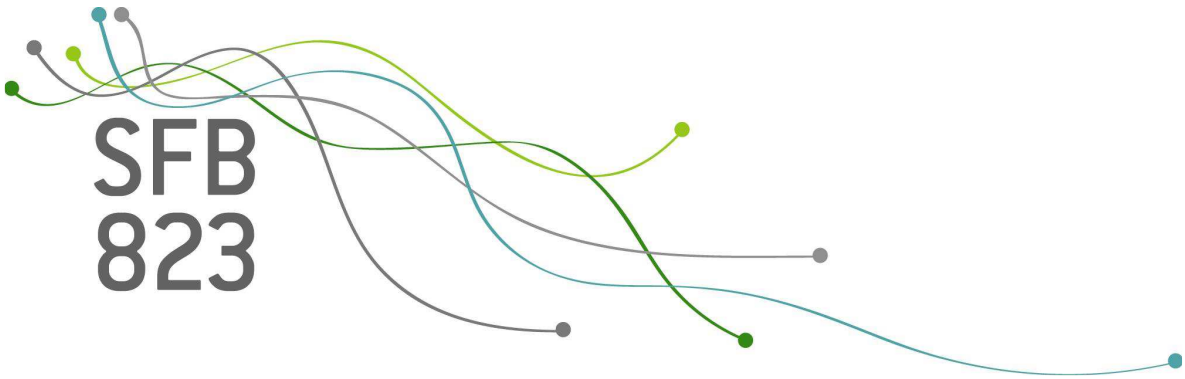
SFB 823

Discussion Paper

A global-local prior for time-varying parameter VARs and monetary policy

Jan Prüser

Nr. 20/2020



A global-local prior for time-varying parameter VARs and Monetary Policy*

Jan Prüser[a]

[a]TU Dortmund†

July 3, 2020

Time-varying parameter VARs have become the workhorse models in empirical macroeconomics. These models are usually equipped with tightly parametrized prior distributions which favor a small and gradual change in parameters. Do such prior distributions suppress some degree of time variation in the VAR coefficients? We address this question by proposing a flexible global-local prior. It turns out that the conventional prior may suppress economically relevant patterns of time variation. Using the global-local prior, we observe that parameter change can be abrupt rather than smooth. We find that, during the chairmanship of Paul Volcker, the Fed has been fighting inflation pressures by raising the interest rate in response to a negative supply shock. However, during the chairmanship of Alan Greenspan, this policy came to an end. In contrast, using the conventional prior, we do not detect this pattern.

Keywords: TVP-VAR, Global-local Prior, Monetary Policy
JEL classification: C11, C32, C54, E52

*I gratefully acknowledge the support of the German Research Foundation (DFG) through SFB 823, within Project A4.

†Fakultät Statistik, 44221 Dortmund, Germany, e-mail: prueser@statistik.tu-dortmund.de

1. Introduction

There is an ongoing debate about the causes of the Great Moderation. Some authors (e.g., Primiceri (2005), Sims and Zha (2006)) emphasized that the variance of the exogenous shocks was higher in the 70s and early 80s than in the rest of the sample. Other authors (e.g., Boivin and Giannoni (2006), Cogley and Sargent (2001) and Lubik and Schorfheide (2004)) emphasize the changes in the transmission mechanism, i.e., the way macroeconomic variables respond to shocks. Particular attention has been given to monetary policy. It is often argued that the reaction of the Fed to inflation has changed over time (e.g., under the Volcker chairmanship, the Fed was more aggressive in fighting inflation pressures). The time-varying parameter VAR (TVP-VAR) of Cogley and Sargent (2005) and Primiceri (2005) allows both the VAR coefficients and the shock size to change over time. This provides a flexible framework for the estimation and interpretation of time variation in the systematic and non-systematic part of monetary policy and their effect on the rest of the economy. Primiceri (2005), Canova and Gambetti (2006), Benati and Mumtaz (2007), Gambetti et al. (2008), Koop et al. (2009) and Chan and Eisenstat (2018) find evidence that points towards a decisive role played by a time-varying shock size.

The TVP-VAR is flexible and can capture many different forms of structural instabilities and the evolving nonlinear relationships between the variables. In order to regularize the degree of time variation of the parameters, TVP-VARs are typically equipped with tightly parametrized prior distributions, favoring a small and constant gradual change in parameters. Is it possible that such prior distributions suppress some degree of time variation in the VAR coefficients? We address this question by proposing a flexible global-local prior distribution to regularize the degree of time variation in the parameters. In addition to a gradual change in parameters, it can also favor abrupt changes of different size if empirically warranted. This flexibility turns out to be important for uncovering economically relevant patterns of time variation. We contribute to the literature by studying the influence of the different prior distributions on the empirical findings. Finally, we compare both prior distributions in a small simulation study to shed some light on the implied dynamics by the different prior.

We find that the conventional prior may suppress some degree of time variation in the coefficients. Using the more flexible global-local prior, we observe that parameter changes can be more abrupt rather than smooth. In particular we provide empirical evidence that the response of monetary policy to supply shocks has changed over time. It turns out that during the chairmanship of Paul Volcker (1979-1987) the Fed has been fighting inflation pressures by raising the interest rate in response to a negative supply shock. During the

chairmanship of Alan Greenspan (1987-2006), this policy came to an end. These changes are not observed with the conventional prior. However, both prior distributions confirm the view that the variance of the exogenous shocks was higher in the 70s and early 80s than in the rest of the sample. Overall, our proposed prior reveals that both the shock size and systematic monetary policy have changed over time.

On a broader level, by proposing a global-local prior for TVP-VARs, we contribute to the literature of modeling TVP-VARs. The prior specification of Koop et al. (2009) can discriminate between very few (but usually large) breaks or many (usually small) breaks in the parameters. They find evidence for gradual change in all of their parameters and reinforce the findings of Primiceri (2005). Koop et al. (2009) use the algorithm of Gerlach et al. (2000) to estimate a single latent indicator for each period to discriminate between time constancy and parameter variation in the autoregressive coefficients, the covariances, and the log-volatilities, respectively. This assumption, however, implies that either all autoregressive parameters change over a given time frame or none of them do. Huber et al. (2019) circumvent this issue by avoiding the computationally intensive simulation of the latent indicators by proposing a straightforward approximation to these indicators. A potential shortcoming of both prior specifications is that they favor either no parameter change or parameter change of a constant size. Thus, both prior specifications assumes two extreme cases: either there are few larger changes or many small changes. As is often the case we argue that the truth lies in between. We find evidence for both: many small breaks and a few large breaks of different size. Another related approach is proposed by Cogley et al. (2010). They use a prior specification that allows the size of parameter change to change gradually over time. Thus, it might not be ideally suited to model cases of abrupt parameter changes. Moreover, it requires a careful parametrization of the prior distributions, see Cogley et al. (2010) for a discussion. In contrast to other approaches, our prior specification does not require any tuning of prior hyperparameters.

The remainder of this paper is organized as follows. Section 2 lays out and discusses the econometric framework. Section 4 provides an overview of the data, contains a simulation study and presents the empirical findings. Section 4 concludes.

2. Econometric Framework

2.1. TVP-VAR

Let $\boldsymbol{y}_t$ be an $n \times 1$ vector of endogenous variables. We follow Chan and Eisenstat (2018) and present the model in structural form:

$$\boldsymbol{B}_{0t}\boldsymbol{y}_t = \boldsymbol{\mu}_t + \boldsymbol{B}_{1t}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{B}_{pt}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

where $\boldsymbol{\mu}_t$ is an $n \times 1$ vector of time-varying intercepts, $\boldsymbol{B}_{1t}, \ldots, \boldsymbol{B}_{pt}$ are $n \times n$ VAR coefficient matrices, $\boldsymbol{B}_{0t}$ is an $n \times n$ lower triangular VAR coefficient matrix with ones on the diagonal and $\boldsymbol{\Sigma}_t$ is a diagonal matrix with time-varying variances. Following Primiceri (2005), Cogley and Sargent (2005), Canova and Gambetti (2006), Benati and Mumtaz (2007), Gambetti et al. (2008), Koop et al. (2009) and Chan and Eisenstat (2018), we set $p = 2$.

Define $\boldsymbol{\beta}_t = \text{vec}((\boldsymbol{\mu}_t, \boldsymbol{B}_{1t}, \ldots, \boldsymbol{B}_{pt})')$ a $k_\beta \times 1$ vector and $\boldsymbol{\gamma}_t$ a $k_\gamma \times 1$ vector containing the free elements of $\boldsymbol{B}_{0t}$ stacked by rows. Note that $k_\beta = n^2p + p$ and $k_\gamma = n(n-1)/2$. Equation (1) can be written as:

$$\boldsymbol{y}_t = \tilde{\boldsymbol{X}}_t\boldsymbol{\beta}_t + \boldsymbol{W}_t\boldsymbol{\gamma}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), \tag{2}$$

where $\tilde{\boldsymbol{X}}_t = \boldsymbol{I}_n \otimes (1, \boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-p})$ and $\boldsymbol{W}_t$ is an $n \times k_\gamma$ matrix that contains appropriate elements of $-\boldsymbol{y}_t$. For example, in our empirical application we have $n = 3$, $\boldsymbol{W}_t$ has the form

$$\boldsymbol{W}_t = \begin{pmatrix} 0 & 0 & 0 \\ -y_{1t} & 0 & 0 \\ 0 & -y_{1t} & -y_{2t} \end{pmatrix}, \tag{3}$$

where y_{it} is the i-th element of $\boldsymbol{y}_t$ for $i = 1, 2$.

Let $\boldsymbol{X}_t = (\tilde{\boldsymbol{X}}_t, \boldsymbol{W}_t)$ and $\boldsymbol{\theta}_t = (\boldsymbol{\beta}'_t, \boldsymbol{\gamma}'_t)'$ be of dimension $k_\theta = k_\beta + k_\gamma$. The model can be written more compactly as

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t). \tag{4}$$

2.2. Hierachical Prior Specification

The estimation of the TVP-VAR is not feasible with the information provided by the likelihood alone. Combining the likelihood with a hierarchical prior makes estimation feasible. The vector of time-varying parameters in turn follows the following random walk process:

$$\theta_{j,t} = \theta_{j,t-1} + \eta_t, \quad \eta_t \sim N(0, V_{\theta_{tj}}), \quad \text{for} \quad j = 1, \ldots, k_\theta. \tag{5}$$

Let $\boldsymbol{\Sigma}_t = \text{diag}(\exp(h_{1t}), \ldots, \exp(h_{nt}))$. The log-volatilities also evolve according to random walks,

$$h_{i,t} = h_{i,t-1} + \xi_t, \quad \xi_t \sim N(0, V_{h_{ti}}), \quad \text{for} \quad i = 1, \ldots, n. \tag{6}$$

This specification allows the parameters to take on a different value in each period. Thereby it is possible to detect structural breaks or regime changes without the need for specifying a fixed number of breaks or regime changes prior to estimation. The prior variances $V_{\theta_{tj}}$ and $V_{h_{ti}}$ are crucial as they control the amount of time variation. Typically, researchers put a tight prior on them and assume them to be constant over time in order to favor gradual changes in the parameters. For a discussion see Cogley and Sargent (2005) and Primiceri (2005). We treat initial conditions $\boldsymbol{\theta}_0$ and $\boldsymbol{h}_0$ as parameters to be estimated and assume the priors for both are Gaussian: $\boldsymbol{\theta}_0 \sim N(\mathbf{0}, 10 \times \boldsymbol{I}_{k_\theta})$ and $\boldsymbol{h}_0 \sim N(\mathbf{0}, 10 \times \boldsymbol{I}_n)$.

2.2.1. Inverse Gamma Prior

In our benchmark model we use the prior specification exactly as in Chan and Eisenstat (2018). This prior specification favors a gradual change (many small breaks) in the parameters. As in most empirical applications (see citations above), the prior variances are assumed to be constant, that is, $V_{\theta_{tj}} = V_{\theta_j}$ and $V_{h_{ti}} = V_{h_i}$, $\forall t$. Both V_{θ_j} and V_{h_i} have inverse gamma priors

$$V_{\theta_j} \sim IG(v_{\theta_j}, S_{\theta_j}), \quad V_{h_i} \sim IG(v_{h_i}, S_{h_i}). \tag{7}$$

The degree of freedom parameters are assumed to be small: $v_{\theta_j} = v_{h_i} = 5$. The scale parameters are set so that the prior mean of V_{θ_j} is 0.01^2 if it is associated with a VAR coefficient and 0.1^2 for an intercept. Similarly, the implied mean of V_{h_i} is 0.1^2.

2.2.2. Horseshoe Prior

In this paper we propose the use of a more flexible prior and set:

$$V_{\theta_{tj}} = \tau_{\theta_j} \lambda_{\theta_{tj}}, \tag{8}$$

$$V_{h_{ti}} = \tau_{h_i} \lambda_{h_{ti}}. \tag{9}$$

The idea is that the global component τ regularizes the overall amount of time variation and the local component λ allows for abrupt breaks. If the local components are close to one for all time periods the prior favors a gradual change. On the other hand, if the local component is larger than one at some point in time it temporarily favors a larger

change. Hence, this prior can favor many small breaks (gradual change), but it can also favor a few large breaks of different size or a mix of both. We will use the data to guide us and estimate the global and local components. Gelman (2006) provides strong arguments for using the half-Cauchy distribution over an inverse gamma distribution for the scale parameters. Polson and Scott (2012) show that it has excellent frequentist risk properties. We therefore follow Carvalho et al. (2010) and use half-Cauchy distributions for the global and local components

$$\sqrt{\tau_{\theta_j}} \sim C^+(0,1), \quad \sqrt{\lambda_{\theta_{tj}}} \sim C^+(0,1), \tag{10}$$

$$\sqrt{\tau_{h_i}} \sim C^+(0,1), \quad \sqrt{\lambda_{h_{ti}}} \sim C^+(0,1). \tag{11}$$

Thus, we propose a global-local prior in form of the horseshoe prior to regularize the degree of time variation.

3. Empirical Results and Simulation

3.1. Data

We use the gross domestic product (GDP) deflator, real GDP and a short-term interest rate as variables for $\boldsymbol{y}_t$. Our sample covers the quarters from 1958Q1 to 2019Q4. The GDP deflator and real GDP are sourced from the Federal Reserve Bank of St. Louis economic database. Both are transformed to annualized growth rates. For the short-term interest rate (denoted by **R** in our figures), a typical choice is the effective federal funds rate. After the financial crisis the effective federal funds rate approaches the zero lower bound. In consequence, it becomes uninformative with respect to the stance of monetary policy. We circumvent this issue by using the shadow rate developed by Wu and Xia (2016) for our main results. The results, however, are similar when we use the effective federal funds rate instead.

3.2. Parameter change: Smooth or abrupt?

Figure 1 shows the estimated VAR coefficients for both prior specifications. For the VAR coefficients we find much more time variation in case of the horseshoe prior as well as more abrupt rather than smooth parameter changes than in case of the inverse gamma prior. Figure 2 shows the volatilities of the residuals for both prior specifications. Using the horseshoe prior, we find that the residual volatility of GDP growth changes quite abruptly from higher to lower volatility from the Great Inflation period to the Great Moderation period. In contrast, this change is substantially smoother for the inverse gamma prior where the change from higher to lower volatility starts in 1980 and ends

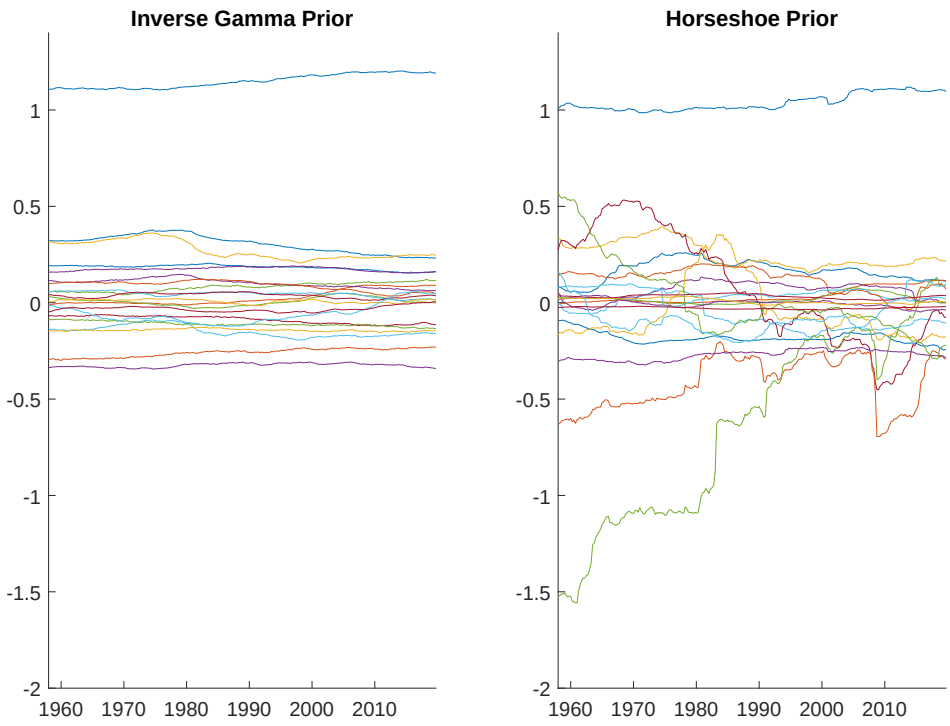


Figure 1: Median VAR coefficients.

in 1985. We find a similar difference between both prior specifications for the residual volatility of the interest rate. However, for inflation we find a smooth change for both prior specifications. These findings demonstrate the flexibility of our proposed horseshoe prior, uncovering various dynamics for the different variables in our analysis.

3.3. Simulation

To shed some light on the implied dynamics by the different priors, we compare both prior specifications in a simulation using one DGP with a gradual change in parameters and one with an abrupt change in parameters. The data for both DGPs are generated from equation (2). We use the point estimates of the TVP-VAR with inverse gamma prior for the first DGP (DGPsmooth) and the point estimates of the TVP-VAR with the horseshoe prior for the second DGP (DGPabrupt). As point estimates we employ the median of the marginal posterior distributions for our main results. The results are similar when using the mean instead of the median. In order to evaluate the simulation results, we calculate the mean squared error (MSE) and mean absolute error (MAE) of the model parameters for both prior distributions and both DGPs using $S = 500$ replications

$$MSE_\theta = \frac{1}{S \times k_\theta \times T} \sum_{s=1}^{S} \sum_{j=1}^{k_\theta} \sum_{t=1}^{T} (\theta_{jt} - \hat{\theta}^r_{jt})^2, \tag{12}$$

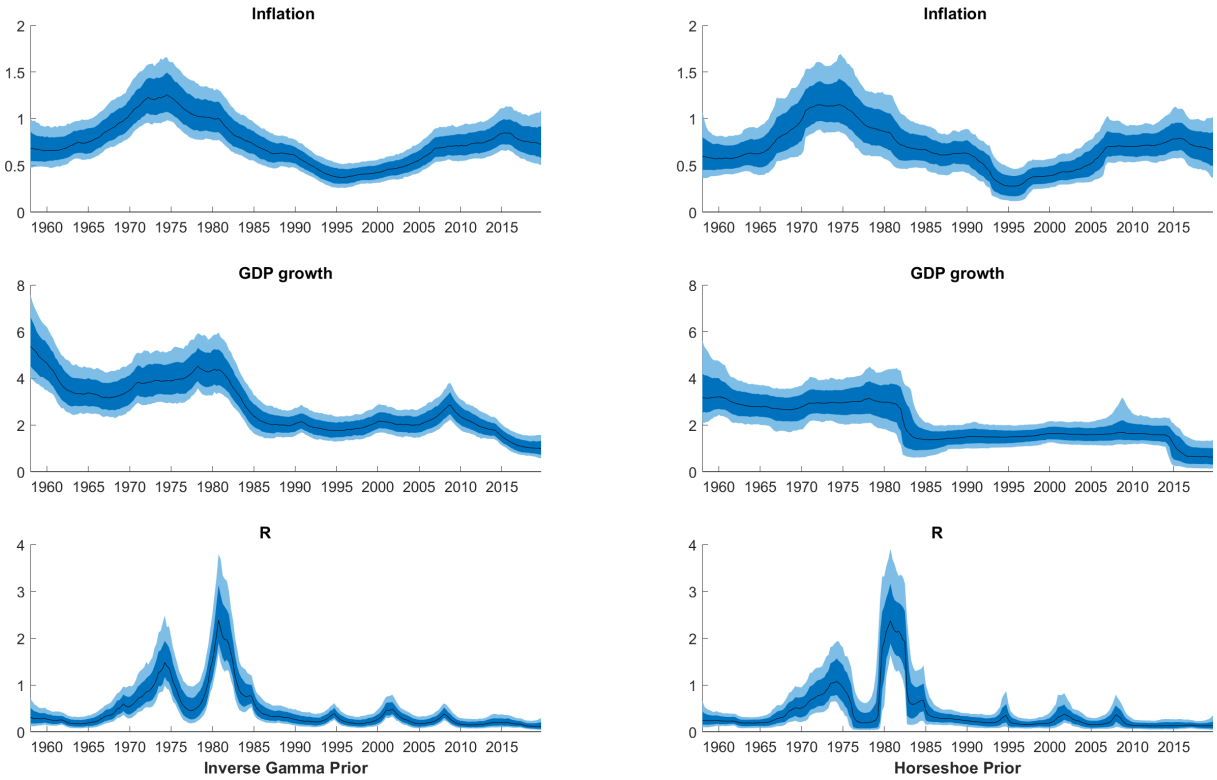


Figure 2: Median residual volatilities with 68% and 90% credible bands.

$$MAE_\theta = \frac{1}{S \times k_\theta \times T} \sum_{s=1}^{S} \sum_{j=1}^{k_\theta} \sum_{t=1}^{T} |\theta_{jt} - \hat{\theta}^r_{jt}|, \tag{13}$$

$$MSE_\sigma = \frac{1}{S \times n \times T} \sum_{s=1}^{S} \sum_{i=1}^{n} \sum_{t=1}^{T} (\sigma_{it} - \hat{\sigma}^r_{it})^2, \tag{14}$$

$$MAE_\sigma = \frac{1}{S \times n \times T} \sum_{s=1}^{S} \sum_{i=1}^{n} \sum_{t=1}^{T} |\sigma_{it} - \hat{\sigma}^r_{it}|, \tag{15}$$

where $\sigma_{it} = \exp(h_{it}/2)$ and both $\hat{\sigma}_{it}$ and $\hat{\theta}_{jt}$ denote point estimates. For our main results we again use the median of the marginal posterior distributions but the results are similar for the mean of the marginal posterior distributions. The results are documented in Table 1. Unsurprisingly, the inverse gamma prior gives more precise estimates for the smooth DGP and the horseshoe prior gives more precise estimates for the abrupt DGP. Furthermore, the estimation precision seems to be reasonable for both prior specifications. For the θ coefficients, the simulation reveals that, if the DGP is unknown, the inverse gamma prior is associated with a higher estimation risk. In the sense that if the data are generated by the smooth DGP, it has the lowest MSE and MAE. If the data are generated by the abrupt DGP it has the highest MSE and MAE. Thus, the horseshoe prior may be the safer option as if the data are generated from the abrupt DGP (the DGP for which both priors have a higher MSE and MAE), it has both a lower MSE and MAE than the inverse gamma prior. These results hold if we calculate the MSE and MAE of θ by excluding the intercepts. The reason why the MSE and MAE are higher for the intercepts

Table 1: Simulation results

	Horseshoe		Inverse gamma	
	MAE	MSE	MAE	MSE
θ				
DGPabrupt	0.18	0.21	0.31	1.18
DGPsmooth	0.18	0.16	0.11	0.04
θ excluding intercepts				
DGPabrupt	0.11	0.03	0.13	0.05
DGPsmooth	0.11	0.03	0.08	0.02
σ				
DGPabrupt	0.16	0.06	0.18	0.09
DGPsmooth	0.22	0.13	0.17	0.08

The MSE and MAE are defined in equations (12) to (15).

is that they exhibit a large degree of time variation. For the MSE and MAE of σ we find the opposite results. Here the horseshoe prior is associated with a higher estimation risk.

3.4. Impulse Response Functions

We follow the approach suggested in Canova and Nicoló (2002) and Uhlig (2005) in order to identify a demand, supply and monetary policy shock in an economically plausible fashion, and use a common/conventional set of sign restrictions on the contemporaneous responses of the variables in $\boldsymbol{y}_t$. The set of sign restrictions can be found in Table 2. Technically, these sign restrictions are implemented by using the algorithm of Rubio-Ramirez et al. (2010). First, we draw a $n \times n$ matrix, $\boldsymbol{J}$, from independent N(0,1) random variables. Second, calculating $\boldsymbol{Q}$ from the QR decomposition of $\boldsymbol{J}$ provides a candidate structural impact matrix such as $\boldsymbol{A}_{0,t} = \boldsymbol{B}_{0t}^{-1}\boldsymbol{\Sigma}_t^{1/2}\boldsymbol{Q}$. The candidate matrix $\boldsymbol{A}_{0,t}$ is accepted if it satisfies the sign restrictions. Up to this point, the shocks are only set identified. Therefore, we follow the suggestion in Fry and Pagan (2011) and collect for each draw from the posterior 1000 candidates $\boldsymbol{A}_{0,t}$ which satisfy the restrictions. Out of this set of 'admissible models' we select the one with elements closest to the median across these 1000 candidates. For more details see Fry and Pagan (2011).

Time variation in the IRFs can result due to time variation in the shock size and/or due to a time-varying response pattern. In order to study if the response of the Fed to demand or to supply shocks has changed over time we keep the shock size constant. That is, we keep $\boldsymbol{\Sigma}_t^{1/2}$ constant for the calculation of the IRFs by fixing it at its posterior mean. Note however that we have estimated the model by allowing time variation in $\boldsymbol{\Sigma}_t^{1/2}$. Calculations (not shown in this paper) of the IRFs without this restriction reveal clearly

Table 2: Imposed sign restrictions on impact

Variable	Demand Shock	Supply Shock	Monetary Policy Shock
Inflation	+	-	-
GDP growth	+	+	-
Interest rate	+	?	+

The plus/minus sign represents a positive/negative sign restriction and the question mark leaves the sign unrestricted.

that the shock size of all three shocks has changed over time. This finding holds for both prior distributions. That the shock size has changed over time is however not much of a debate. Therefore our aim is to study if the response of the Fed to demand and supply shocks has changed over time and, in addition, if the results change when using the more flexible prior specification.

Figure 3 shows the median response of the Fed to a demand and supply shock for horizons 1 to 20 over the entire sample period and for both prior specifications. Figure A.1 shows the median response of the Fed, along with 68% and 90% credible bands, to a demand shock at specific points in time for both prior specifications. Finally, Figure A.2 shows the median response of the Fed, along with 68% and 90% credible bands, to a supply shock at specific time points for both prior specifications. The figures reveal that the response of the Fed to a demand shock is stronger than to a supply shock throughout the sample. The response of the Fed to a demand shock has changed over time for both prior specifications. In particular, the persistence of the response has changed over time. The change of the persistence over time is a bit stronger for the horseshoe prior. Furthermore, the persistence of the response is, overall, smaller for the horseshoe prior. While for the response of the Fed to a demand shock both prior distributions give roughly similar results, the opposite is the case for the response of the Fed to a supply shock.

When a supply shock is temporary, central banks in general face a trade-off between stabilizing inflation and economic activity. Therefore the Fed can respond in three possible ways: no policy response, policy that stabilizes inflation or policy that stabilizes economic activity. Figure A.1 reveals different results for the two prior specifications. Using the horseshoe prior we discover a time-varying pattern of the Fed's response. At the beginning and at the end of the sample, we find some evidence that the Fed pays more attention to stabilizing economic activity. Most interestingly, we find that, during the chairmanship of Paul Volcker (1979-1987), the Fed fought inflation pressures by raising the interest rate in response to a negative supply shock. Note that we can equivalently say that the Fed has stabilized inflation by lowering the interest rate in response to a positive supply shock which is shown in our figures. Note that this policy started shortly

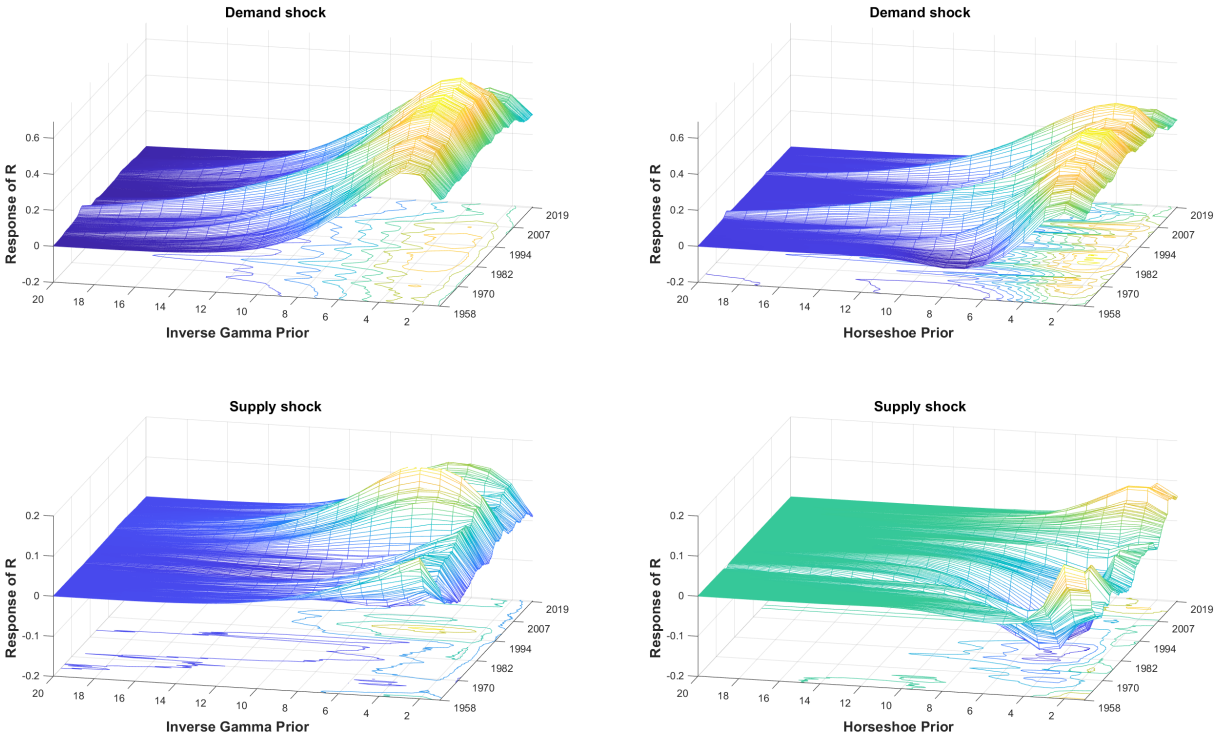


Figure 3: Comparison of median impulse response functions.

before the beginning of the period of the Great Moderation in 1982. During the chairmanship of Alan Greenspan (1987-2006) this policy came to an end. In contrast, using the inverse gamma prior we do not identify this pattern over time. Here we find that the Fed either does not respond to a supply shock or if it does, then it pays more attention to stabilize economic activity. Overall we find stronger empirical evidence that the Fed fought inflation more aggressively under Paul Volcker using the horseshoe prior than for the inverse gamma prior.

The full set of all impulse responses can be found in the online appendix. Here we note that the horseshoe prior reveals a higher degree of time variation also for other impulse response functions. In particular we find that the response of GDP growth to a demand shock and monetary policy shock has changed over time. This pattern of time variation is again suppressed by the inverse gamma prior. In contrast we observe time variation for the response of inflation to a demand shock for both prior specifications. Finally, we observe for both prior specifications that the persistence of the interest rate to a monetary policy shock has smoothly increased over time. This illustrates the great flexibility of the horseshoe prior in terms of the types of parameter change allowed for. And this flexibility does not cost us much in terms of estimation precision in the sense that the credible intervals for the two approaches have similar width or in some cases the credible intervals are even smaller than the credible bands obtained with the inverse gamma prior.

4. Conclusion

We have studied the influence of the prior distribution on the time-varying pattern of parameters in a TVP-VAR. In order to regularize the degree of time variation, researchers typically use tightly parametrized prior distributions favoring a gradual change in the parameters. We find that such priors may suppress some amount of time variation. Using the proposed horseshoe prior we find that parameter changes can be more abrupt rather than smooth. From an economic perspective we find that, during the chairmanship of Paul Volcker, the Fed fought inflation pressures by raising the interest rate in response to a negative supply shock. During the chairmanship of Alan Greenspan this policy came to an end. In contrast, using the inverse gamma prior, we do not observe this pattern over time. Hence, we provide evidence that the prior choice matters for the empirical findings provided by the TVP-VAR and that the conventional inverse gamma prior may suppress economically relevant patterns of time variation.

References

Benati, L. and Mumtaz, H. (2007). US evolving macroeconomic dynamics: A structural investigation. *European Central Bank Working Paper*, (746).

Boivin, J. and Giannoni, M. (2006). Has monetary policy become more effective? *The Review of Economics and Statistics*, 88(3):445–462.

Canova, F. and Gambetti, L. (2006). Structural changes in the US economy: Bad luck or bad policy? *CEPR Discussion Paper No. 5457*.

Canova, F. and Nicoló, G. (2002). Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics*, 49(6):1131–1159.

Carvalho, C., Polson, N., and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chan, J. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532.

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

Cogley, T., Primiceri, G., and Sargent, T. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2(1):43–69.

Cogley, T. and Sargent, T. (2001). Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual*, 16:331–388.

Cogley, T. and Sargent, T. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Fry, R. and Pagan, A. (2011). Sign restrictions in Structural Vector Autoregressions: A critical review. *Journal of Economic Literature*, 49(4):938–960.

Gambetti, L., Pappa, E., and Canova, F. (2008). The structural dynamics of US output and inflation: What explains the changes? *Journal of Money, Credit and Banking*, 40(2-3):369–388.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533.

Gerlach, R., Carter, C., and Kohn, R. (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, 95(451):819–828.

Huber, F., Kastner, G., and Feldkircher, M. (2019). Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models. *Journal of Applied Econometrics*, 34(5):621–640.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393.

Koop, G., Leon-Gonzalez, R., and Strachan, R. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4):997–1017.

Lubik, T. and Schorfheide, F. (2004). Testing for indeterminacy: An application to US monetary policy. *American Economic Review*, 94(1):190–217.

Makalic, E. and Schmidt, D. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

Polson, N. and Scott, J. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Rubio-Ramirez, J., Waggnoner, D., and Zha, T. (2010). Structural Vector Autoregressions: Theory of identification and algorithms of inference. *The Review of Economics Studies*, 77(2):665–696.

Sims, C. and Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, 96(1):54–81.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Wu, J. and Xia, F. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

Appendix A. Figures

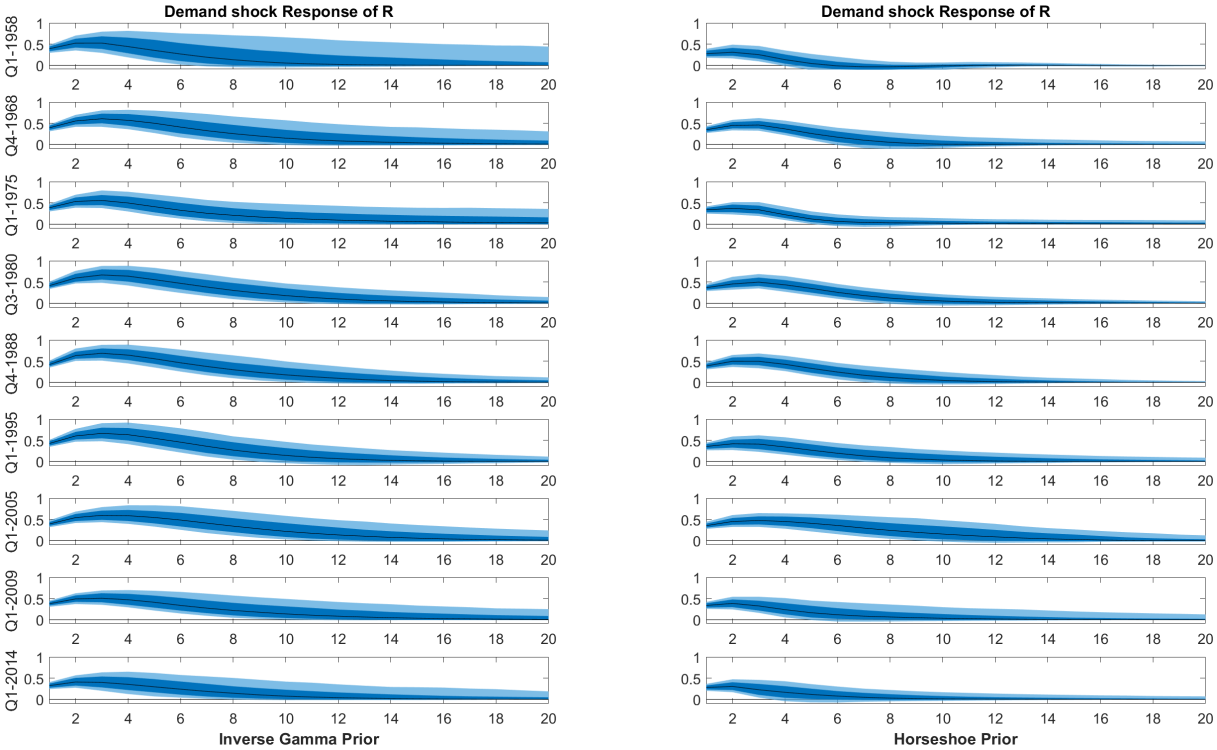


Figure A.1: Comparison of the Fed's median response to a demand shock with 68% and 90% credible bands.

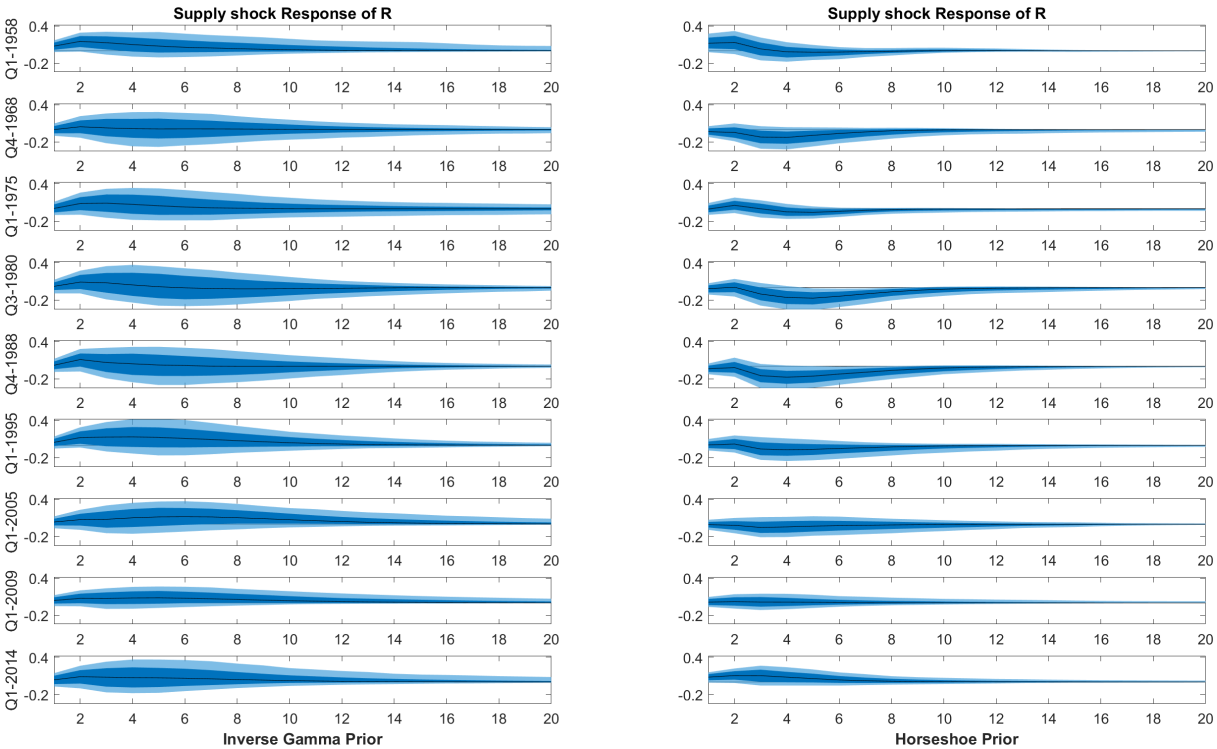


Figure A.2: Comparison of the Fed's median response to a supply shock with 68% and 90% credible bands.

Appendix B. Gibbs-sampler

In order to estimate the TVP-VAR model we use the Gibbs-sampler as described in the appendix of Chan and Eisenstat (2018). Let $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_T')'$, $\boldsymbol{\Sigma}_\theta = \text{diag}(V_{\theta_{11}}, \ldots, V_{\theta_{1k_\theta}}, \ldots, V_{\theta_{T1}}, \ldots, V_{\theta_{Tk_\theta}})$ and $\boldsymbol{\Sigma}_h = \text{diag}(V_{h_{11}}, \ldots, V_{h_{1n}}, \ldots, V_{h_{T1}}, \ldots, V_{h_{Tn}})$. Then a sample of the posterior can be obtained by sequentially drawing from the conditional posterior distributions:

1. $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \boldsymbol{h}_0)$;
2. $p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \boldsymbol{h}_0)$;
3. $p(\boldsymbol{\theta}_0, \boldsymbol{h}_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h)$;
4. $p(\boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\theta}_0, \boldsymbol{h}_0)$.

In order to derive the conditional posterior in step 1 it is useful to write the TVP-VAR in (2) more compactly as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{16}$$

where $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_T')'$, $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$ and $\boldsymbol{X} = \text{diag}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)$. The hierarchical prior in (5) can be written more compactly as

$$\boldsymbol{H}_\theta \boldsymbol{\theta} = \tilde{\boldsymbol{\alpha}}_\theta + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\theta), \tag{17}$$

where $\tilde{\boldsymbol{\alpha}}_\theta = (\boldsymbol{\theta}_0', \boldsymbol{0}, \ldots, \boldsymbol{0})'$ and

$$\boldsymbol{H}_\theta = \begin{pmatrix} \boldsymbol{I}_{k_\theta} & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ -\boldsymbol{I}_{k_\theta} & \boldsymbol{I}_{k_\theta} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \ldots & -\boldsymbol{I}_{k_\theta} & \boldsymbol{I}_{k_\theta} \end{pmatrix}. \tag{18}$$

It follows that the hierarchical prior for $\boldsymbol{\theta}$ follows a Gaussian distribution

$$(\boldsymbol{\theta}|\boldsymbol{\Sigma}_\theta, \boldsymbol{\theta}_0) \sim N(\boldsymbol{\alpha}_\theta, (\boldsymbol{H}_\theta' \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{H}_\theta)^{-1}), \tag{19}$$

where $\boldsymbol{\alpha}_\theta = \boldsymbol{H}_\theta^{-1} \tilde{\boldsymbol{\alpha}}_\theta$. Using standard linear regression results the conditional posterior of $\boldsymbol{\theta}$ can be shown to be

$$(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\theta}_0, \boldsymbol{h}_0) \sim N(\hat{\boldsymbol{\theta}}, \boldsymbol{K}_\theta^{-1}), \tag{20}$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{K}_\theta^{-1} \boldsymbol{d}_\theta$ with

$$\boldsymbol{K}_\theta = \boldsymbol{H}_\theta' \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{H}_\theta + \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}, \quad \boldsymbol{d}_\theta = \boldsymbol{H}_\theta' \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{H}_\theta \boldsymbol{\alpha}_\theta + \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{y}. \tag{21}$$

The precision matrix $\boldsymbol{K}_\theta$ is a band matrix, i.e., the nonzero elements are all confined within a narrow band along the main diagonal. This structure allows us to use the precision sampler of Chan and Jeliazkov (2009) to draw from $N(\hat{\boldsymbol{\theta}}, \boldsymbol{K}_\theta^{-1})$ efficiently.

To implement Step 2, we use the auxiliary mixture sampler of Kim et al. (1998) in combination with the precision sampler of Chan and Jeliazkov (2009) to sequentially draw each slice of $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{iT})'$, $i = 1, \ldots, n$.

The initial conditions $\boldsymbol{\theta}_0$ and $\boldsymbol{h}_0$ are conditionally independent and their conditional posterior is Gaussian:

$$(\boldsymbol{\theta}_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h) \sim N(\hat{\boldsymbol{\theta}}_0, \boldsymbol{K}_{\theta_0}^{-1}), \quad (\boldsymbol{h}_0|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h) \sim N(\hat{\boldsymbol{h}}_0, \boldsymbol{K}_{h_0}^{-1}), \tag{22}$$

where $\boldsymbol{K}_{\theta_0} = \text{diag}(V_{\theta_{11}}, \ldots, V_{\theta_{1k_\theta}})^{-1} + \frac{1}{10}\boldsymbol{I}_{k_\theta}$, $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{K}_{\theta_0}^{-1}\text{diag}(V_{\theta_{11}}, \ldots, V_{\theta_{1k_\theta}})^{-1}\boldsymbol{\theta}_1$, $\boldsymbol{K}_{h_0} = \text{diag}(V_{h11}, \ldots, V_{h_{1n}})^{-1} + \frac{1}{10}\boldsymbol{I}_n$ and $\hat{\boldsymbol{h}}_0 = \boldsymbol{K}_{h_0}^{-1}\text{diag}(V_{h_{11}}, \ldots, V_{h_{1n}})^{-1}\boldsymbol{h}_1$.

Finally we implement step 4 to draw $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{\Sigma}_h$. In the benchmark model equipped with inverse gamma prior the elements of $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{h}_\theta$ are conditionally independent and follow inverse gamma distributions:

$$(V_{\theta_j}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\theta}_0, \boldsymbol{h}_0) \sim IG\left(v_{\theta_j} + \frac{T}{2}, S_{\theta_j} + \frac{1}{2}\sum_{t=1}^{T}(\theta_{jt} - \theta_{j,t-1})^2\right), \quad j = 1, \ldots, k_\theta, \tag{23}$$

$$(V_{h_j}|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{\theta}_0, \boldsymbol{h}_0) \sim IG\left(v_{h_i} + \frac{T}{2}, S_{h_i} + \frac{1}{2}\sum_{t=1}^{T}(h_{it} - h_{i,t-1})^2\right), \quad i = 1, \ldots, n. \tag{24}$$

Estimating the TVP-VAR model equipped with the horseshoe prior requires a modification of the Gibbs-sampler. In order to get known conditional posterior distributions for the elements of $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{\Sigma}_h$ for the horseshoe prior we follow Makalic and Schmidt (2016) and exploit the scale mixture representation of the half-Cauchy distribution. The scalar mixture representation stems from the fact that if X and w are random variables such that $X^2|w \sim IG(\frac{1}{2}, \frac{1}{w})$ and $w \sim IG(\frac{1}{2}, 1)$, then $X \sim C^+(0, 1)$. Since the Gaussian and inverse gamma distributions are conjugate distributions, it is straightforward to derive the posteriors of the hyperparameters.

The conditional posterior distributions of the hyperparameters from the scale mixture representation of the horseshoe prior are conditionally independent and follow inverse

gamma distributions:

$$(\tau_{\theta_j}|\boldsymbol{\theta}, \boldsymbol{\theta}_0, v_{\tau_{\theta_j}}, \lambda_{\theta_{tj}}) \sim IG\left(\frac{T+1}{2}, \frac{1}{v_{\tau_{\theta_j}}} + \frac{1}{2}\sum_{t=1}^{T}\frac{(\theta_{jt} - \theta_{j,t-1})^2}{\lambda_{\theta_{tj}}}\right), \quad j = 1, \ldots, k_\theta, \tag{25}$$

$$(\lambda_{\theta_{tj}}|\boldsymbol{\theta}, \boldsymbol{\theta}_0, v_{\lambda_{\theta_{tj}}}, \tau_{\theta_{tj}}) \sim IG\left(1, \frac{1}{v_{\lambda_{\theta_{tj}}}} + \frac{1}{2}\frac{(\theta_{jt} - \theta_{j,t-1})^2}{\tau_{\theta_j}}\right), \quad j = 1, \ldots, k_\theta, \tag{26}$$

$$(v_{\tau_{\theta_j}}|\tau_{\theta_j}) \sim IG\left(1, 1 + \frac{1}{\tau_{\theta_j}}\right), \quad j = 1, \ldots, k_\theta, \tag{27}$$

$$(v_{\lambda_{\theta_{tj}}}|\lambda_{\theta_{tj}}) \sim IG\left(1, 1 + \frac{1}{\lambda_{\theta_{tj}}}\right), \quad j = 1, \ldots, k_\theta, \tag{28}$$

$$(\tau_{h_i}|\boldsymbol{h}, v_{\tau_{h_i}}, \lambda_{ht_i}) \sim IG\left(\frac{T+1}{2}, \frac{1}{v_{\tau_{h_i}}} + \frac{1}{2}\sum_{t=1}^{T}\frac{(h_{it} - h_{i,t-1})^2}{\lambda_{h_{ti}}}\right), \quad i = 1, \ldots, n, \tag{29}$$

$$(\lambda_{h_{ti}}|\boldsymbol{h}, \boldsymbol{h}_0, v_{\lambda_{h_{ti}}}, \tau_{h_i}) \sim IG\left(1, \frac{1}{v_{\lambda_{h_{ti}}}} + \frac{1}{2}\frac{(h_{it} - h_{i,t-1})^2}{\tau_{h_i}}\right), \quad i = 1, \ldots, n, \tag{30}$$

$$(v_{\tau_{h_i}}|\tau_{h_i}) \sim IG\left(1, 1 + \frac{1}{\tau_{h_i}}\right), \quad i = 1, \ldots, n, \tag{31}$$

$$(v_{\lambda_{h_{ti}}}|\lambda_{h_{ti}}) \sim IG\left(1, 1 + \frac{1}{\lambda_{h_{ti}}}\right), \quad i = 1, \ldots, n. \tag{32}$$

Thus, the implementation of the horseshoe prior is straightforward, as all hyperparameters can be drawn from the inverse gamma distribution. This implies also that the computation burden relative to the inverse gamma prior does not increase much.

Appendix C. Figures for online appendix

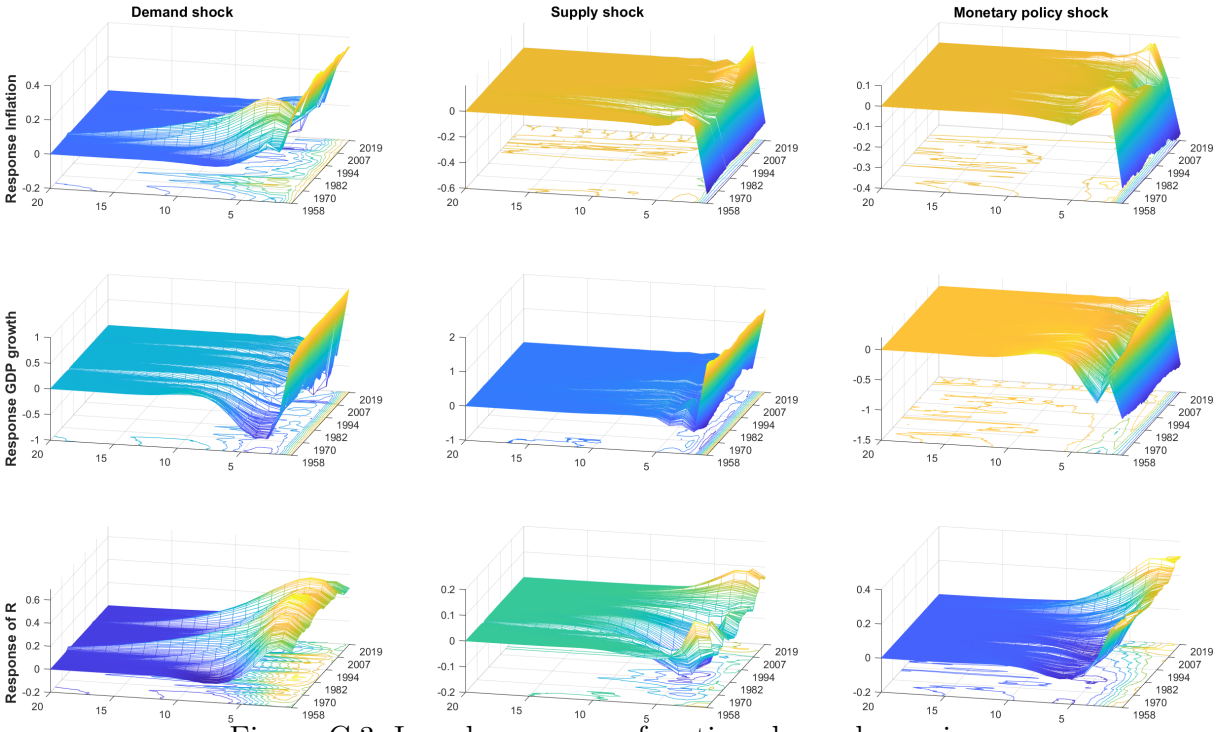


Figure C.3: Impulse response functions horseshoe prior.

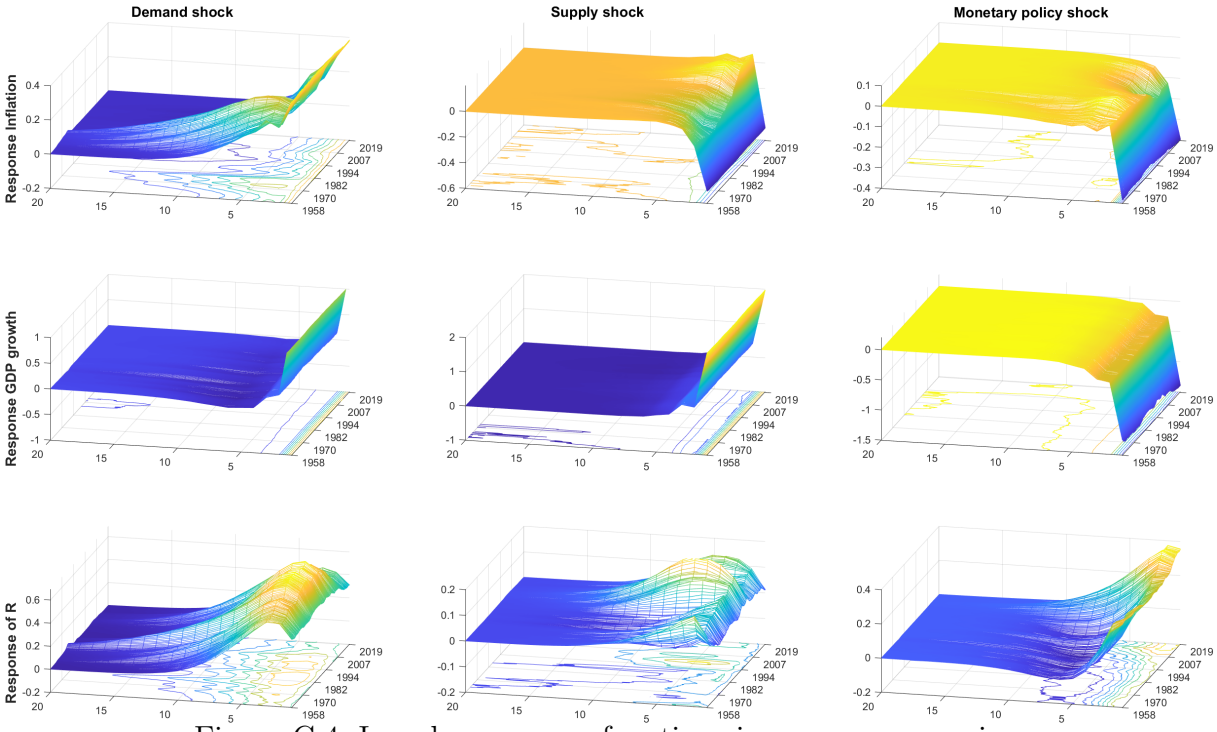


Figure C.4: Impulse response functions inverse gamma prior.

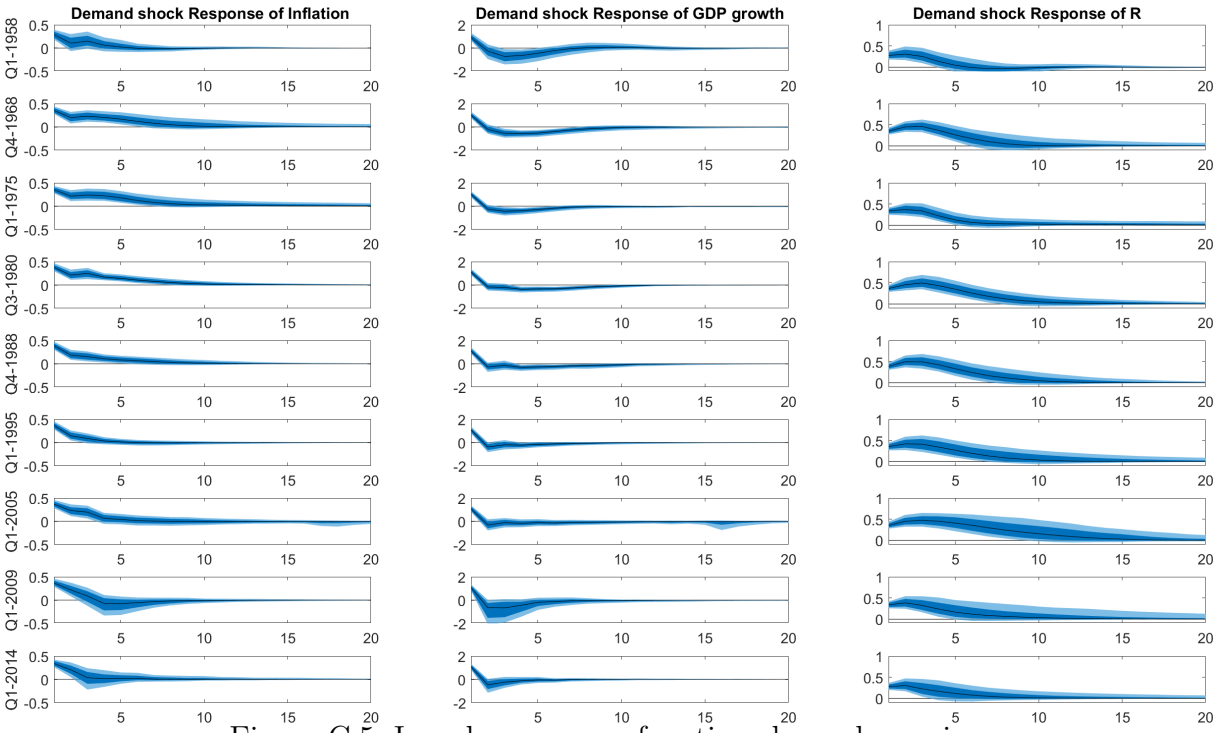


Figure C.5: Impulse response functions horseshoe prior.

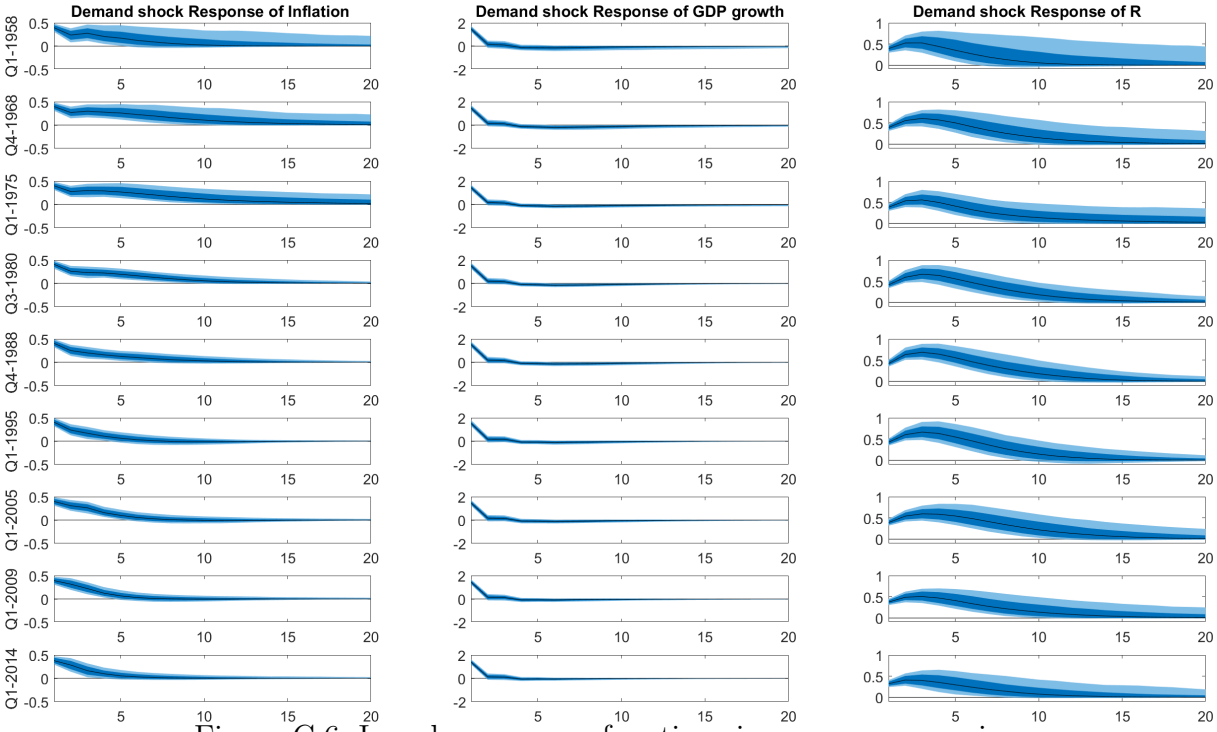


Figure C.6: Impulse response functions inverse gamma prior.

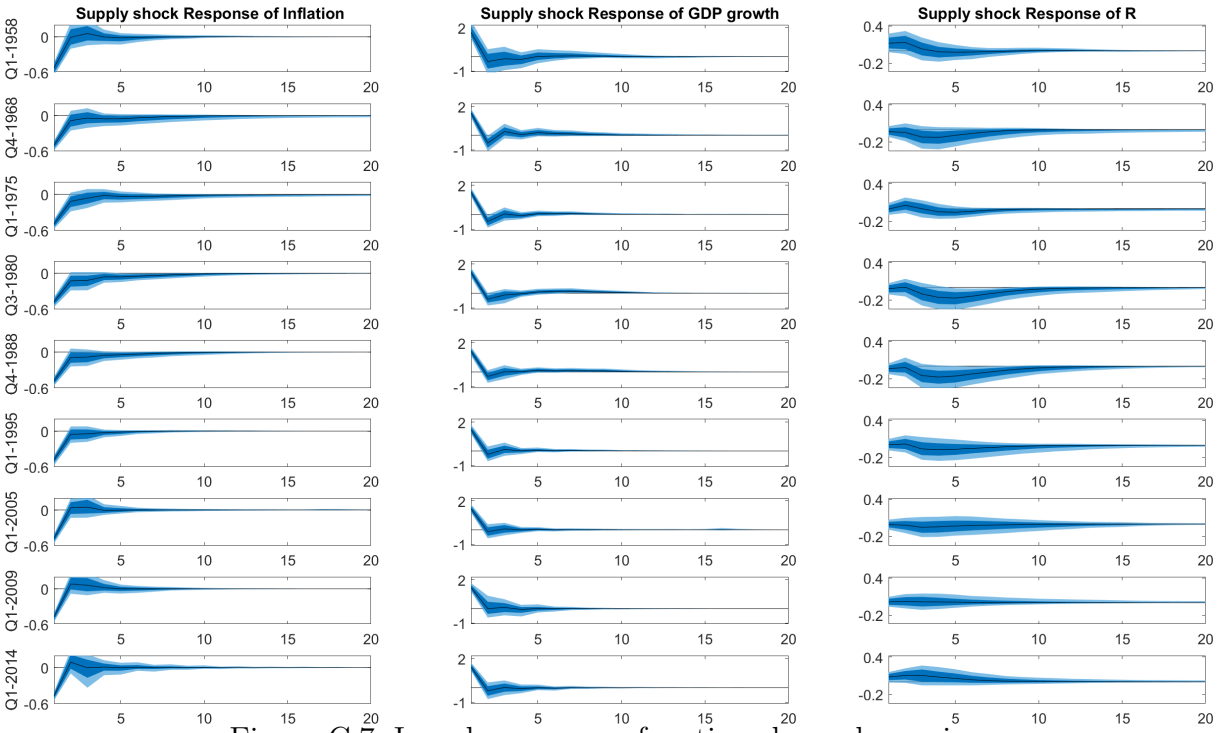


Figure C.7: Impulse response functions horseshoe prior.

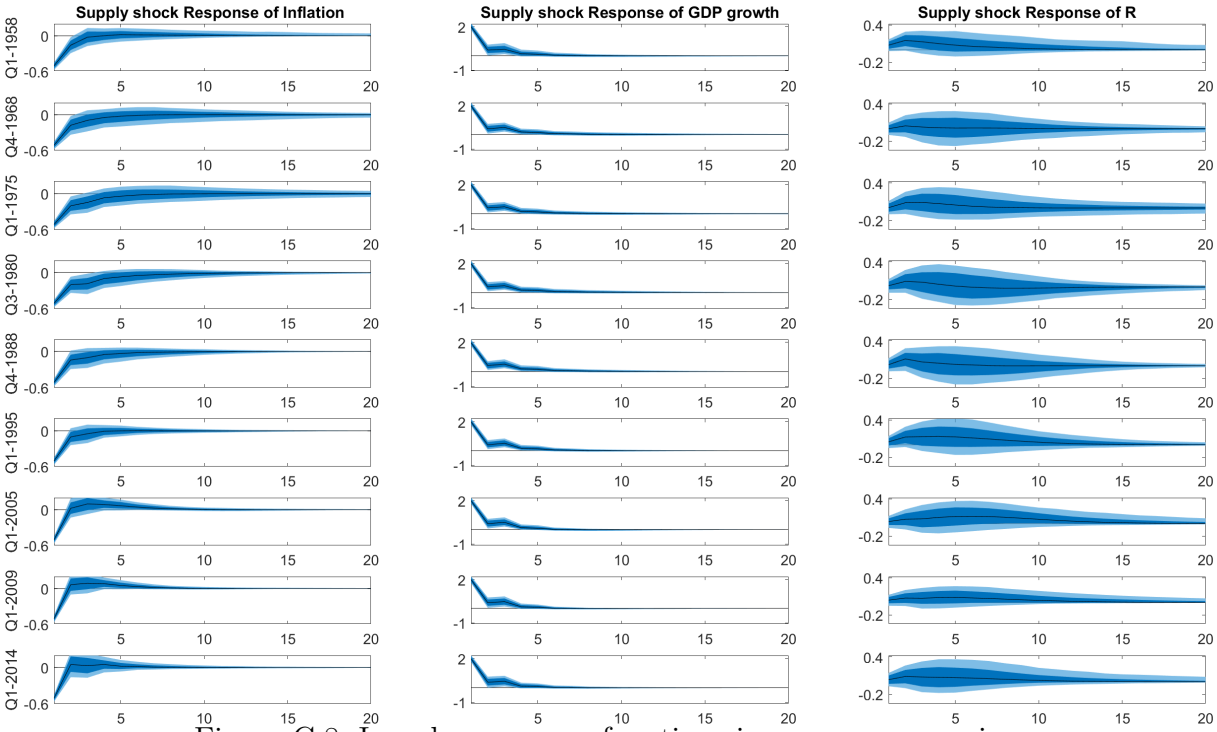


Figure C.8: Impulse response functions inverse gamma prior.

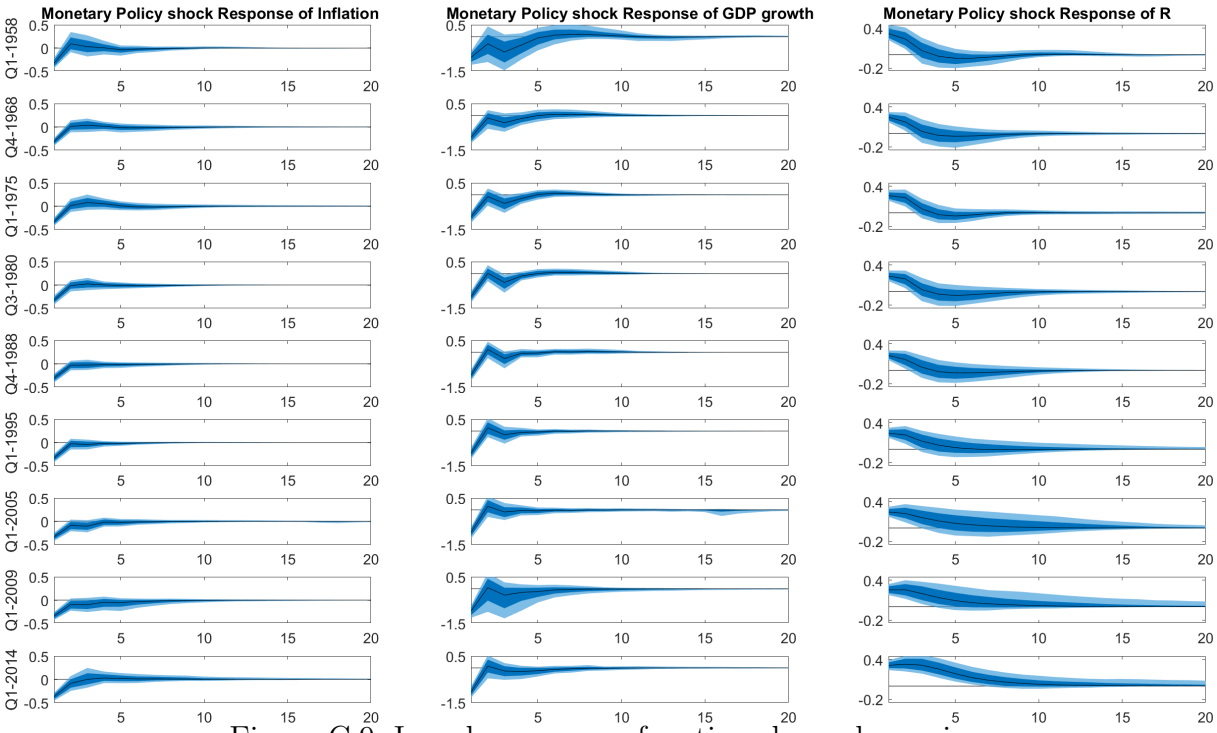


Figure C.9: Impulse response functions horseshoe prior.

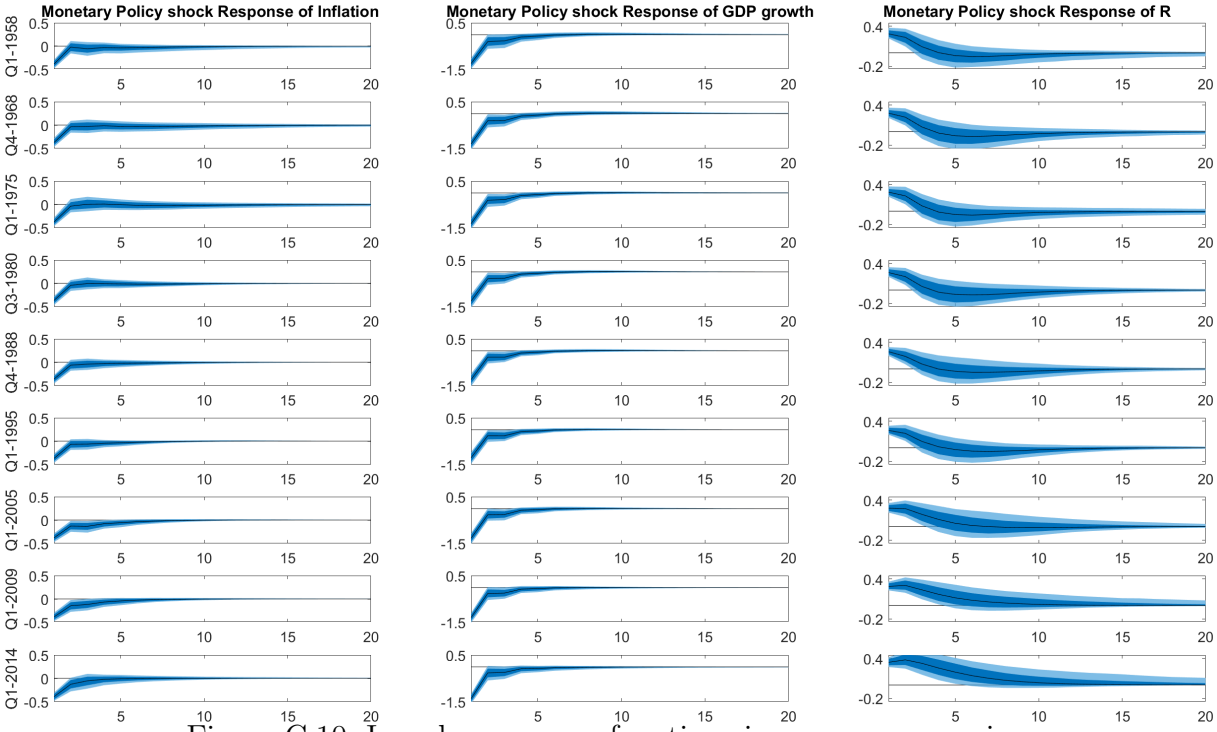


Figure C.10: Impulse response functions inverse gamma prior.