Tractable Bayesian Estimation of Smooth Transition Vector Autoregressive Models

-

ONLINE APPENDIX

Martin Bruns[*]   Michele Piffer[†]

September 4, 2023

---

[*]University of East Anglia, UK, School of Economics.

[†]King's Business School, King's College London, UK. Corresponding author e-mail: m.b.piffer@gmail.com

Contents

List of Figures

List of Tables

A Sampler

The sampler refers to the posterior distribution summarized in the Appendix of the paper. Detailed derivations are available on the authors' webpage.

Conditioning on (γ, c, ψ), generating posterior draws for $(\boldsymbol{\theta}, \Sigma)$ is not a problem: sampling directly from $\tilde{p}(\boldsymbol{\theta}, \Sigma | Y, \gamma, c, \psi)$ is possible as long as the conditioning values for (γ, c, ψ) do not imply identification problems for $(\boldsymbol{\theta}, \Sigma)$. We hence build the posterior sampler as follows. We employ a standard Metropolis-Hastings algorithm to explore $\tilde{p}(\gamma, c, \psi | Y)$, borrowing from the existing literature to set the starting point of the sampler, the tuning parameter and the scaling matrix. Then, we append the chain from the Metropolis-Hastings algorithm with an accept-reject algorithm that directly draws from $\tilde{p}(\boldsymbol{\theta}, \Sigma | Y, \gamma, c, \psi)$.

More precisely, iteration i of the algorithm starts with draws $\left(\gamma^{(i-1)}, c^{(i-1)}, \psi^{(i-1)}\right)$ and proceeds with the following five steps:

Algorithm:

1) draw (γ^*, c^*, ψ^*) from

$$
\begin{pmatrix} \gamma^* \\ c^* \\ \psi^* \end{pmatrix} = \begin{pmatrix} \gamma^{(i-1)} \\ c^{(i-1)} \\ \psi^{(i-1)} \end{pmatrix} + \begin{pmatrix} v_\gamma \\ v_c \\ v_\psi \end{pmatrix}, \tag{A.1}
$$

with (v_γ, v_c, v_ψ) a 3×1 Normal random vector with mean zero and covariance matrix equal to $c_{\text{tune}} \cdot S$, with S a scaling matrix

$$
S = \begin{pmatrix} s_1 & s_4 & s_5 \\ s_4 & s_2 & s_6 \\ s_5 & s_6 & s_3 \end{pmatrix}; \tag{A.2}
$$

2) compute

$$\alpha = \min\left(1, \frac{\tilde{p}(\gamma^*, c^*, \psi^*|Y)}{\tilde{p}(\gamma^{i-1}, c^{i-1}, \psi^{i-1}|Y)}\right), \tag{A.3}$$

then set

$$\begin{pmatrix} \gamma^{(i)} \\ c^{(i)} \\ \psi^{(i)} \end{pmatrix} = \begin{pmatrix} \gamma^* \\ c^* \\ \psi^* \end{pmatrix}, \qquad \text{with probability } \alpha, \tag{A.4}$$

$$\begin{pmatrix} \gamma^{(i)} \\ c^{(i)} \\ \psi^{(i)} \end{pmatrix} = \begin{pmatrix} \gamma^{(i-1)} \\ c^{(i-1)} \\ \psi^{(i-1)} \end{pmatrix}, \qquad \text{with probability } 1 - \alpha; \tag{A.5}$$

3) keep track of how frequently Step 2 accepts the candidate draw (γ^*, c^*, ψ^*) from Step 1. Adjust the tuning parameter c_{tune} adaptively in accordance with the acceptance rate associated with the first i draws, unless the chain has already run for a desired number of draws, n_1;

4) if the chain for (γ, c, ψ) has already been running for n_1 draws

   4a) draw Σ^* from $\tilde{p}(\Sigma|Y, \gamma^{(i)}, c^{(i)}, \psi^{(i)})$;

   4b) draw $\boldsymbol{\theta}^*$ from $\tilde{p}(\boldsymbol{\theta}|Y, \Sigma^*, \gamma^{(i)}, c^{(i)}, \psi^{(i)})$;

   4c) if $\text{I}\{\boldsymbol{\theta}^*\} = 1$ store $(\boldsymbol{\theta}^*, \Sigma^*, \gamma^{(i)}, c^{(i)}, \psi^{(i)})$;

5) return to Step 1, set $i = i+1$ and keep running the chain on (γ, c, ψ) until a desired number of draws n_2 has been stored in Step 4c).

In our application, we run a preliminary importance sampler on $\tilde{p}(\gamma, c, \psi|Y)$ to initialize our sampler. We use the marginal mean of the chain for (γ, c, ψ) generated by the

importance sampler to start our sampler.[1] We then set the scaling matrix S from equation (A.2) equal to the empirical covariance matrix of the preliminary draws for (γ, c, ψ). We then keep S constant throughout the algorithm, initialize c_{tune} at 2, and adaptively adjust it within the first n_1 draws of the algorithm using the same adaptive procedure as in Waggoner et al. (2016). Once the first n_1 draws are generated, the chain proceeds until n_2 draws are generated for $(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi)$. To speed up the sampler, we run multiple chains for Steps 4-5 and exploit parallelization techniques. Standard tests are used to assess the convergence of the chains for (γ, c).

The approach outlined above is particularly tractable in the special case in which the model features homoskedastic innovations, as in the baseline specification of the application in the paper. Under this condition, which implies setting $\psi = 0$, $\tilde{p}(\gamma, c|Y)$ can be visually inspected using a 3-dimensional plot before starting the sampler, a procedure that can help detect identification problems in the distribution of interest. Figure 1 in the paper shows the marginal posterior $\tilde{p}(\gamma, c|Y)$ in the baseline application of the paper. The distribution is convex with a unique maximum.

B   Nonlinear impulse responses

In this section we explain how we compute nonlinear impulse responses in the spirit of Koop et al. (1996), also known as generalized impulse responses. We then discuss how we construct the counterfactual exercises. To simplify the illustration, we concentrate on the

---

[1] We start the chain for (c, ψ) at the marginal mean from the preliminary importance sampler. For γ, we found it useful to follow the suggestion by Livingston and Nur (2017) and initialize the algorithm at relatively high values of γ, in order to stay away from the unidentified region. We start the chain for γ at twice the marginal mean from the preliminary importance sampler.

case of a single shock of interest and use the model

$$\boldsymbol{y}_t = \qquad g_{t-1} \cdot \left(\Pi_a \boldsymbol{x}_{t-1} + \boldsymbol{b}_a m_t + D_a \boldsymbol{q}_t\right) + \qquad \text{(B.6a)}$$

$$+ (1 - g_{t-1}) \cdot \left(\Pi_b \boldsymbol{x}_{t-1} + \boldsymbol{b}_b m_t + D_b \boldsymbol{q}_t\right) + \sqrt{h_{t-1}} \cdot \tilde{\boldsymbol{u}}_t,$$

$$\tilde{\boldsymbol{u}}_t \sim N(\boldsymbol{0}, \Sigma), \qquad \text{(B.6b)}$$

$$g_{t-1} = g(z_{t-1}, \gamma, c) \quad = \frac{1}{1 + e^{-\gamma(z_{t-1} - c)}}, \quad \gamma > 0, \qquad \text{(B.6c)}$$

$$h_{t-1} = h(z_{t-1}, \gamma, c, \psi) = g_{t-1} + e^{\psi} \cdot (1 - g_{t-1}), \quad \psi \geq 0. \qquad \text{(B.6d)}$$

The model features a single shock of interest (m_t is a scalar) and a set of shocks that are controlled for ($\boldsymbol{q}_t$ is a vector, including a constant term). In the special case of a homoskedastic model, $\psi = 0$. To simplify the discussion below, rewrite equation (B.6a) as

$$\boldsymbol{y}_t = g_{t-1} \cdot \left(\Pi_a \boldsymbol{x}_{t-1} + D_a \boldsymbol{q}_t\right) + (1 - g_{t-1}) \cdot \left(\Pi_b \boldsymbol{x}_{t-1} + D_b \boldsymbol{q}_t\right) + \qquad \text{(B.7)}$$

$$+ \boldsymbol{b}_{imp,t} \cdot m_t + \sqrt{h_{t-1}} \cdot \tilde{\boldsymbol{u}}_t,$$

$$\boldsymbol{b}_{imp,t} = g_{t-1} \cdot \boldsymbol{b}_a + (1 - g_{t-1}) \cdot \boldsymbol{b}_b, \qquad \text{(B.8)}$$

where $\boldsymbol{b}_{imp,t}$ is the impulse vector, which depends on t via g_{t-1}.

Our application requires computing nonlinear impulse responses for recessions and for expansions as well as assessing their statistical difference. Call H the number of horizons for which the impulse responses are computed. We proceed as follows:

1) draw

$$\Psi^{(d)} = \left(\Pi_a^{(d)}, \Pi_b^{(d)}, \boldsymbol{b}_a^{(d)}, \boldsymbol{b}_b^{(d)}, D_a^{(d)}, D_b^{(d)}, \Sigma^{(d)}, \gamma^{(d)}, c^{(d)}, \psi^{(d)}\right), \qquad \text{(B.9)}$$

from the joint posterior distribution;

2) draw

$$U^{(d)} = \left\{ \tilde{\boldsymbol{u}}_\tau, m_\tau, \boldsymbol{q}_\tau \right\}_{\tau=0}^{H}, \tag{B.10}$$

from $\tilde{\boldsymbol{u}}_\tau \sim N(\boldsymbol{0}, \Sigma^{(d)})$, $m_\tau \sim N(\mu_m, \sigma_m^2)$ and $q_{j,\tau} \sim N(\mu_{j,q}, \sigma_{j,q}^2)$, with (μ_m, σ_m) the mean and standard deviation of the instrument m_t in the estimation sample, and $(\mu_{j,q}, \sigma_{j,q})$ the mean and standard deviation of the instrument $q_{j,t}$, $\forall\, j$ in the estimation sample;

3) randomly select two periods $(t^{\text{rec}}, t^{\text{exp}})$ within the subsets of periods associated with a recession or an expansion. For both t^{p}, $\text{p} = \{rec,\ exp\}$,

3a) set $\boldsymbol{x}_0$ equal to $\boldsymbol{x}_t$ associated with $t = t^{\text{p}}$;

3b) set $\tilde{\boldsymbol{z}}'_{0,ns}$ equal to $\tilde{\boldsymbol{z}}'_{t,ns}$ associated with $t = t^{\text{p}}$ and compute the corresponding value of z_0;

3c) compute $\boldsymbol{b}^{(d)}_{imp,t^{\text{p}}}$, the impulse vector associated with $\Psi^{(d)}$ and the selected period p;

3d) generate the data $\{\boldsymbol{y}_\tau^{\text{without},(\text{d})}\}_{\tau=0}^{H}$ recursively using the parameter values $\left(\Pi_a^{(d)}, \Pi_b^{(d)}, \boldsymbol{b}^{(d)}_{imp,t^{\text{p}}}, D_a^{(d)}, D_b^{(d)}\right)$ by feeding $U^{(d)}$ into the model, updating $(\tilde{\boldsymbol{z}}'_{\tau,ns}, z_\tau, g_\tau, h_\tau, \boldsymbol{b}^{(d)}_{imp,t^{\text{p}}+\tau})$ as τ progresses towards H;

3e) repeat Steps 3a-3d with the only difference that m_0, the first entry of $\{m_\tau\}_{\tau=0}^{H}$, is now augmented by the shock $\bar{\epsilon}$, which in our application is either expansionary (if $\tau^{\text{p}} = \tau^{\text{rec}}$) or contractionary (if $\tau^{\text{p}} = \tau^{\text{exp}}$). Generate the data $\{\boldsymbol{y}_\tau^{\text{with},(\text{d})}\}_{\tau=0}^{H}$;

3f) compute the impulse response as

$$\text{IRF}_\tau^{\text{p}}(t^{\text{p}}, U^{(d)}, \Pi^{(d)}) = \boldsymbol{y}_\tau^{\text{with},(\text{d})} - \boldsymbol{y}_\tau^{\text{without},(\text{d})}, \tag{B.11}$$

9

which can be interpreted as the additional effect that the shock $\bar{\epsilon}$ has on the dynamics of the variables in the model;

4) compute the sum

$$\mathrm{IRF}_\tau^+(t^{\mathrm{rec}}, t^{\mathrm{exp}}, \mathrm{U}^{(d)}, \Pi^{(d)}) = \mathrm{IRF}_\tau^{\mathrm{rec}}(t^{\mathrm{rec}}, \mathrm{U}^{(d)}, \Pi^{(d)}) + \mathrm{IRF}_\tau^{\mathrm{exp}}(t^{\mathrm{exp}}, \mathrm{U}^{(d)}, \Pi^{(d)});$$

$$(\mathrm{B.12})$$

5) repeat Steps 2-4 a total of N times and compute

$$\mathrm{IRF}_\tau^{\mathrm{rec}}(\Pi^{(d)}), \tag{B.13}$$

$$\mathrm{IRF}_\tau^{\mathrm{exp}}(\Pi^{(d)}), \tag{B.14}$$

$$\mathrm{IRF}_\tau^+(\Pi^{(d)}), \tag{B.15}$$

as the average of

$$\mathrm{IRF}_\tau^{\mathrm{rec}}(t^{\mathrm{rec}}, \mathrm{U}^{(d)}, \Pi^{(d)}), \tag{B.16}$$

$$\mathrm{IRF}_\tau^{\mathrm{exp}}(t^{\mathrm{exp}}, \mathrm{U}^{(d)}, \Pi^{(d)}), \tag{B.17}$$

$$\mathrm{IRF}_\tau^+(t^{\mathrm{rec}}, t^{\mathrm{exp}}, \mathrm{U}^{(d)}, \Pi^{(d)}), \tag{B.18}$$

across the N replications. (B.16)-(B.18) generalize the impulse responses by averaging out $\mathrm{U}^{(d)}$ as well as the exact periods $(t^{\mathrm{rec}}, t^{\mathrm{exp}})$ capturing recessions and expansions;

6) repeat the full algorithm from Step 1 to generate a distribution along posterior draws.

By construction, for each posterior draw $\Psi^{(d)}$, the impact effect of the shock equals $\boldsymbol{b}_{imp,t^{\mathrm{p}}}^{(d)}$. We set the size of the shock $\bar{\epsilon}$ equal to $+0.25$ or -0.25 divided by $b_{imp,t^{\mathrm{p}},1}^{(d)}$, with $b_{imp,t^{\mathrm{p}},j}^{(d)}$ the j

entry of $\boldsymbol{b}_{imp,t^{\mathrm{p}}}^{(d)}$. This generates impulse responses associated with a 25 basis point increase or decrease in the interest rate, which is ordered first in our applications. The application in the paper sets $N = 20$. To speed up computation, we do not compute the nonlinear impulse responses for all posterior draws but for a random subsample of 200 draws. We verified that the results are not affected by this choice.

C   Simulation exercise

To validate the performance of the posterior sampler we conduct a simulation exercise using artificially generated data. We employ a data generating process which is based on the real data we use in our application. We then assess whether the algorithm performs well at recovering the true conditionally linear impulse responses.

The data generating process

We simulate 50 datasets from the following model:

$$\boldsymbol{y}_t^{sim} = \quad g(z_{t-1,ns}^{sim}, \gamma_{ns}^{true}, c_{ns}^{true})\big(\Pi_a^{true} \boldsymbol{x}_{t-1}^{sim} + \boldsymbol{b}_a^{true} \epsilon_t^{sim} + \boldsymbol{d}_a^{true}\big) \tag{C.19a}$$

$$+ \big(1 - g(z_{t-1,ns}^{sim}, \gamma_{ns}^{true}, c_{ns}^{true})\big)\big(\Pi_b^{true} \boldsymbol{x}_{t-1}^{sim} + \boldsymbol{b}_b^{true} \epsilon_t^{sim} + \boldsymbol{d}_b^{true}\big) + \boldsymbol{u}_t^{sim},$$

$$\boldsymbol{u}_t^{sim} \sim N(\boldsymbol{0}, \Sigma_t^{sim}), \tag{C.19b}$$

$$\epsilon_t^{sim} \sim N(0, 1), \tag{C.19c}$$

$$\Sigma_t^{sim} = g(z_{t-1,ns}^{sim}, c_{ns}^{true}, \gamma_{ns}^{true})\Sigma_a^{true} + (1 - g(z_{t-1,ns}^{sim}, c_{ns}^{true}, \gamma_{ns}^{true}))\Sigma_b^{true}, \tag{C.19d}$$

$$g(z_{t-1,ns}^{sim}, \gamma_{ns}^{true}, c_{ns}^{true}) = \frac{1}{1 + e^{-\gamma_{ns}^{true}(z_{t-1,ns}^{sim} - c_{ns}^{true})}}, \qquad \gamma_{ns}^{true} > 0. \tag{C.19e}$$

We now explain the exact specification of the above data generating model, highlighting the few differences from the model estimated in the paper and illustrating how we set the true values of the parameters.

The first challenge that arises in simulating data from the STVAR model with an endogenous transition variable is that the transition variable is standardized using observed moments of the data. The researcher first specifies the nonstandardized transition variable $z_{t-1,ns}$, which only depends on data until $t - 1$. He then calculates (μ_z, σ_z) as the sample mean and standard deviation of $z_{t-1,ns}$ on the full estimation sample $\{z_{t-1,ns}\}_{t=1}^T$, computes the standardized transition variable

$$z_{t-1} = \frac{z_{t-1,ns} - \mu_z}{\sigma_z}, \tag{C.20}$$

and estimates the model using the transition function

$$g(z_{t-1}, \gamma, c) = \frac{1}{1 + e^{-\gamma(z_{t-1}-c)}}, \quad \gamma > 0. \tag{C.21}$$

This standardization has the advantage that it makes (γ, c) scale free. However, when simulating data with an endogenous transition variable, this specification is not viable, because the scaling parameters (μ_z, σ_z) are not known until the full simulated dataset is generated, creating a circularity problem. To circumvent this limitation, we equivalently rewrite the transition function as

$$g(z_{t-1}, \gamma, c) = \frac{1}{1 + e^{-\gamma(z_{t-1}-c)}}, \qquad \gamma > 0, \tag{C.22}$$

$$= \frac{1}{1 + e^{-\frac{\gamma}{\sigma_z}(\sigma_z z_{t-1} - \sigma_z c)}}, \tag{C.23}$$

$$= \frac{1}{1 + e^{-\gamma_{ns}(z_{t-1,ns}-c_{ns})}}, \quad \gamma_{ns} > 0, \tag{C.24}$$

where

$$\gamma_{ns} = \gamma/\sigma_z, \qquad \gamma = \sigma_z \gamma_{ns}, \tag{C.25}$$

$$c_{ns} = c\sigma_z + \mu_z, \qquad c = (c_{ns} - \mu_z)/\sigma_z, \tag{C.26}$$

$$z_{t-1} = \frac{z_{t-1,ns} - \mu_z}{\sigma_z}, \qquad z_{t-1,ns} = \sigma_z z_{t-1} + \mu_z. \tag{C.27}$$

The data generating process sets parameter values for the nonstandardized parametrization (γ_{ns}, c_{ns}) and generates the data recursively using equation (C.24) rather than (C.22), updating the nonstandardized transition variable $z_{t-1,ns}^{sim}$ at each iteration. Once the data is generated, one can specify the transition function in terms of the standardized (rather than nonstandardized) transition variable and estimate the standardized equivalent of (γ_{ns}, c_{ns}).

A second challenge consists of choosing plausible parameter values for the data gener-

ating process. We estimate a VAR(1) on real data, as in Bruns and Lütkepohl (2022). We set the true parameter values as close as possible to posterior draws from this real data application, which features a homoskedastic model. We choose $(\gamma^{true}, c^{true})$ among the posterior draws $\{\gamma_i^{post}\}_{i=1}^{n_2}$ and $\{c_i^{post}\}_{i=1}^{n_2}$ such that $\{\gamma^{true}, c^{true}\}$ imply that the companion form of the autoregressive coefficients Π_a and Π_b, on average across regimes, has a minimal eigenvalue, thereby ensuring a stationary DGP. We obtain $\gamma^{true} = 16.10$ and $c^{true} = -1.02$. Note that these are standardized values, because the application uses a standardized transition variable. We then use the moments (μ_z, σ_z) from the underlying nonstandardized transition variable to compute the corresponding nonstandardized values $\gamma_{ns}^{true} = 1458.80$ and $c_{ns}^{true} = -0.0081$. It is these values that are used to simulate the datasets. We then set $\Theta^{true} = [\Pi_a^{true}, \boldsymbol{b}_a^{true}, \boldsymbol{d}_a^{true}, \Pi_b^{true}, \boldsymbol{b}_b^{true}, \boldsymbol{d}_b^{true}]$ equal to the elements of the posterior draw of $\Theta = [\Pi_a, B_a, D_a, \Pi_b, B_b, D_b]$ corresponding to the posterior draw for $(\gamma^{true}, c^{true})$. For $(\boldsymbol{b}_a^{true}, \boldsymbol{b}_b^{true})$ we normalize the (1,1) element to 0.25 and obtain

$$
\boldsymbol{b}_a^{true} = \begin{bmatrix} 0.2500 \\ 0.0666 \\ -0.0551 \\ -0.0001 \\ -0.0024 \end{bmatrix} \qquad \boldsymbol{b}_b^{true} = \begin{bmatrix} 0.2500 \\ 2.8106 \\ -0.5681 \\ 0.0107 \\ -0.0019 \end{bmatrix}. \tag{C.28}
$$

All in all, this implies that our simulation exercise builds on the same number of variables as the application of the paper, namely 5 endogenous variables, and mimics their dynamics. We discard the columns of Θ relating to additional control variables in $\boldsymbol{q}_t$ and set $\boldsymbol{q}^{sim} = 1$.

A third challenge is to assess the model's robustness to time-variation in the variance of the error term. As pointed out in the paper, the baseline specification of our application assumes a homoskedastic error covariance matrix Σ and controls for instruments for other shocks in $\boldsymbol{q}_t$. In this simulation exercise, we aim to assess whether the model succeeds in

recovering important posterior moments even if the underlying data generating process is heteroskedastic, and even if no additional controls are introduced in the model. We do this to further assess the strength of our methodology in applied work. We generate data by drawing innovations $\boldsymbol{u}_t$ featuring heteroskedasticity as in equation (C.19d), which is more general than the type of heteroskedasticity allowed for in the model of the paper. We then assess if our model can recover the true impulse responses even under the simplified scenario of homoskedasticity without additional controls in $\boldsymbol{q}_t$ apart from the constant term. We set Σ_a^{true} and Σ_b^{true} using the following steps:

1. compute the reduced-form residuals as $U = Y - \Theta^{true} W(\gamma^{true}, c^{true})$, where (Y, W) are the same matrices as from the real-data application, except that we omit control variables in $\boldsymbol{q}_t$;

2. divide U into two equally sized subsamples using

$$U_a = U, \quad \text{if} \quad g(z_{t-1}, \gamma^{true}, c^{true}) \leq q_{z,0.5},$$

$$U_b = U, \quad \text{if} \quad g(z_{t-1}, \gamma^{true}, c^{true}) > q_{z,0.5},$$

where $q_{z,0.5}$ is the median of the transition function conditioning on $(\gamma^{true}, c^{true})$;

3. compute

$$\hat{\Sigma}_a = \frac{U_a U_a'}{T/2}, \tag{C.29}$$

$$\hat{\Sigma}_b = \frac{U_b U_b'}{T/2}; \tag{C.30}$$

4. Set $\Sigma_a^{true} = \hat{\Sigma}_a$ and $\Sigma_a^{true} = \hat{\Sigma}_a$.

We obtain

$$\Sigma_a^{true} = \begin{bmatrix} 35.7820 & -11.9630 & -0.8206 & 0.0213 & 0.1785 \\ -11.9630 & 38.2796 & -3.6655 & -0.0281 & -0.1401 \\ -0.8206 & -3.6655 & 3.2895 & 0.0030 & -0.0007 \\ 0.0213 & -0.0281 & 0.0030 & 0.0021 & -0.0004 \\ 0.1785 & -0.1401 & -0.0007 & -0.0004 & 0.0164 \end{bmatrix} \times 10^{-3}, \qquad (\text{C.31})$$

$$\Sigma_b^{true} = \begin{bmatrix} 32.0004 & -8.2123 & -0.1955 & 0.0187 & 0.0478 \\ -8.2123 & 26.0259 & -3.1626 & -0.0110 & -0.0452 \\ -0.1955 & -3.1626 & 1.2612 & -0.0057 & 0.0038 \\ 0.0187 & -0.0110 & -0.0057 & 0.0022 & -0.0006 \\ 0.0478 & -0.0452 & 0.0038 & -0.0006 & 0.0126 \end{bmatrix} \times 10^{-3}. \qquad (\text{C.32})$$

The analysis above completes the illustration of the data generating process from equations (C.19a)-(C.19e). To generate data, we set the initial value $\boldsymbol{x}_0^{sim}$ building on the first observation of the real data application of the paper, and set $z_{0,ns}^{sim}$ equal to the first observation of the nonstandardized transition variable in the application. We then use the model recursively to generate 577 observations. We discard the first 100 observations and split the remaining ones into $T_{training} = 127$ observations as a training sample and $T_{estimation} = 350$ observations for estimation. The training and estimation sample length is chosen as in the real-data application. We repeat the procedure 50 times to generate 50 pseudo datasets.
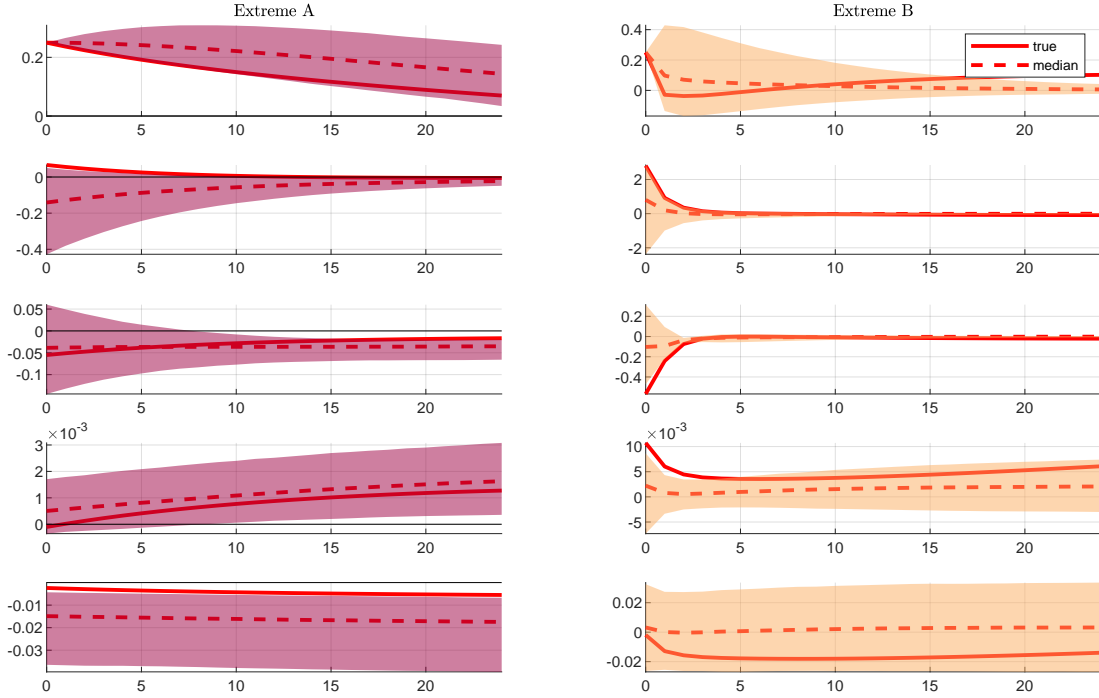
On each dataset, we estimate a homoskedastic model that includes all 5 variables from the data generating process, a constant, 1 lag for the autoregressive part of the model, and a noisy measure of the shock ϵ_t^{sim} constructed as

$$m_t = \phi^{true}\epsilon_t^{sim} + \sigma_\nu^{true}\nu_t, \qquad \nu_t \sim \text{ iid } N(0,1). \qquad (\text{C.33})$$

This measurement error in m_t, which is assumed to be independent of ϵ_t, allows us to assess whether we are still able to recover important posterior moments in a situation where we do not observe the true structural shock but a noisy measure. We set σ_ν^{true} equal to the standard deviation of the proxy in our application and calibrate ϕ^{true} to ensure that the population correlation between m_t and ϵ_t^{sim} is 0.25. We introduce this measurement error, following e.g. Bruns and Lütkepohl (2022), to assess the model's ability to deal with noisy shock measures, and calibrate the correlation to match applications such as Gertler and Karadi (2015) and Jarociński and Karadi (2020). We add no proxies capturing the structural shocks other than the shock of interest. We use the same prior specifications as in the baseline application of the paper.
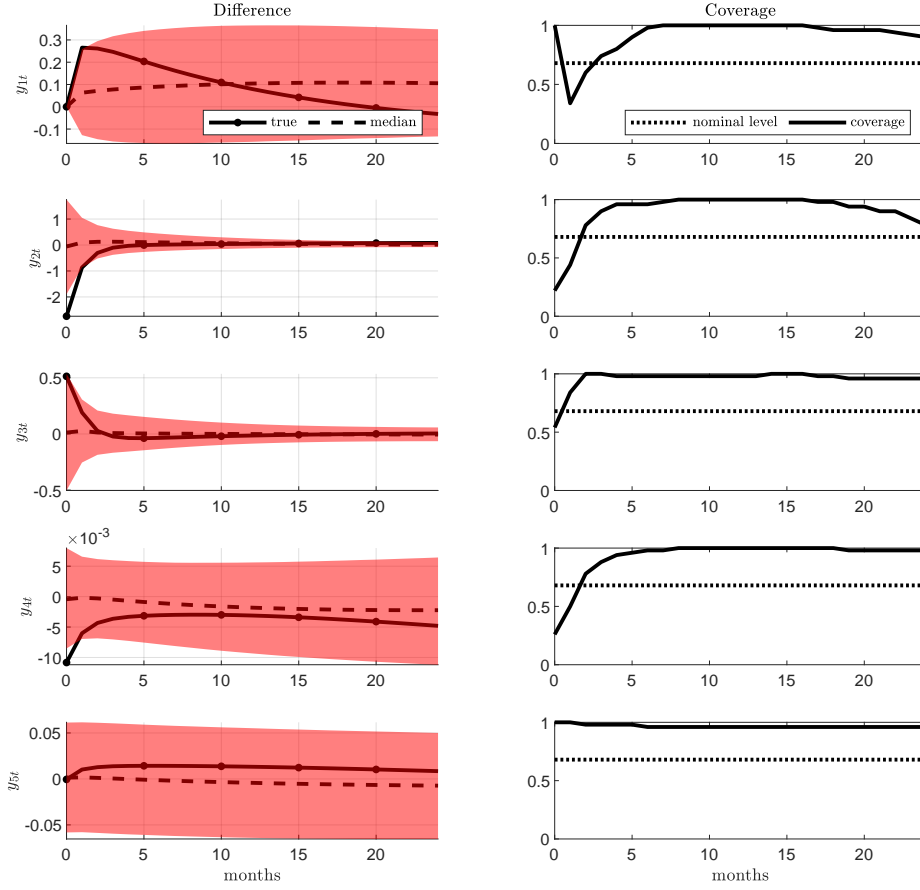
Figure C.1: Simulation Study: Conditionally linear impulse responses

Note: The figure shows the conditionally linear impulse responses conditioning on the Θ_{true} as well as the median and 68% bands of the posterior conditionally linear impulse responses for one of the 50 Monte Carlo replications. The impact effect of the first variable is normalized to match the true impact effect. The proxy has a 0.25 correlation with the structural shock.
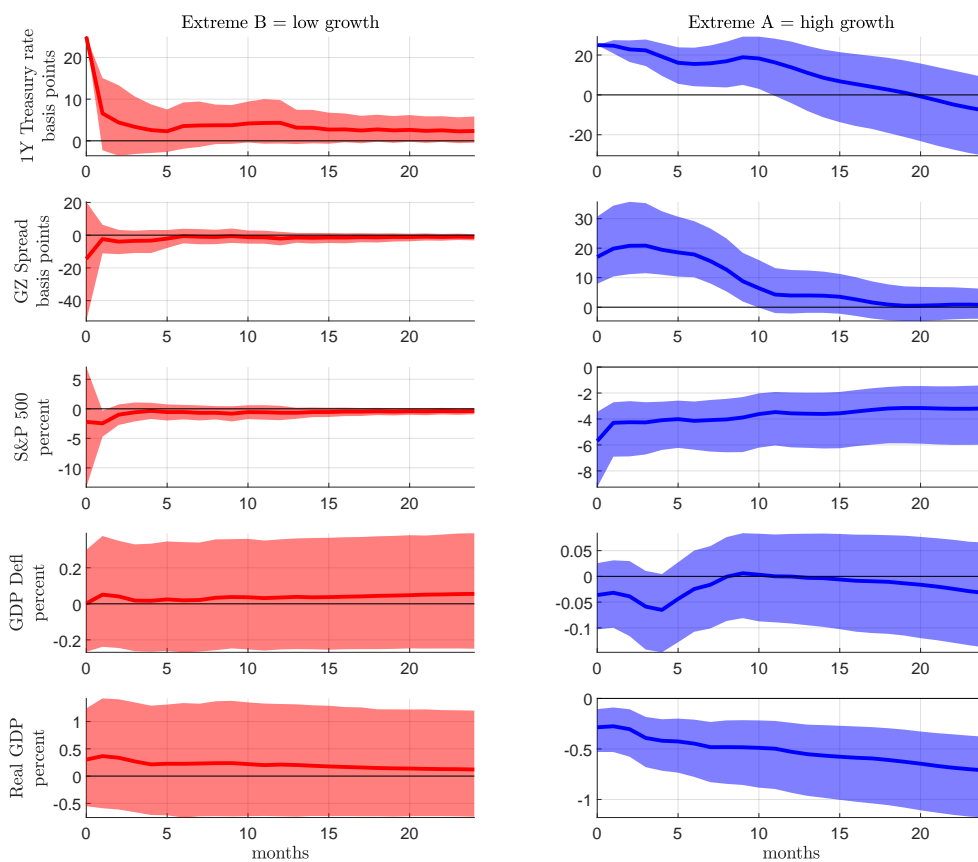
The parameters $[\Pi_a^{true}, \boldsymbol{b}_a^{true}, \Pi_b^{true}, \boldsymbol{b}_b^{true}]$ pin down the true conditionally linear impulse

Figure C.2: Simulation Study: Difference between Regime A and Regime B and Coverage

Note: The left panel shows the difference in the conditionally linear impulse responses between regime A and regime B pooled across 50 Monte Carlo replications (dashed), as well as the true difference (solid line), and the associated pooled 68% bands. The right panel shows the coverage (solid line) and the nominal level (dotted). The impact effect of the first variable is normalized to match the true impact effect. The proxy has a 0.25 correlation with the structural shock.

responses that we aim to correctly estimate using simulated data. Figure C.1 shows the true conditionally linear impulse responses as well as the median and posterior 68% bands of posterior impulse responses for one of the 50 simulations as an illustration. The true impulse response largely lies within the 68% bands for all variables in the model. This confirms that the posterior sampler performs well at recovering the true impulse responses. It also suggests that heteroskedasticity in the innovations of the model is not always necessary for the model to uncover the true conditionally linear impulse responses to the shock of

18

interest even if the data are generated using a process with heteroskedasticity.

A key requirement of the model is the ability to detect differences in the impulse responses. Figure C.2 (left panel) shows that the pointwise median difference between regime A and regime B, pooled across simulations, lies within the pooled 68% confidence bands and is close to the true difference, further confirming the homoskedastic model's ability to recover the parameters of a hetereoskedastic DGP. Lastly, figure C.2 (right panel) reports the coverage rate of this difference, i.e. the share, across Monte Carlo replications, for which the 68% bands contain the true difference. Overall the model performs well at matching the nominal level, with a tendency to be conservative at intermediate horizons.

D   Additional figures and tables

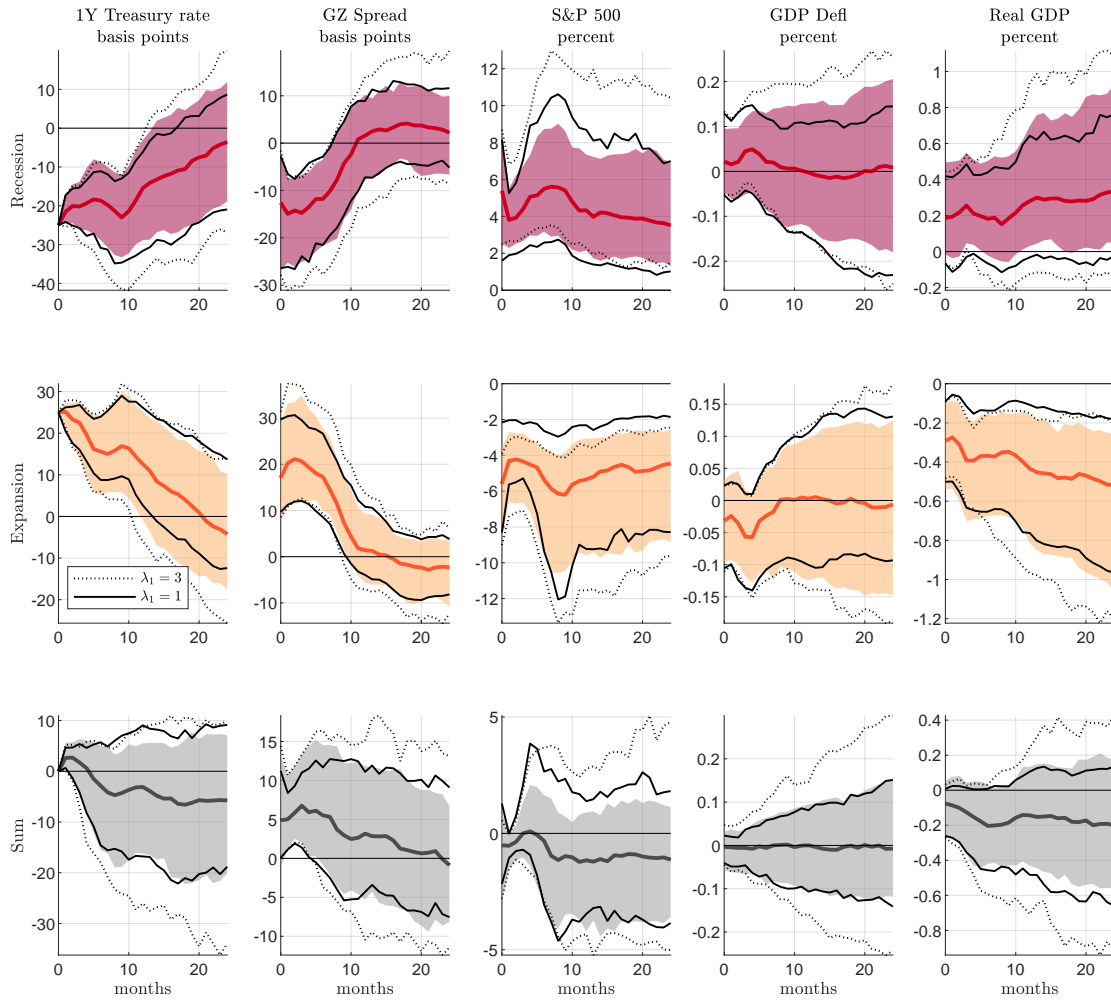**Figure D.3:** Conditionally linear impulse response

Note: The figure shows the pointwise median and 68% bands of the conditionally linear impulse responses.

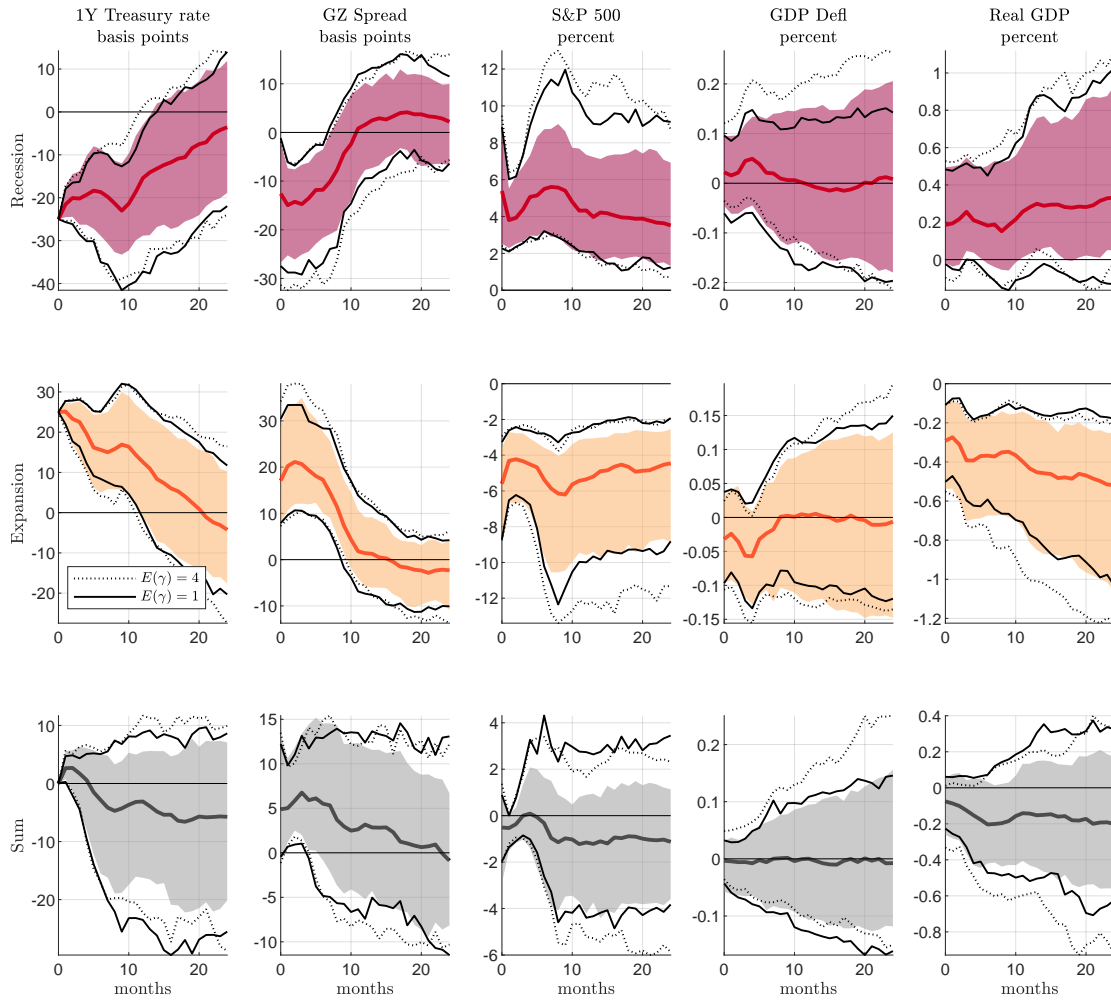**Figure D.4:** Comparison of different types of prior dependence of $\boldsymbol{\theta}$ on γ

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline prior specification, which uses $\gamma_\gamma = 2$, $c_\gamma = 1$, see equation (6) in the paper.

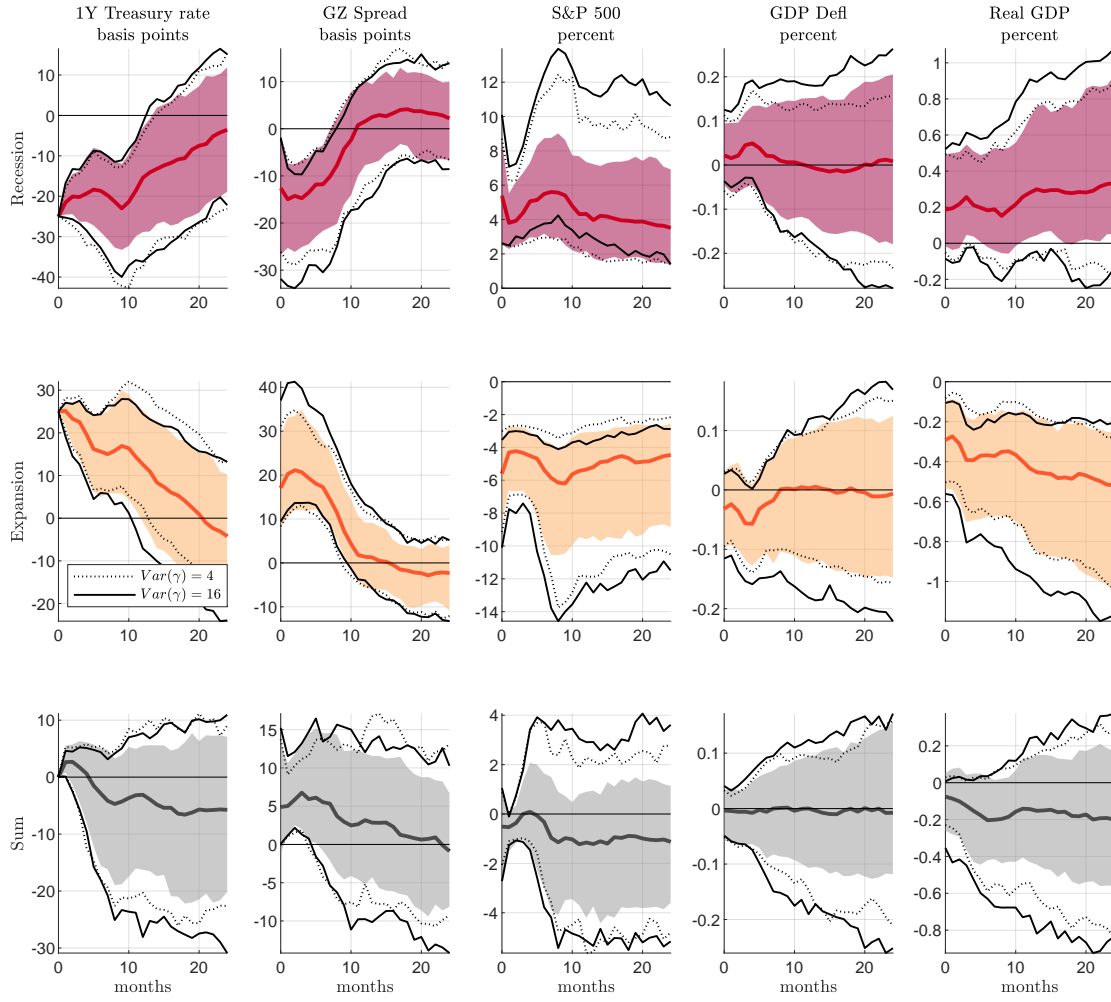**Figure D.5:** Comparison of different prior variances for $\boldsymbol{\theta}$

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline prior specification for $\boldsymbol{\theta}$, which uses $\lambda_1 = 2$.

Figure D.6: Comparison of different prior means for γ

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline prior specification for γ, which uses $E(\gamma) = 2$ before truncation at 0.01.
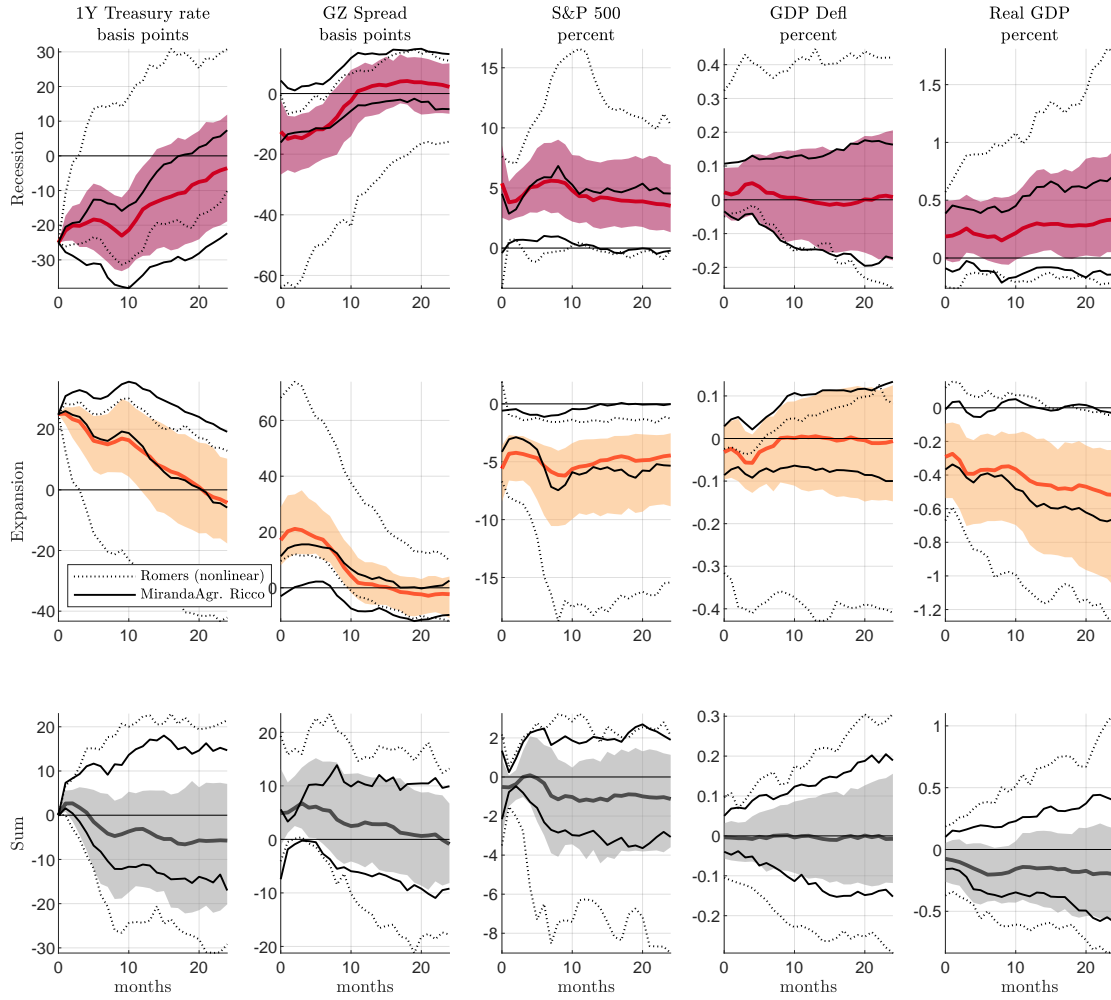
Figure D.7: Comparison of different prior variances for γ

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline prior specification for γ, which uses $Var(\gamma) = 9$ before truncation at 0.01.
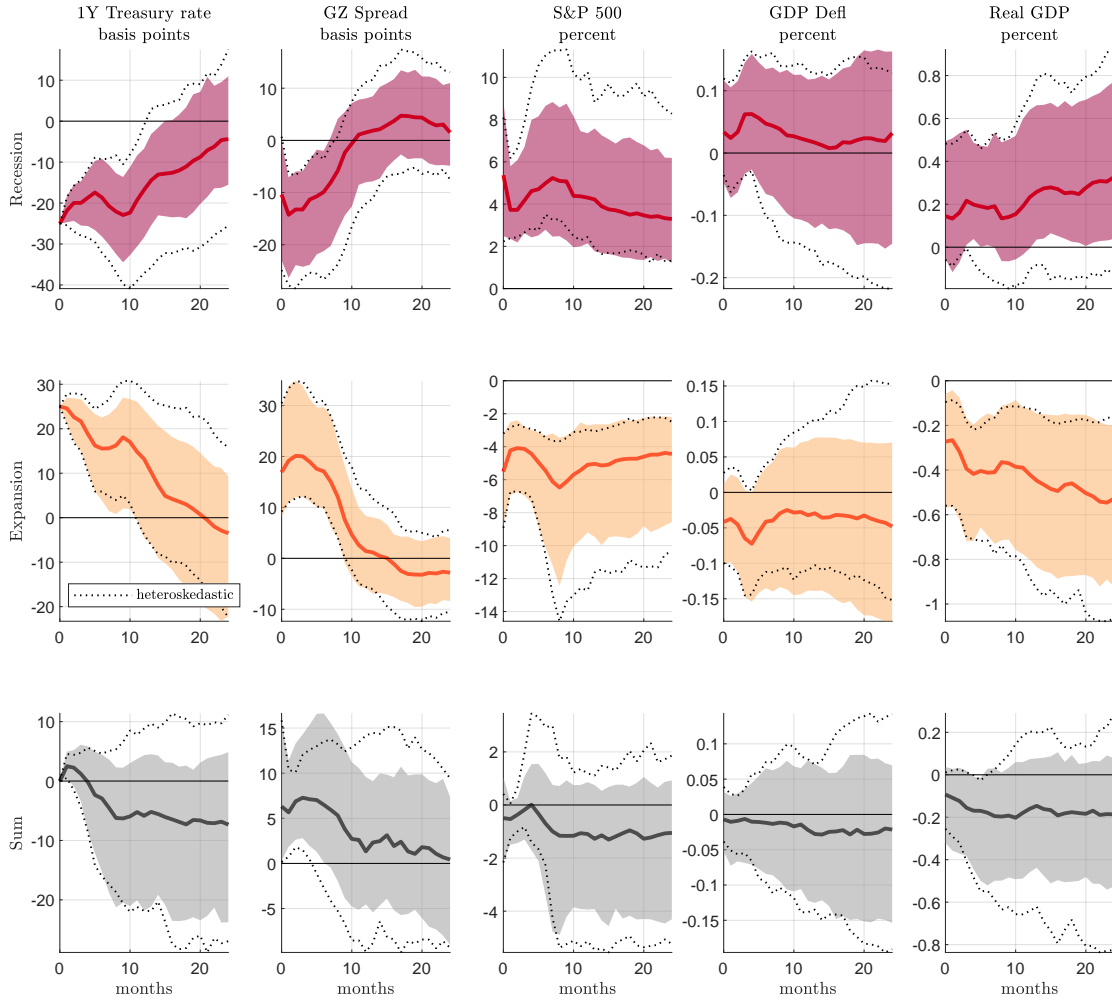
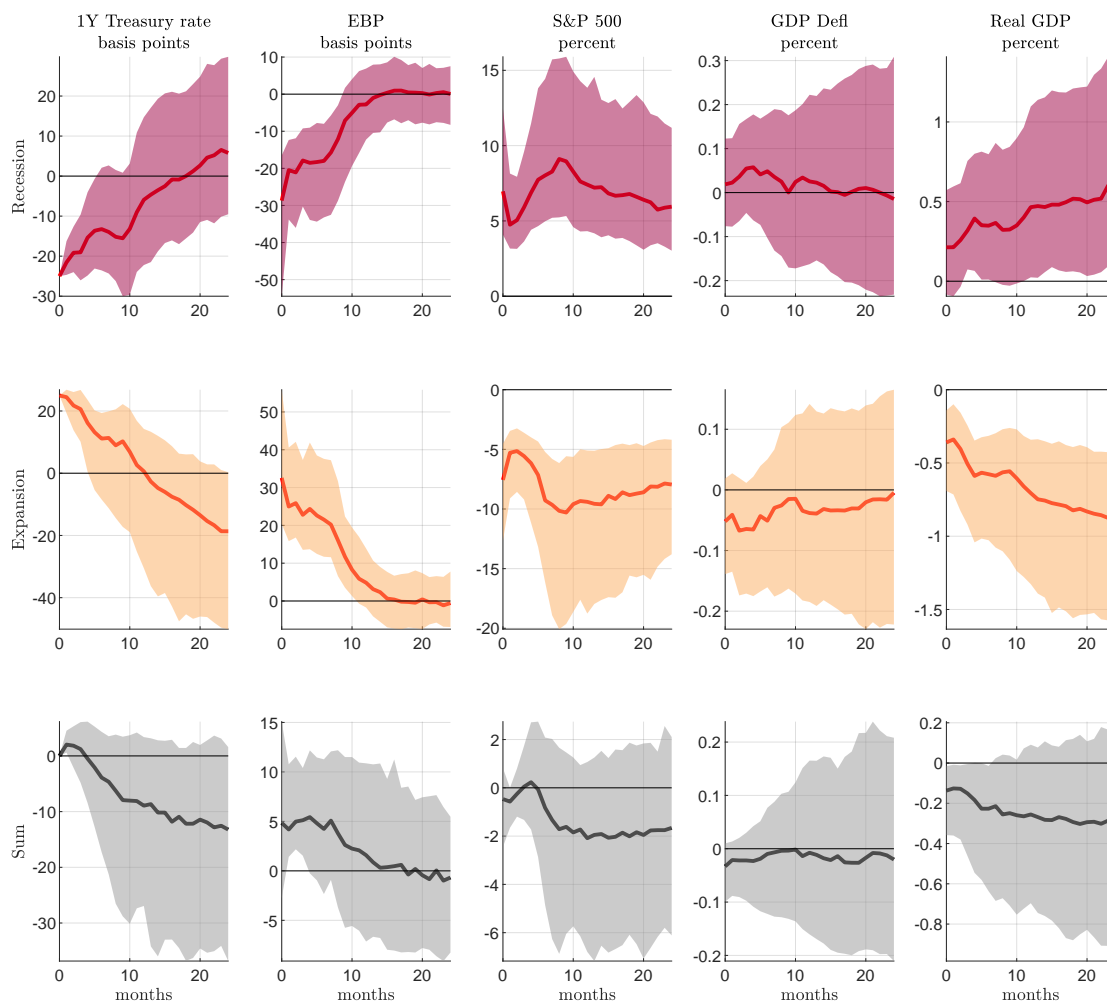**Figure D.8:** Comparison of different instruments for the monetary shocks

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline specification, which uses poor man's version of the instrument by Jarociński and Karadi (2020). The alternative cases use as instrument the exact nonlinear estimates of the Romer and Romer shocks by Tenreyro and Thwaites (2016) or the instrument by Miranda-Agrippino and Ricco (2021).

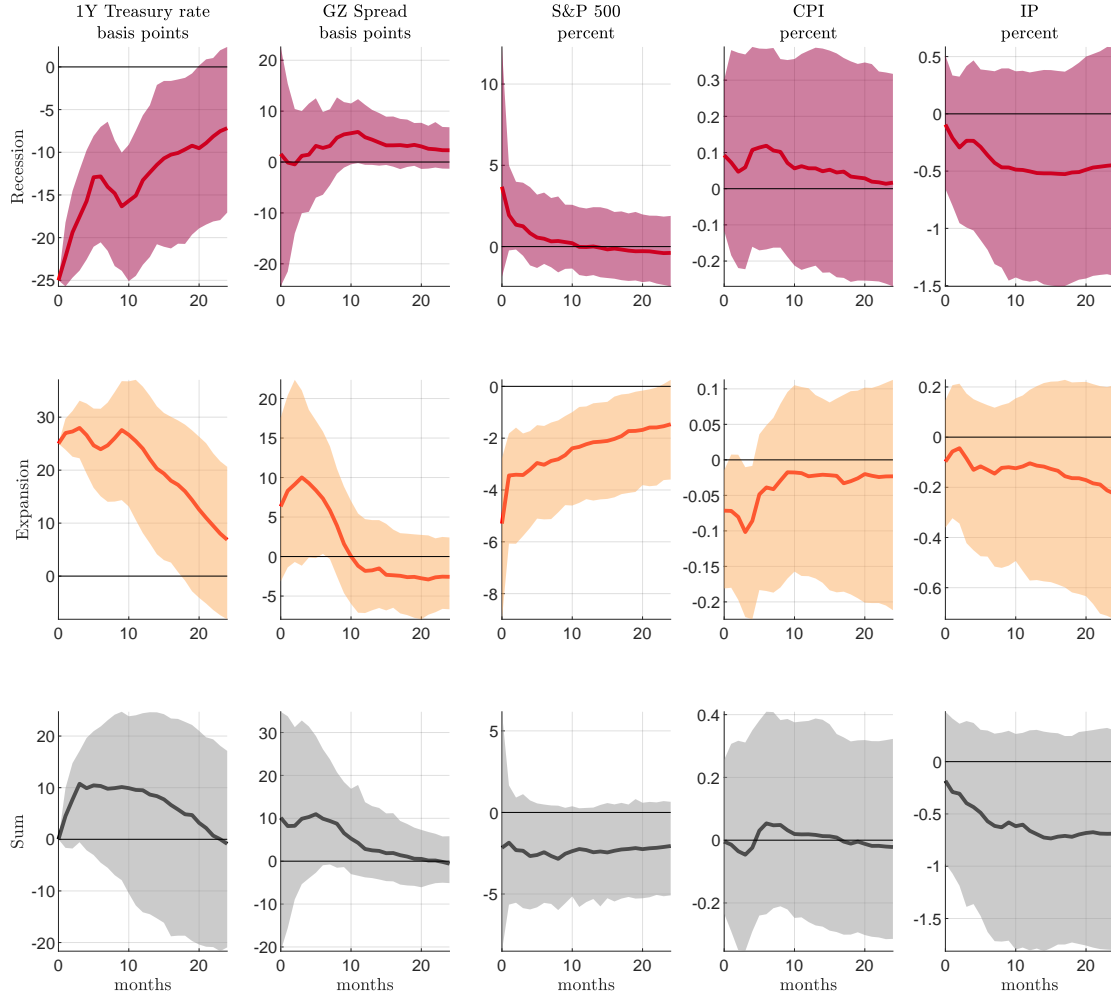**Figure D.9:** Comparison of homoskedastic and heteroskedastic models

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline specification, which assumes homoskedasticity ($\psi = 0$). The alternative case allows for heteroskedasticity.

Figure D.10: Comparison when the GZ Spread is replaced with the GZ Excess Bond Premium
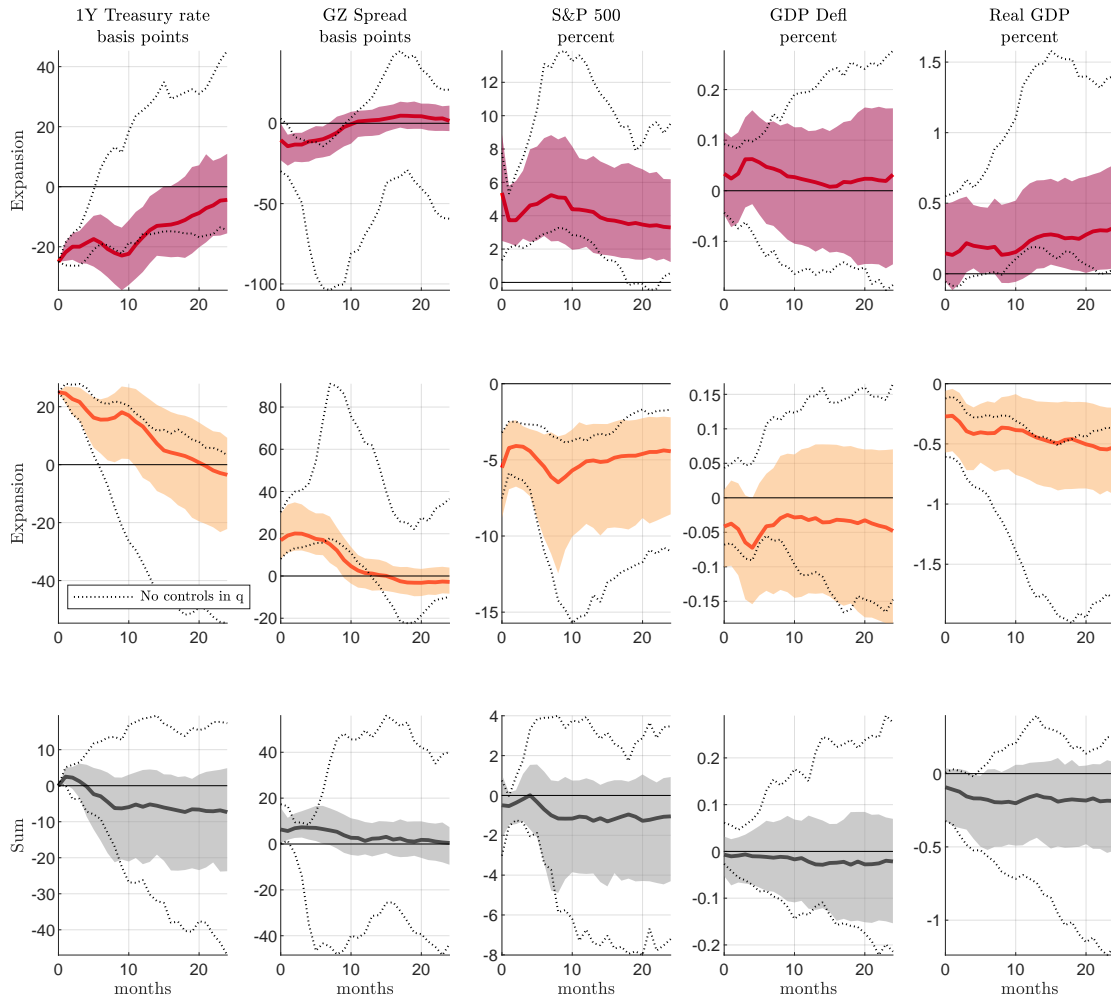

Note: The figure shows pointwise 68% bands only.

Figure D.11: Comparison when replacing real GDP with industrial production and the GDP deflator with CPI inflation
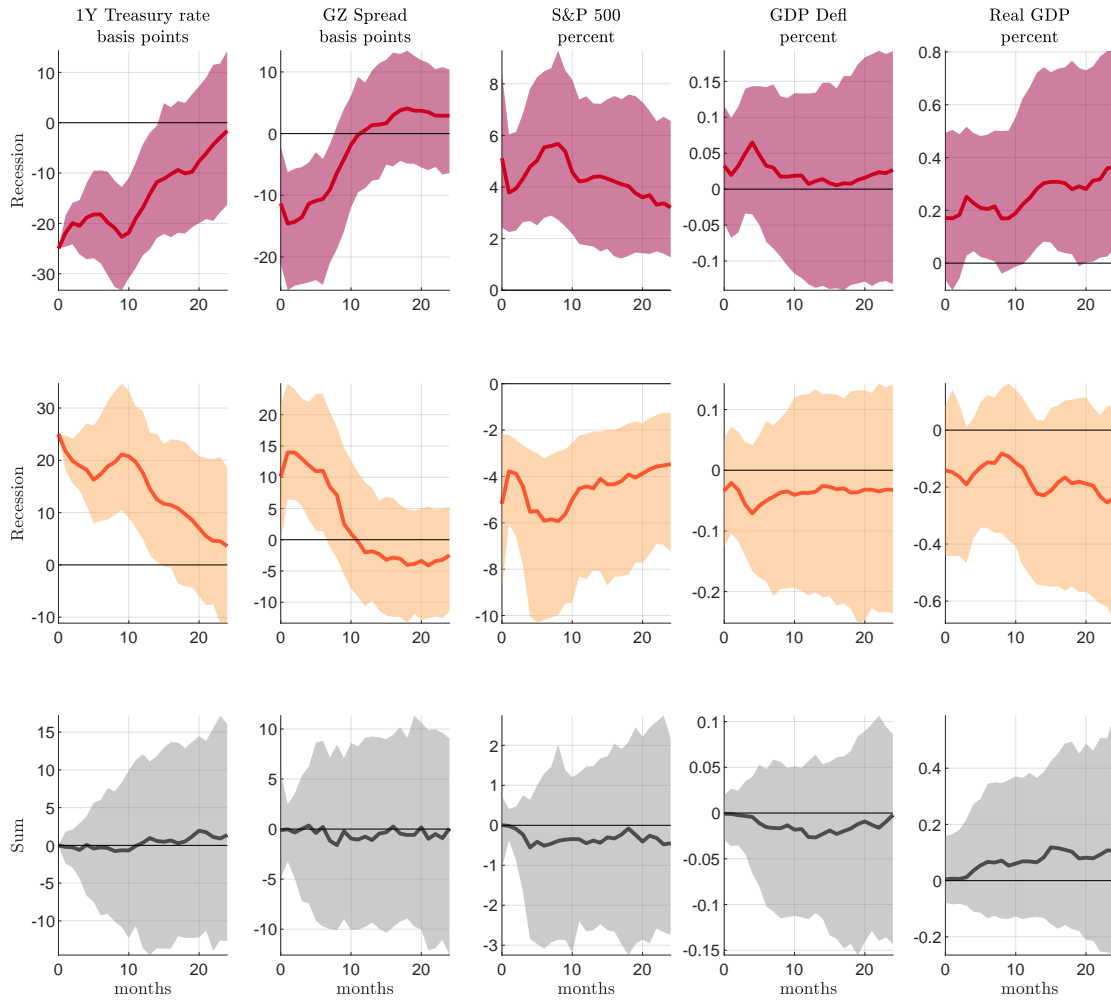

Note: The figure shows pointwise 68% bands only.

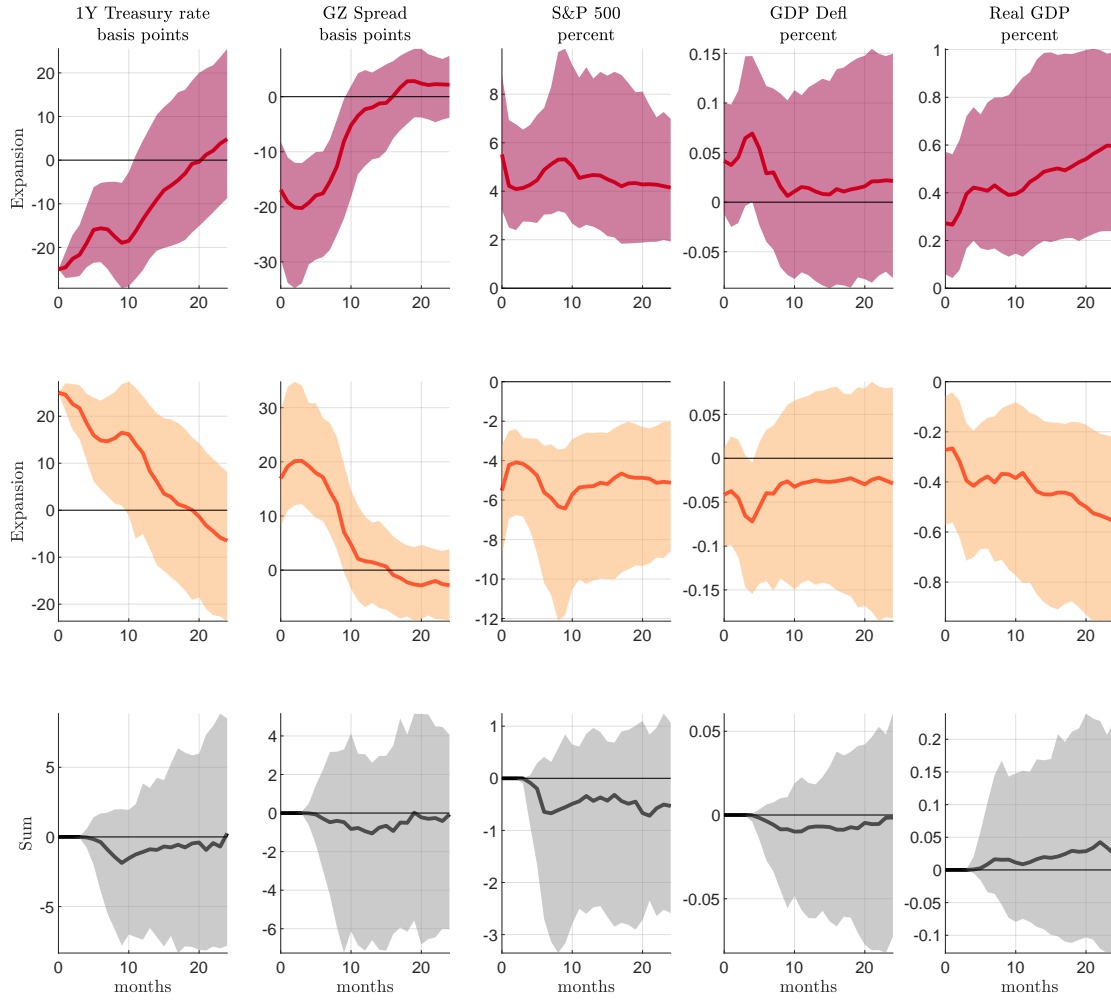**Figure D.12:** Comparison when adding no instruments in $\boldsymbol{q}_t$

Note: The figure shows pointwise 68% bands only. The shaded area corresponds to the baseline specification, which includes four instruments in $\boldsymbol{q}_t$

Figure D.13: Expansionary versus contractionary shock, both given in recession

Note: The figure shows pointwise median and 68% bands. The shock is always given in a low quartile of the transition variable, to document that the nonlinearity documented in the paper is not driven by the sign of the shock.

Figure D.14: Expansionary versus contractionary shock, both given in expansion

Note: The figure shows pointwise median and 68% bands. The shock is always given in a high quartile of the transition variable, to document that the nonlinearity documented in the paper is not driven by the sign of the shock.

Table D.1: Convergence criteria for the posterior draws of γ, c, ψ

Parameter	Raftery and Lewis (1992) n^*	Geweke (1992) p-value
γ	13,645	0.81
c	18,659	0.50

Notes: For the test statistic by Raftery and Lewis (1992), convergence is found if $n^* < 20,000$. For the test statistic by Geweke (1992), convergence is found if p-value > 0.1.

Table D.2: Strength in the nonlinearities of the nonlinear impulse responses, using the Romer and Romer shocks as in Tenreyro and Thwaites (2016)

	horizons				
	$0:2$	$3:5$	$6:8$	$9:11$	$12:18$
One-year interest rate	0.675	0.505	0.415	0.395	0.335
GZ Spread	0.685	0.745	0.64	0.55	0.34
S&P500	0.115	0.225	0.215	0.235	0.135
GDP Deflator	0.355	0.36	0.39	0.415	0.38
Real GDP	0.36	0.35	0.36	0.355	0.3

Notes: For each variable of the model we select a set of the response horizons (e.g. from 0 to 2, from 3 to 5 etc.) and report the share of total posterior draws for which the sum of the impulse responses in recession and expansion is positive in each of the selected horizons. The results from this table correspond to the analysis when identifying the monetary shock using the nonlinear estimates of the Romer and Romer shocks from Tenreyro and Thwaites (2016).
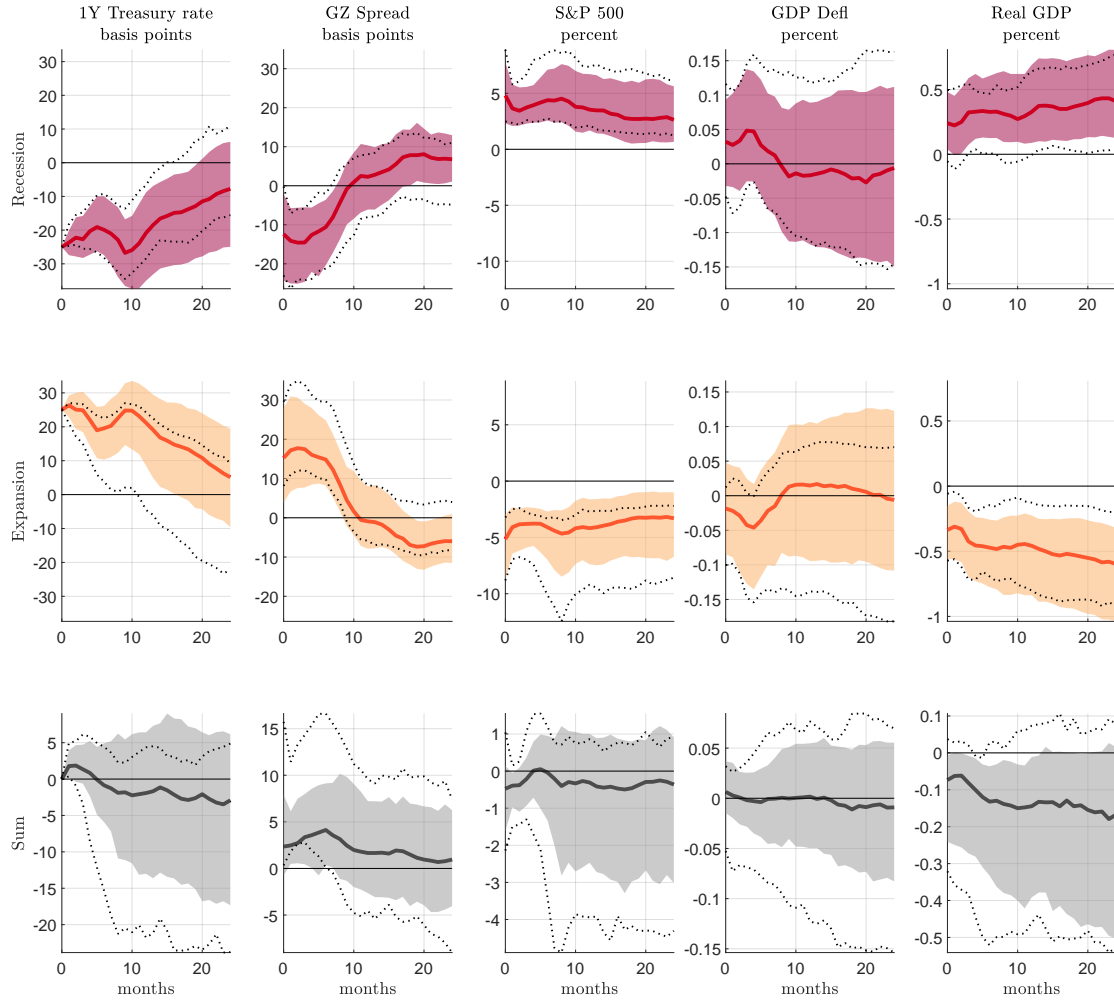
Table D.3: Strength in the nonlinearities of the nonlinear impulse responses, using the instrument by Miranda-Agrippino and Ricco (2021)

	horizons				
	$0:2$	$3:5$	$6:8$	$9:11$	$12:18$
One-year interest rate	0.875	0.605	0.415	0.485	0.42
GZ Spread	0.39	0.665	0.565	0.475	0.345
S&P500	0.18	0.45	0.275	0.345	0.205
GDP Deflator	0.51	0.505	0.475	0.525	0.49
Real GDP	0.345	0.33	0.32	0.335	0.325

Notes: For each variable of the model we select a set of the response horizons (e.g. from 0 to 2, from 3 to 5 etc.) and report the share of total posterior draws for which the sum of the impulse responses in recession and expansion is positive in each of the selected horizons. The results from this table correspond to the analysis when identifying the monetary shock using the instrument by Miranda-Agrippino and Ricco (2021).
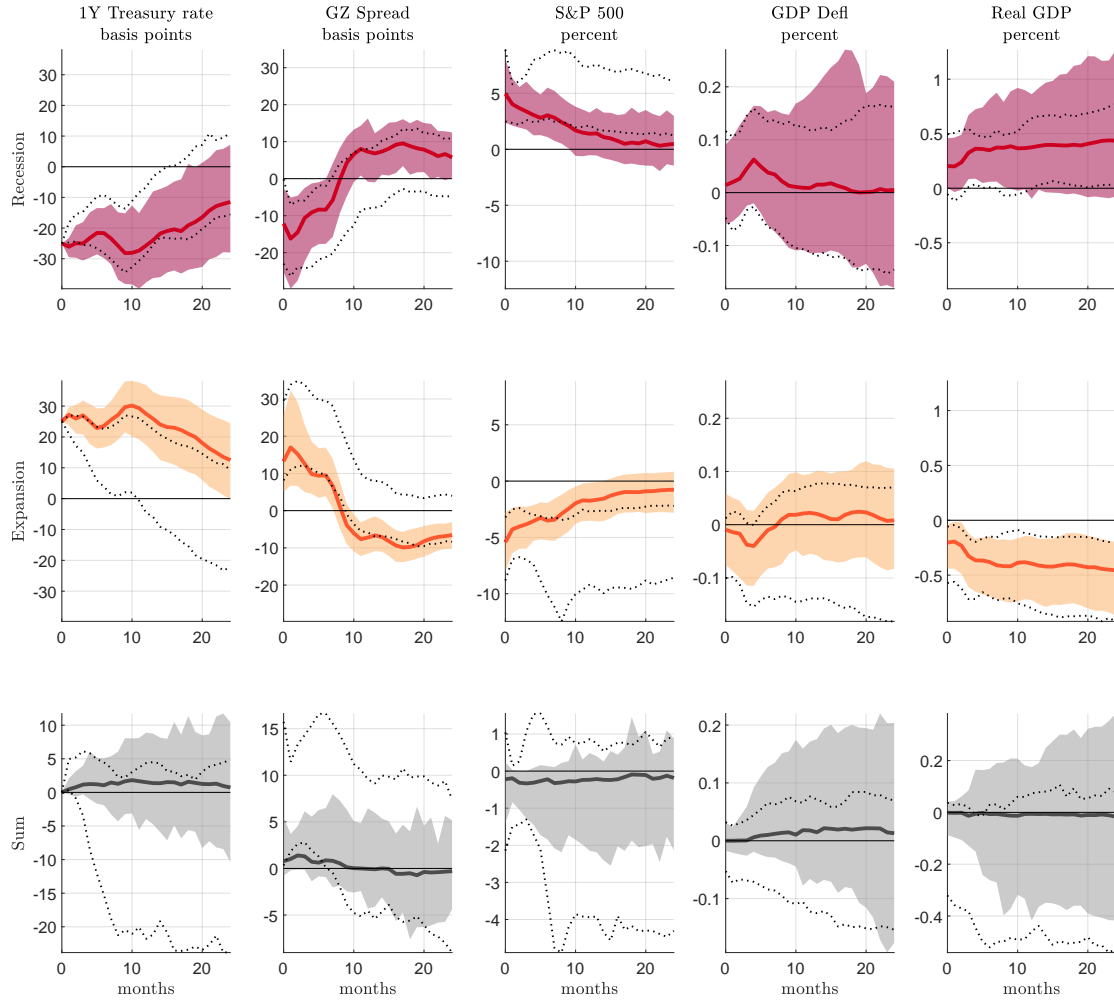
Figure D.15: Nonlinear IRFs associated with the top-left plot in Figure 4 in the paper (calibration, $(\gamma, c) = (3.5, -2.479)$)
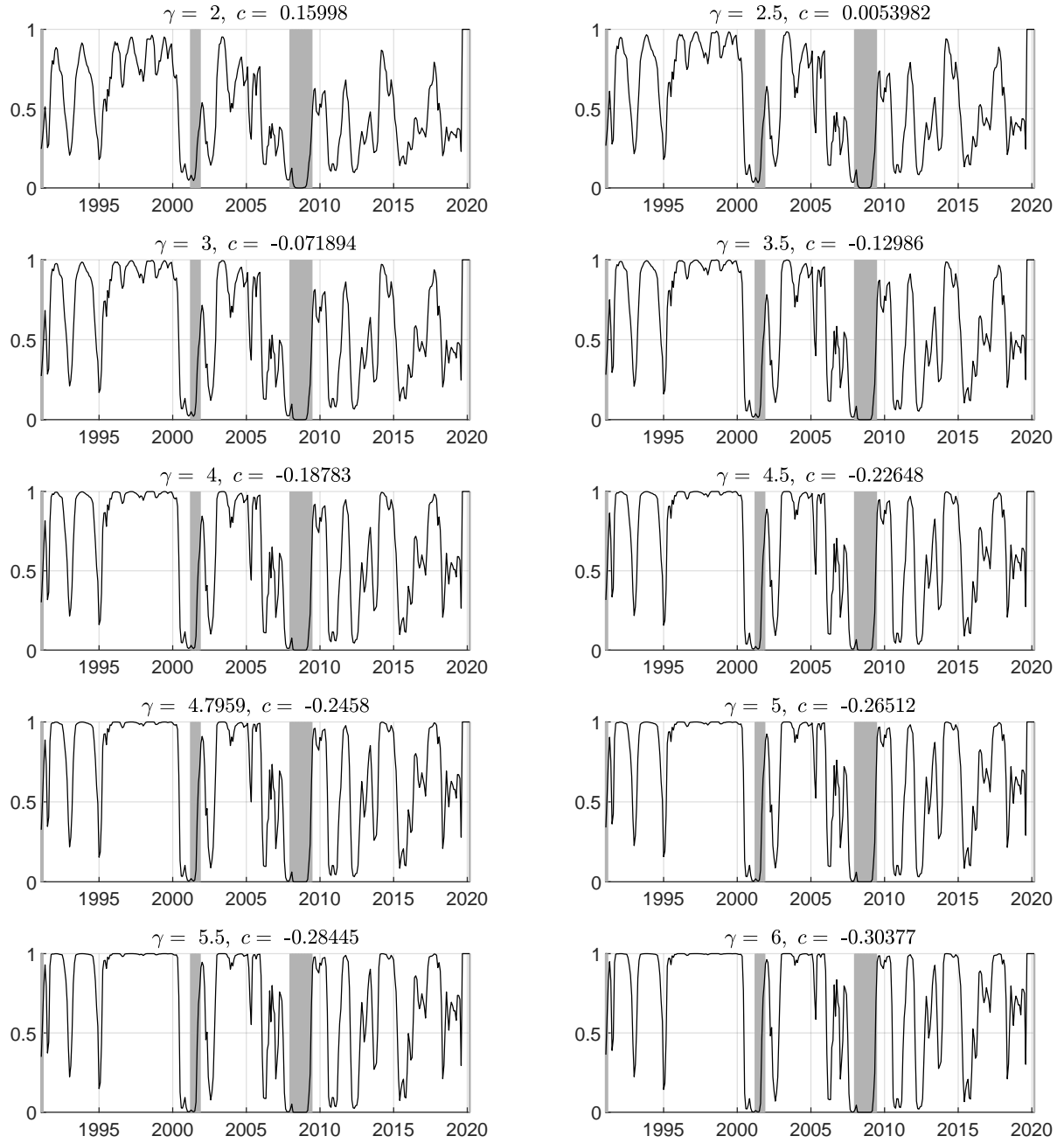

Note: The figure shows pointwise median and 68% bands, using the first definition of the nonlinear impulse responses. The shaded area corresponds to the specification under calibration, while the dashed lines correspond to the results under baseline estimation.

Figure D.16: Nonlinear IRFs associated with the top-right plot in Figure 4 in the paper (calibration, $(\gamma, c) = (6, -4.1297)$)

Note: The figure shows pointwise median and 68% bands, using the first definition of the nonlinear impulse responses. The shaded area corresponds to the specification under calibration, while the dashed lines correspond to the results under baseline estimation.

Figure D.17: Transition function under the calibration from the middle and bottom row of Figure 5 in the paper

Note: The shaded areas indicate NBER recessions.

References

Bruns, M. and H. Lütkepohl (2022). Comparison of local projection estimators for proxy vector autoregressions. *Journal of Economic Dynamics and Control 134*, 104277.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Geweke, J. (1992). Evaluating the accurating of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, 169–193.

Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises: the role of information shocks. *American Economic Journal: Macroeconomics 12*(2), 1–43.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1), 119–147.

Livingston, G. and D. Nur (2017). Bayesian inference for smooth transition autoregressive (STAR) model: A prior sensitivity analysis. *Communications in Statistics-Simulation and Computation 46*(7), 5440–5461.

Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics 13*(3), 74–107.

Raftery, A. and S. Lewis (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics 4*, 763–773.

Tenreyro, S. and G. Thwaites (2016). Pushing on a string: US monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics 8*(4), 43–74.

Waggoner, D. F., H. Wu, and T. Zha (2016). Striated Metropolis–Hastings sampler for high-dimensional models. *Journal of Econometrics 192*(2), 406–420.