters Vector Autoregressions

Consider a general TVP-VAR model:

$$y_t = X_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, H_t^{-1}), \tag{1}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \mathcal{Q}_t), \tag{2}$$

$$\beta_0 \sim \mathcal{N}(\beta, \mathcal{P}_0), \tag{3}$$

Different versions of TVP-VARs often differ in their specification of the dynamics of the matrices H_t and $\mathcal{Q}_t$ in (1) and (2).

Time-Varying Parameters Vector Autoregressions

The standard approach proposed by Primiceri (2005) fixes $\mathcal{Q}_t = Q$ and

$$y_t = X_t \beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, H_t^{-1}), \tag{1}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, Q), \tag{2'}$$

$$\beta_0 \sim \mathcal{N}(\beta, \mathcal{P}_0), , \tag{3}$$

$$H_t^{-1} = A_t^{-1} R_t A_t^{-1'}, \tag{4}$$

where R_t is diagonal with elements, r_{it}, evolving as:

$$\ln r_{it} = \ln r_{it-1} + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{5}$$

and A_t is a lower triangular with ones in its diagonal and non-zero off-diagonals elements, a_{i,j_t}, given by:

$$\alpha_{i,j_t} = \alpha_{i,j_{t-1}} + \zeta_{i,j_t}, \quad \zeta_{i,j_t} \sim \mathcal{N}(0, \omega_{i,j_t}). \tag{6}$$

Order Dependence

In a bivariate version of this TVP-VAR we have

$$R_t = \begin{bmatrix} r_{1t} & 0 \\ 0 & r_{2t} \end{bmatrix}, \qquad A_t^{-1} = \begin{bmatrix} 1 & 0 \\ -a_{21_t} & 1 \end{bmatrix},$$

and thus H_t^{-1} is given by

$$H_t^{-1} = A_t^{-1} R_t A_t^{-1'} = \begin{bmatrix} r_{1t} & -r_{1t}a_{21_t} \\ -r_{1t}a_{21_t} & r_{2t} + r_{1t}a_{21_t}^2 \end{bmatrix}, \tag{7}$$

which shows that the distribution of $h_{1,1,t}^{-1} = r_{1t}$ is lognorrmal, while $h_{2,2,t}^{-1} = r_{2t} + r_{1t}a_{21_t}^2$ is not lognormally distributed.

- Changing the order of the variables in y_t changes the model for each time series in the VAR, thus structural inference, impulse-responses and model comparison are order specific.

Wishart Multivariate Stochastic Volatility

- Uhlig (1994, 1997) developed a state-space model whose latent states are restricted to the manifold of symmetric positive-definite matrices.

- By modeling directly the evolution of precision matrices, this approach is order invariant, and can be used for structural inference in TVP-VAR models (see Bognanni, 2018).

- Additionally, this approach is amenable to forward filtering and backward sampling, and requires a smaller number of parameters.

Wishart Multivariate Stochastic Volatility

- The law of motion for H_t proposed by Uhlig (1997) is given by:

$$H_t = \frac{1}{\lambda} H_{t-1}^{1/2'} \Theta_{t-1} \, H_{t-1}^{1/2}, \qquad \Theta_t \sim \mathcal{B}_k\left(\frac{v}{2}, \frac{1}{2}\right), \tag{8}$$

$$H_1 \sim \mathcal{W}_k(v, \Sigma_0^{-1}/\lambda), \tag{9}$$

- v, determines the variance of the transition density (8) and governs the time-variation of H_t.

- the smaller v is, the larger is the variance of Θ_t and the more the process H_t fluctuates.

- As $v \to \infty$, realizations of Θ_t concentrate around the identity matrix, and the model approaches a homoscedastic VAR.

- The scalar parameter λ can be seen as a discounting factor.

Heteroscedastic β_t

- It is possible to exploit the conjugacy of the Wishart and Normal distributions to account for heteroscedasticity in the dynamics of the VAR parameters in a parsimonious and order invariant way defining $\mathcal{Q}_t$ as:

$$\mathcal{Q}_t = H_t^{-1} \otimes Q, \qquad (10)$$

  where Q here is of lower dimension, and H_t evolves as a Wishart multivariate stochastic volatility.

- The Kronecker structure in (10) means that the covariances between the shocks to β_t have an equation-specific component coming from H_t, as well as a parameter-specific component defined by Q.

Inference in TVP-VAR

- Estimation and inference in TVP-VARs is difficult since they are fundamentally nonlinear.
- Bayesian approach facilitates inference in TVP-VARs as it allows the use of powerful computational algorithms.
- Prior information also helps to discipline the behavior of the model, which is specially relevant in scenarios with a large number of latent variables and parameters.
- However, it is not easy to come out with sensible priors for all elements of a large parameter set.

Priors on the variance of VAR parameters

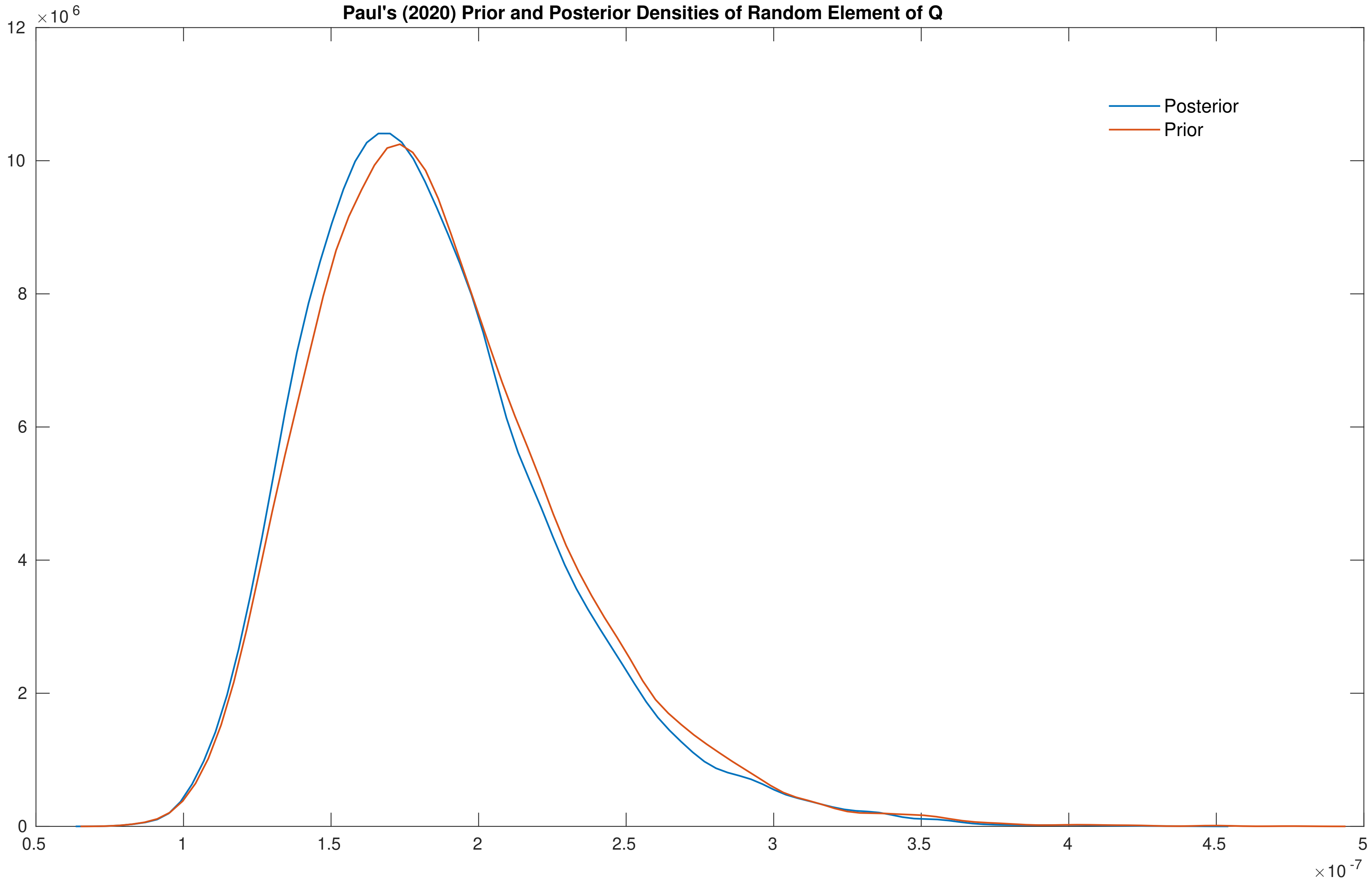
- The flexibility of TVP-VARs comes with the risk of overfitting, which increases with the growing number of coefficients, as many of them might actually be constant.

- However, the usual inverted Wishart prior for $\mathcal{Q}$ implies that all parameters of the mean equation are time-varying.

- Often in the literature $\mathcal{Q}$ is distributed a priori as $i\mathcal{W}(4 \times 10^{-3} I, 40)$, which implies that $P(q_{i,i} < 10^{-5}) = 0$.

- In the context of TVP regressions, Bitto and Frühwirth-Schnatter (2019) show that shrinkage priors can accurately learn from the data which coefficients are time-varying and which are static.

Inverse Wishart Prior



Bitto & Frühwirth-Schnatter Shrinkage Priors

- We use normal-gamma shrinkage priors on the elements of $Q^{1/2}$ and of β to both reduce the number of time-varying parameters and to shrink static parameters towards zero.

- The prior is based on a diagonal Q with diagonal elements q_i, and Gaussian marginal priors for both $q_i^{1/2}$ and β_i along with local and global shrinkage parameters:

$$q_i^{1/2}|\xi_i^2 \sim \mathcal{N}(0, \xi_i^2), \qquad \xi_i^2|a^\xi, \kappa^2 \sim \mathcal{G}(a^\xi, a^\xi \kappa^2/2), \tag{11}$$

$$\beta_i|\tau_i^2 \sim \mathcal{N}(0, \tau_i^2), \qquad \tau_i^2|a^\tau, \lambda^2 \sim \mathcal{G}(a^\tau, a^\tau \lambda^2/2). \tag{12}$$

- $q_i^{1/2} \in \mathbb{R}$ is the positive **and** negative square root of q_i, which facilitates shrinking q_i towards 0.

- To learn a^ξ, a^τ, κ, and λ from the data, we assume that these parameters follow gamma distributions .

Posterior Based on Shrinkage Prior

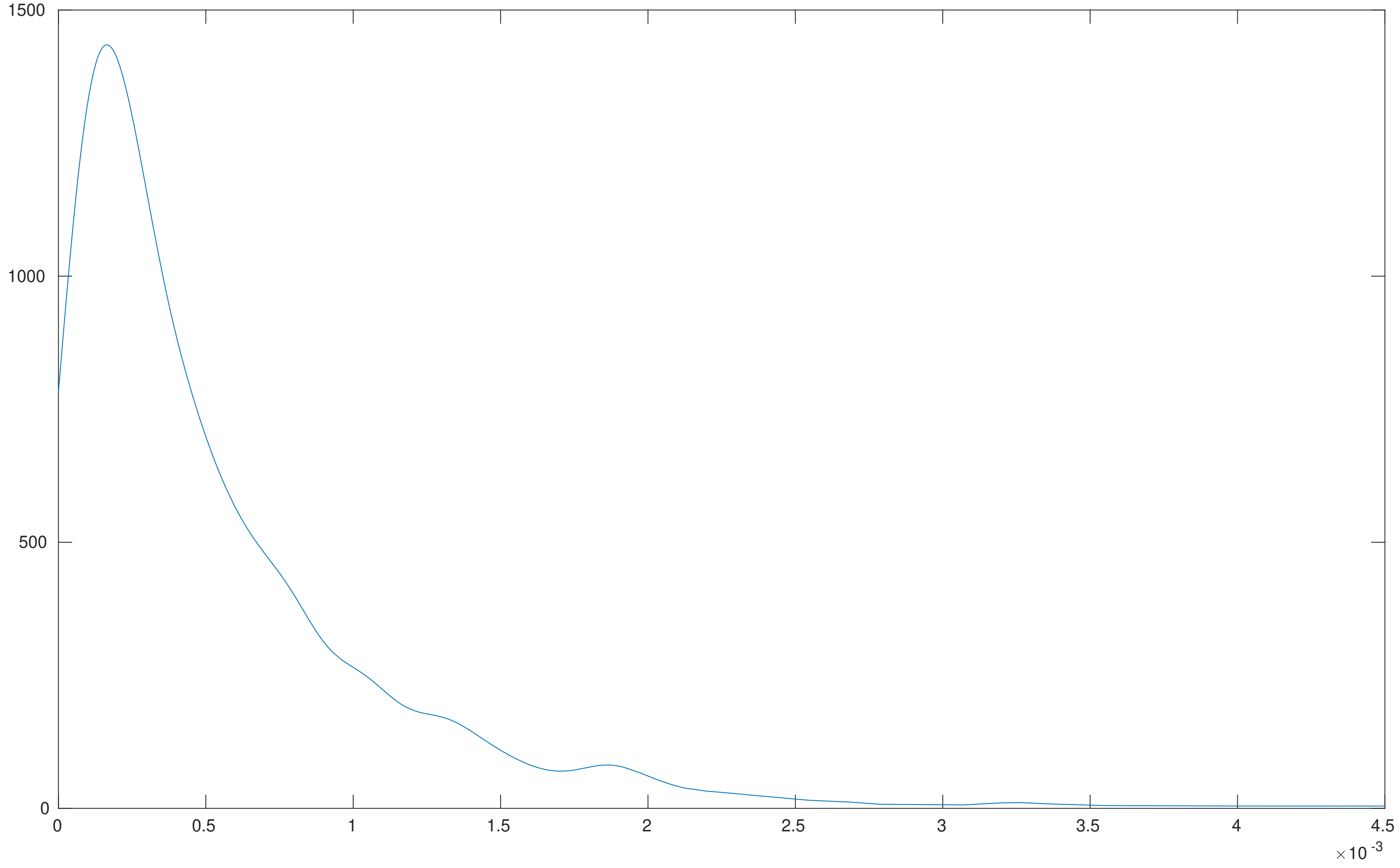


Figure 1: Posterior estimates of a non-zero element elements of Q.

Posterior Based on Shrinkage Prior

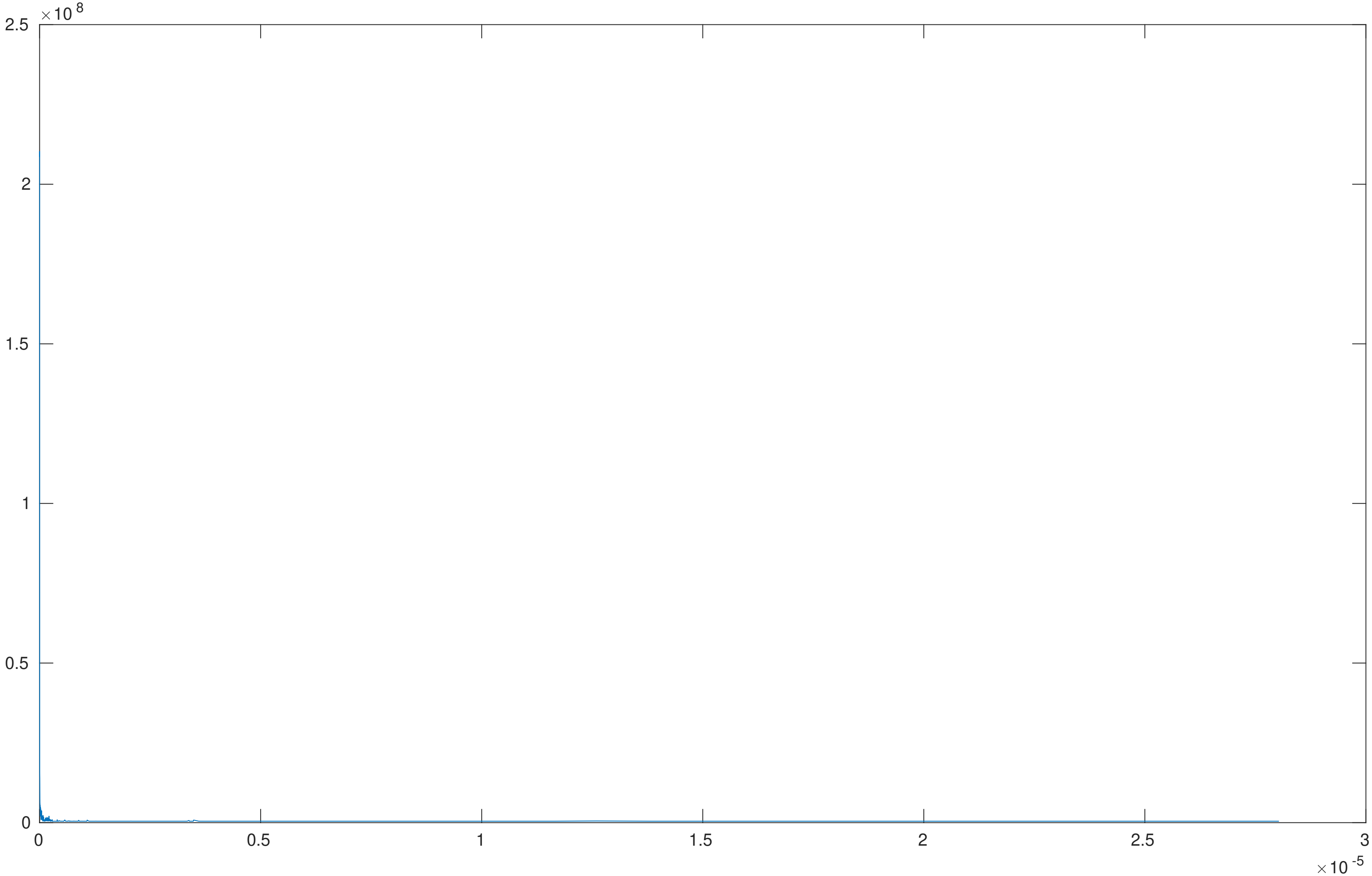


Figure 2: Posterior estimates of an elements of Q which is shrunk to 0.

Effects of Monetary Policy on Stock Market Bubbles

- Gali and Gambetti (2015) argue based on the theory of rational bubbles that a tightening of monetary policy can either increase or decrease asset prices.

- They use a heteroscedastic TVP-VAR model to provide evidence that stock prices end up increasing persistently in response to a contractionary monetary policy.

- Paul (2020) develops a different identification scheme for monetary policy shocks and argues that Gali and Gambetti's result is based on their identification choice.

- Paul (2020) identifies monetary policy shocks via an exogenous variable constructed based on high-frequency surprises.

- However, Paul's TVP-VAR does not account for SV.

Empirical Application

- We use Paul's (2020) application to test different TVP-VAR speciffications.

- Estimation is based on a monthly data from the U.S. consisting of:
  1. the FED funds rate (i_t),
  2. log real stock price index (S&P 500) (q_t),
  3. log real dividends (d_t),
  4. log real S&P/Case-Shiller national home price index (hp_t),
  5. log consumer price index (p_t),
  6. log real industrial production (ip_t),
  7. exogenous surprises (z_t) extracted from 30-day FED funds futures.
- The sample period ranges from Nov. 1988 until Sep. 2017, with 347 observations for each of the 7 time series.

- We fix the lag length of the estimated TVP-VAR models to $p = 3$.

Bayesian Estimation

- The proposed Gibbs algorithm to estimate TVP-VAR models uses the efficient data augmentation scheme based on the ancillarity-sufficiency interweaving strategy of Yu and Meng (2011).

- Moreover, we exploit the fact that, under the Wishart MSV, the conditional posterior of $H_{1:T}$ is available in closed-form and allows sampling of $H_{1:T}$ and the degrees of freedom v in one block through collapsed Gibbs moves.

- The Bayesian posterior analysis is based on 25,000 iterations of the Gibbs sampler and the first 5,000 are discarded as burn-in.

Model Comparison

- We use Paul’s (2020) application to test different TVP-VAR speciffications.

- We analyze 6 different specifications:

  1. VAR-SV, for which $\beta_t = \beta \, \forall t$
  2. TVP-VAR from Paul (2020), for which $H_t = H$ and $\mathcal{Q}_t = Q$,
  3. TVP-VAR with $H_t = H$ and $\mathcal{Q}_t = Q$ and shrinkage priors,
  4. TVP-VAR-SV with $\mathcal{Q}_t = Q$ and H_t following (4),
  5. TVP-VAR-WSV with $\mathcal{Q}_t = Q$ and H_t following Wishart MSV,
  6. TVP-VAR-2WSV where H_t follows (8) and $\mathcal{Q}_t$ follows (10).

- Models are compared using the Log-Predictive Density Scores.

Model Comparison

- Models are compared via Log-Predictive Density Scores (LPDS).
- LPDS can be interpreted as the log marginal likelihood based on a training sample prior.
- Defining the first t_0 observations as $y_{1:t_0}$, the LPDS is computed as the cumulative log predictive density of the remaining sample:

$$\text{LPDS} = \log p(y_{t_0+1:T}|y_{1:t_0}) = \sum_{t=t_0+1}^{T} \log p(y_t|y_{1:t-1}), \tag{13}$$

$$\begin{aligned}\log p(y_t|y_{1:t-1}) = \quad & \int f_\theta(y_t|\beta_t, H_t) f_\theta(\beta_t|\beta_{t-1}, H_t) f_\theta(H_t|H_{t-1}) \\ & \times \pi(\beta_{0:t-1}, H_{1:t-1}, \theta|y_{t_0:t-1}) d\beta_{t0:t} dH_{t1:t} d\theta,\end{aligned}$$

where $\pi(\beta_{0:t-1}, H_{1:t-1}, \theta|y_{t_0:t-1})$ is the posterior density computed with data up to period $t-1$.

Model Comparison

Log Predictive Density Scores

	VAR-SV	TVP-VAR-Paul	TVP-VAR	TVP-VAR-SV	TVP-VAR-WSV	TVP-VAR-2WSV
LPDS	-705.7	-1128.1	-1141.4	-937.8	-769.4	-622.1
MC S.E.	0.523	0.0731	0.485	1.237	0.863	2.038

LPDS computed between Jan. 1993 until Sep. 2017 using data from Nov. 1988 to Dec. 1992 as a training sample.

Time-Varying Coefficients of TVP-VAR-2WSV

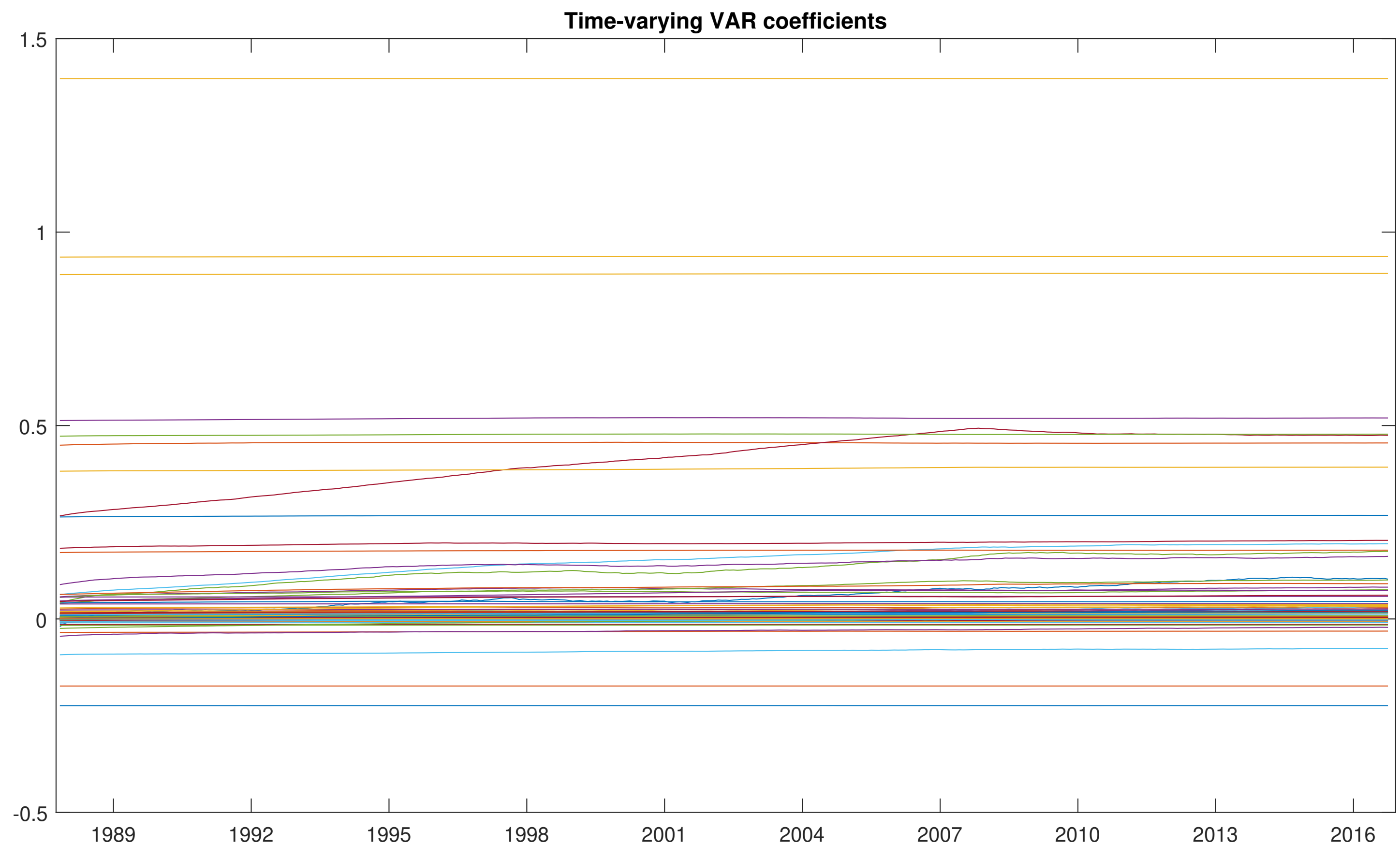


Figure 3: Posterior estimates of β_t.

Stochastic Volatility of TVP-VAR-2WSV

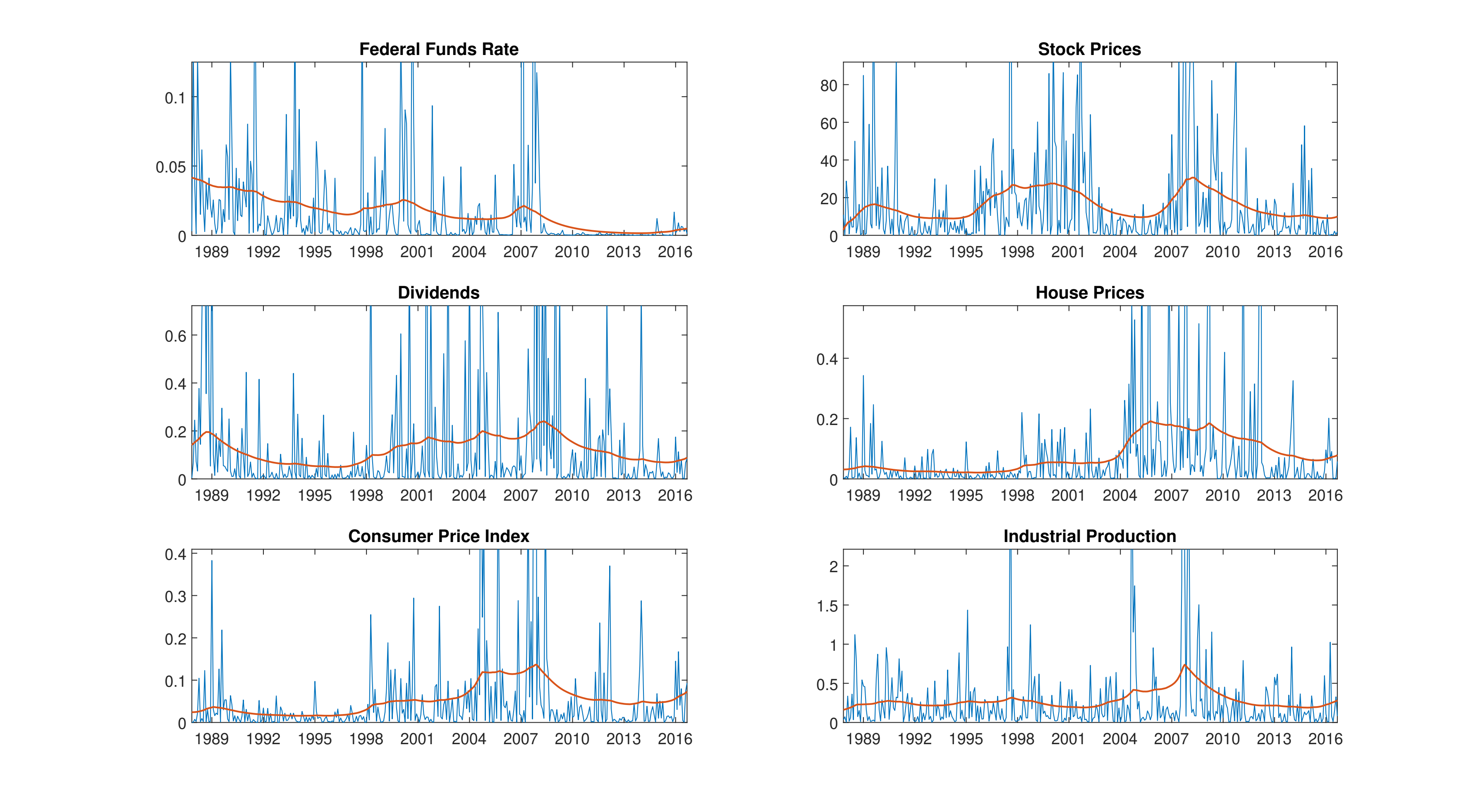


Figure 4: Posterior estimates of ε_t^2 (blue) and of the diagonal elements of H_t^{-1} (red).

Time-Varying Impulse Response Function

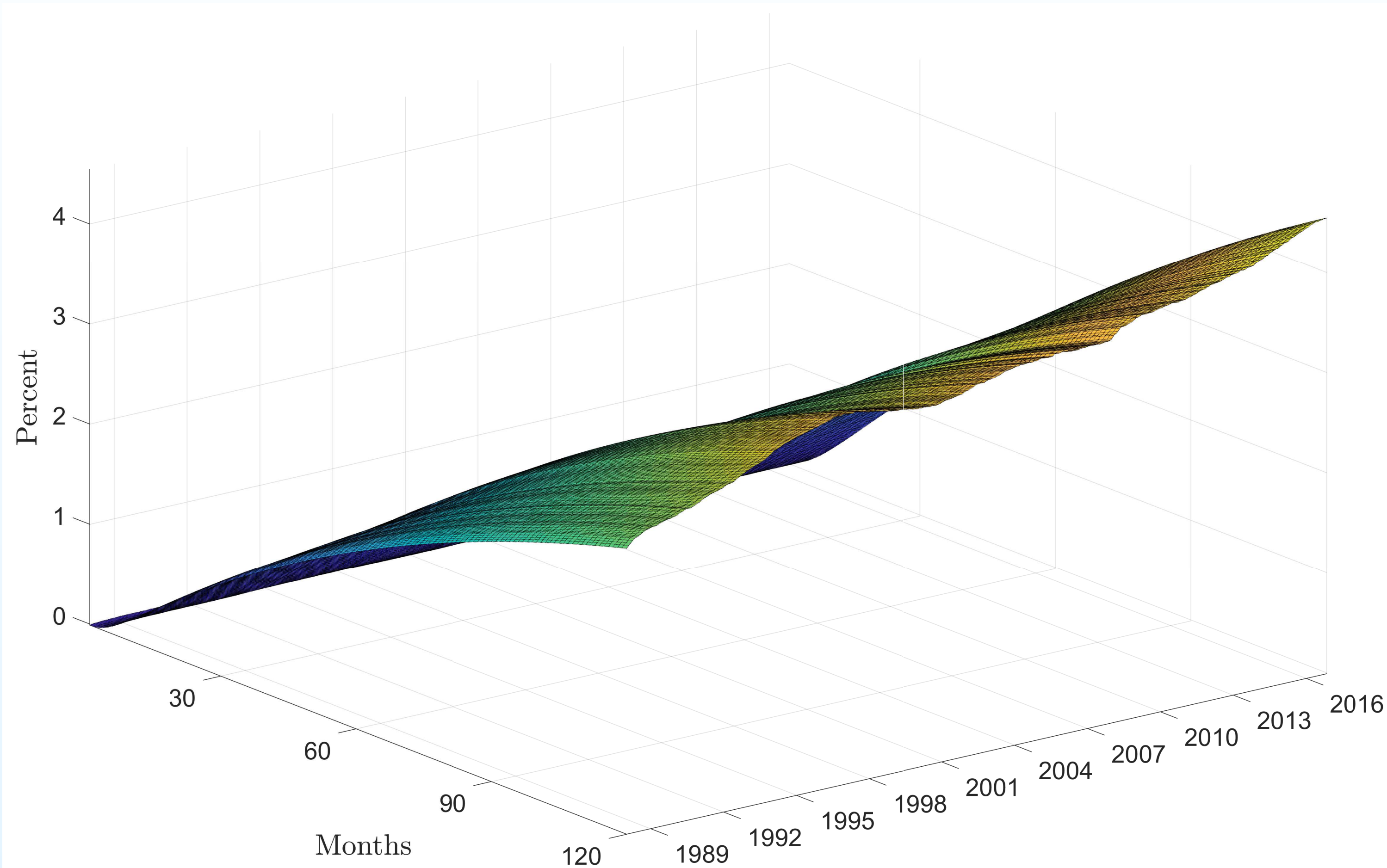


Figure 5: Time-varying effect of a monetary policy shock on stock prices.

Conclusions

- TVP-VARs are very flexible, but the large number of latent variables and parameters used increase the risk of overfitting.
- Specification analysis in TVP-VAR is fundamental to avoid relying on a misspecified model.
- Shrinkage priors are useful to avoid overfitting.
- Wishart multivariate stochastic volatility models are well suited for TVP-VARs, since they yield order invariant models, and capture well the heteroscedasticity present in the data.
- Results similar to those of Galí and Gambetti (2015) can be obtained even with Paul's(2020) identification scheme.