Online Appendix to

'Monetary Policy, External Instruments and Heteroskedasticity'

Thore Schlaak*

Malte Rieth†

Maximilian Podstawski‡

December 8, 2021

*Freie Universität Berlin; schlaak@zedat.fu-berlin.de

†Corresponding author: Martin-Luther-Universität Halle-Wittenberg, Faculty of Law, Economics and Business, Department of Economics, Chair of Monetary Economics, 06099 Halle, Germany; and DIW Berlin; mrieth@diw.de

‡DIW Berlin; maximilian.podstawski@gmail.com

A.1 Notes on computation

Our estimation procedure follows the expectation maximization algorithm described in Herwartz and Lütkepohl (2014) and is based on the following concentrated out log-likelihood function in the maximization step:

$$\mathcal{L}(D, \Lambda_m) = \frac{1}{2} \sum_{m=1}^{M} \left[T_m \log(\det(\tilde{\Sigma}_m)) + \mathrm{tr}\left((\tilde{\Sigma}_m)^{-1} \sum_{t=1}^{T} \xi_{mt|T} u_t u_t' \right) \right],$$

where $\xi_{mt|T},\ t = 1, \ldots, T$ are the model smoothed probabilities with $T_m = \sum_{t=1}^{T} \xi_{mt|T}$.

All estimations in this paper use the statistical software R4.0.5. For maximization of the log-likelihood function the R-package 'nloptr' provides the optimization routine 'slsqp', a sequential (least-squares) quadratic programming algorithm for nonlinearly constrained, gradient-based optimization. This algorithm supports equality constraints and inequality constraints. The former are needed to implement zero restrictions in our model setup on the structural impact matrix. The latter are used to impose a lower bound of 0.001 on the diagonal elements of Λ_m for $m = 1, \ldots, M$ to avoid singularity of the covariance matrix.

The zero restrictions in columns $(K + 1)$ of the autoregressive coefficient matrices Γ_p for $p = 1, \ldots, P$ of model 8 are implemented using restricted ordinary least squares. The restrictions are updated at the end of each maximization step of the EM algorithm.

To generate starting values for the structural parameters D and Λ_m for $m = 2, \ldots, M$ for the estimation algorithm we follow Herwartz and Lütkepohl (2014) with two exceptions. First, we choose starting values of $D = (T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t')^{1/2} \Omega$ with $\hat{u}_t$ being the estimated reduced form residuals of the respective model and Ω being a random orthogonal matrix. Choosing an orthogonal matrix instead of adding a matrix of small random numbers as suggested by Herwartz and Lütkepohl (2014) covers a wider range of the parameter space of possible starting values for D. Second, starting values of Λ_m for $m = 2, \ldots, M$ are chosen as $\Lambda_m = \mathrm{diag}(0.5k, 2.0k, 3.5k, 5.0k)^{m-1}$ with $k = 1, \ldots, 10$. For each k we draw 100 random orthogonal matrices Ω as starting values for D. Thus, the total number of distinct initial parameters for each model amounts to 1000. We check convergence of the estimation algorithm using the relative changes of the log-likelihood function for of each estimated model and choose the model that maximizes the likelihood among all converged models.

In the Monte Carlo study we rely on one draw of starting values for D to limit the computational burden. To make up for choosing only one initial parameter for D we set starting values for $\Lambda = (0.5, 2, 3.5, 5)$ to start the estimation algorithm in the proximity of the true parameter values. We did not encounter convergence problems of the estimation algorithm in the simulation study. The simulations were conducted with 50 cores (Intel Xeon Skylake 6130 processors) on the high performance computing server at Freie Universität Berlin.

A.2 Supplementary results of Monte Carlo study

This section contains supplementary material to the Monte Carlo study carried out in Section 3 of the paper. All relevant information is in the captions and notes of the respective tables.

A.2.1 Supplementary results of baseline simulation

Table A.1: Relative rejection frequencies at nominal significance level of 5% of LR-tests on exogeneity of instrument.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.07	.	.	.	.
	(0.35,0.15)	0.07	0.09	0.40	.	.
	(0.72,0.30)	0.08	0.08	0.32	0.64	0.89
	(1.00,0.40)	0.08	0.08	0.29	0.57	0.81
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	0.05	0.14	0.76	.	.
	(0.72,0.30)	0.05	0.13	0.66	0.97	1.00
	(1.00,0.40)	0.06	0.11	0.57	0.92	1.00

Notes: Based on 500 replications of simulation experiment. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock of interest (ε_t^r) than between the instrument s_t and the endogenous structural shock (ε_t^z). These cases are not taken into account in the analysis.

Table A.2: Relative rejection frequencies at nominal significance level of 10% of LR-tests for *relevance* of instrument for sample sizes $T = 200$ and $T = 500$.

Sample Size	Relevance (β_1,ρ_1,F)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.11	.	.	.	.
	(0.35,0.15)	0.96	0.96	0.96	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.00,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.00,0.40)	1.00	1.00	1.00	1.00	1.00

Notes: Based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the other structural shock (ε_t^z). These cases are not taken into account in the analysis.

Table A.3: Comparison of MSE of impulse responses to monetary policy shock.

Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)														
	(0,0)			**(0.05,0.03)**			**(0.17,0.10)**			**(0.27,0.15)**			**(0.37,0.20)**		
	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0,0)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model C	113.18	74.12	132.97	.	.	.	.	.	.	.	.	.	.	.	.
(0.35,0.15)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.01	1.07	1.01	1.01	1.05	1.01	0.98	0.82	0.98	.	.	.	.	.	.
Model C	37.66	22.56	39.75	37.70	26.73	40.38	35.82	42.07	41.48	.	.	.	.	.	.
(0.72,0.30)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.08	1.25	1.08	1.07	1.18	1.08	1.01	0.73	0.99	0.90	0.42	0.85	0.76	0.20	0.68
Model C	13.18	7.21	12.34	13.38	8.75	12.71	14.30	13.46	14.60	14.43	13.43	15.22	12.95	9.54	13.56
(1.0,0.40)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.18	1.46	1.22	1.18	1.36	1.22	1.10	0.78	1.10	0.98	0.44	0.94	0.81	0.23	0.72
Model C	7.95	4.54	7.41	8.07	5.39	7.60	8.70	7.68	8.69	9.00	7.77	9.24	8.60	6.60	8.85

Notes: The table shows the cumulated MSE of fitted models A-C relative to model A for a propagation horizon up to $h = 25$ and sample size $T = 500$. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument w_t and the target structural shock ε_t^r than between the instrument and the non-targeted structural shock ε_t^z. These cases are not taken into account in the analysis.

Table A.4: Comparison of MSE of impulse responses to monetary policy shock for a propagation horizon up to $h = 5$ and sample size $T = 200$.

Relevance ($\boldsymbol{\beta_1},\boldsymbol{\rho_1}$)	Endogeneity ($\boldsymbol{\beta_2},\boldsymbol{\rho_2}$)														
	(0.0,0.0)			**(0.05,0.03)**			**(0.17,0.10)**			**(0.27,0.15)**			**(0.37,0.20)**		
	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0.0,0.0)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.05	1.05	1.08	.	.	.	.	.	.	.	.	.	.	.	.
Model C	35.52	42.45	64.68	.	.	.	.	.	.	.	.	.	.	.	.
(0.35,0.15)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.05	1.06	1.16	1.04	1.04	1.15	1.04	0.87	1.12	.	.	.	.	.	.
Model C	20.33	21.83	31.57	20.30	21.45	30.73	18.46	20.31	26.79	.	.	.	.	.	.
(0.72,0.30)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.09	1.21	1.18	1.08	1.11	1.10	1.06	0.85	0.99	0.97	0.56	0.95	0.79	0.35	0.90
Model C	10.97	7.92	10.83	10.92	8.11	10.80	10.26	8.89	10.92	8.61	7.48	9.48	6.29	5.88	7.89
(1.0,0.40)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.14	1.34	1.23	1.14	1.24	1.22	1.05	1.08	1.10	1.02	0.76	1.05	0.97	0.48	1.09
Model C	7.19	4.70	6.21	7.15	4.72	6.16	6.99	5.70	6.40	6.61	5.24	5.70	5.62	4.55	5.28

Notes: The table shows the cumulated MSE of fitted models A-C relative to model A. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the other structural shock (ε_t^z). These cases are not taken into account in the analysis.

Table A.5: Comparison of MSE of impulse responses to monetary policy shock for a propagation horizon up to $h = 5$ and sample size $T = 500$.

Relevance ($\boldsymbol{\beta_1},\boldsymbol{\rho_1}$)	Endogeneity ($\boldsymbol{\beta_2},\boldsymbol{\rho_2}$)														
	(0.0,0.0)			**(0.05,0.03)**			**(0.17,0.10)**			**(0.27,0.15)**			**(0.37,0.20)**		
	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0.0,0.0)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model C	115.53	157.32	226.58	.	.	.	.	.	.	.	.	.	.	.	.
(0.35,0.15)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.02	1.06	1.00	1.01	1.02	1.00	0.99	0.78	0.99	.	.	.	.	.	.
Model C	47.14	44.47	61.63	45.83	48.35	60.36	37.40	57.00	50.29	.	.	.	.	.	.
(0.72,0.30)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.07	1.19	1.05	1.06	1.10	1.05	1.05	0.69	1.04	1.02	0.40	1.01	0.93	0.20	0.96
Model C	18.19	11.04	15.11	18.04	12.90	15.03	17.03	16.74	14.26	15.80	15.85	13.17	13.72	11.39	11.55
(1.0,0.40)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.16	1.66	1.31	1.16	1.53	1.32	1.15	0.91	1.33	1.12	0.52	1.30	1.01	0.22	1.03
Model C	10.76	6.59	8.51	10.72	7.67	8.51	10.45	9.56	8.36	10.07	9.21	8.07	9.25	7.62	7.64

Notes: The table shows the cumulated MSE of fitted models A-C relative to model A. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the other structural shock (ε_t^z). These cases are not taken into account in the analysis.

Table A.6: Comparison of MSE of impulse responses to monetary policy shock for a propagation horizon up to $h = 25$ and sample size $T = 200$ including model C*.

Relevance ($\boldsymbol{\beta_1},\boldsymbol{\rho_1}$)	Endogeneity ($\boldsymbol{\beta_2},\boldsymbol{\rho_2}$)														
	(0,0)			**(0.05,0.03)**			**(0.17,0.10)**			**(0.27,0.15)**			**(0.37,0.20)**		
	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0,0)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.02	1.02	1.02	.	.	.	.	.	.	.	.	.	.	.	.
Model C	31.51	24.34	36.22	.	.	.	.	.	.	.	.	.	.	.	.
Model C*	182.39	87.61	179.38	.	.	.	.	.	.	.	.	.	.	.	.
(0.35,0.15)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.01	1.03	1.02	1.01	1.02	1.03	0.95	0.78	0.92	.	.	.	.	.	.
Model C	16.48	12.73	18.02	16.33	13.06	17.87	15.10	13.64	16.44	.	.	.	.	.	.
Model C*	41.69	34.54	46.04	77.64	46.19	76.24	11595.82	6269.91	17061.19	.	.	.	.	.	.
(0.72,0.30)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.06	1.20	1.06	1.06	1.18	1.06	0.98	0.90	0.95	0.87	0.54	0.82	0.68	0.31	0.64
Model C	7.91	5.75	7.57	7.85	5.99	7.54	7.73	6.71	7.79	7.24	5.72	7.48	5.89	4.51	6.30
Model C*	7.73	9.90	8.23	8.77	13.54	9.54	9.21	10.68	9.16	12.22	9.36	12.23	8.42	7.32	9.34
(1.0,0.40)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.21	1.37	1.22	1.15	1.20	1.14	1.05	1.05	1.04	0.96	0.72	0.92	0.81	0.43	0.75
Model C	5.22	3.62	4.87	5.10	3.58	4.72	5.19	4.46	5.01	5.30	4.45	5.17	4.80	3.68	4.76
Model C*	4.86	6.44	5.03	5.50	6.61	5.19	4.50	6.76	4.76	5.43	6.95	5.90	5.75	5.88	6.27

Notes: The table shows the cumulated MSE of fitted models A-C* relative to model A. Model C* includes the instrument ordered first and is identified by a lower triangular Choleski decomposition. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the other structural shock (ε_t^z). These cases are not taken into account in the analysis.

Table A.7: Comparison of MSE of impulse responses to monetary policy shock for a propagation horizon up to $h = 25$ and sample size $T = 500$ including model C*.

	Endogeneity ($\boldsymbol{\beta_2}$,$\boldsymbol{\rho_2}$)														
Relevance ($\boldsymbol{\beta_1}$,$\boldsymbol{\rho_1}$)	**(0,0)**			**(0.05,0.03)**			**(0.17,0.10)**			**(0.27,0.15)**			**(0.37,0.20)**		
	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0,0)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model C	113.17	74.11	132.97	.	.	.	.	.	.	.	.	.	.	.	.
Model C*	1537.89	1682.47	2296.01	.	.	.	.	.	.	.	.	.	.	.	.
(0.35,0.15)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.01	1.07	1.01	1.01	1.05	1.01	0.98	0.82	0.98	.	.	.	.	.	.
Model C	37.66	22.56	39.76	37.70	26.73	40.39	35.82	42.07	41.48	.	.	.	.	.	.
Model C*	56.37	48.58	60.75	62.46	51.49	67.51	179.39	130.00	191.96	.	.	.	.	.	.
(0.72,0.30)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.08	1.25	1.08	1.07	1.18	1.08	1.01	0.73	0.99	0.90	0.42	0.85	0.76	0.20	0.68
Model C	13.18	7.21	12.34	13.38	8.75	12.71	14.30	13.46	14.60	14.44	13.43	15.23	12.95	9.49	13.57
Model C*	12.66	13.60	14.88	10.64	13.99	11.21	13.52	20.71	15.05	16.98	21.14	19.15	18.16	15.15	19.61
(1.0,0.40)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.18	1.46	1.22	1.18	1.36	1.22	1.10	0.78	1.10	0.98	0.44	0.94	0.81	0.23	0.72
Model C	7.95	4.54	7.41	8.07	5.39	7.61	8.70	7.68	8.69	9.00	7.77	9.24	8.60	6.60	8.85
Model C*	6.28	7.48	6.88	5.97	7.77	6.27	7.48	10.95	8.08	9.72	11.62	10.63	11.22	10.20	11.93

Notes: The table shows the cumulated MSE of fitted models A-C* relative to model A. Model C* includes the instrument ordered first and is identified by a lower triangular Choleski decomposition. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the other structural shock (ε_t^z). These cases are not taken into account in the analysis.

A.2.2 Tests for alternative DGPs

This subsection contains robustness checks and extensions of the simulation study.

A.2.2.1 Confounding common shift in variances

Table A.8: Relative rejection frequencies at nominal significance level of 10% of exogeneity test for common confounding shift.

Sample Size	Relevance ($h\beta_1$,$h\rho_1$)	Endogeneity ($h\beta_2$,$h\rho_2$)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.09	.	.	.	.
	(0.35,0.15)	0.12	0.14	0.27	.	.
	(0.72,0.30)	0.14	0.14	0.24	0.31	0.43
	(1.0,0.40)	0.13	0.14	0.19	0.29	0.35
T = 500	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.12	0.17	0.62	.	.
	(0.72,0.30)	0.11	0.13	0.38	0.63	0.80
	(1.0,0.40)	0.10	0.11	0.27	0.48	0.67

Table A.9: Relative rejection frequencies at nominal significance level of 5% of relevance test for confounding common shift.

Sample Size	Relevance ($h\beta_1$,$h\rho_1$)	Endogeneity ($h\beta_2$,$h\rho_2$)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.09	.	.	.	.
	(0.35,0.15)	0.96	0.96	0.97	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.05	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.2 Alternative structural shock variances

Table A.10: Relative rejection frequencies at nominal significance level of 10% of LR-Tests of exogeneity of instrument for alternative structural shock variances.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.20,0.12)	(0.30,0.16)	(0.40,0.22)
T = 200	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.13	0.16	0.26	.	.
	(0.72,0.30)	0.14	0.15	0.22	0.31	0.38
	(1.0,0.40)	0.13	0.13	0.19	0.27	0.35
T = 500	(0,0)	0.11	.	.	.	.
	(0.35,0.15)	0.12	0.14	0.57	.	.
	(0.72,0.30)	0.11	0.13	0.33	0.61	0.76
	(1.0,0.40)	0.09	0.11	0.22	0.43	0.64

Table A.11: Relative rejection frequencies at nominal significance level of 5% of LR-test of instrument relevance for alternative structural shock variances.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.09	.	.	.	.
	(0.35,0.15)	0.93	0.94	0.94	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.3 Censored instrument observations

Table A.12: Relative rejection frequencies at nominal significance level of 10% for exogeneity test with 60% censored instrument observations.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.14	0.15	0.40	.	.
	(0.72,0.30)	0.14	0.17	0.37	0.62	0.82
	(1.0,0.40)	0.15	0.17	0.36	0.55	0.75
T = 500	(0,0)	0.08	.	.	.	.
	(0.35,0.15)	0.08	0.17	0.75	.	.
	(0.72,0.30)	0.10	0.18	0.68	0.96	1.00
	(1.0,0.40)	0.10	0.16	0.61	0.90	1.00

Table A.13: Relative rejection frequencies at nominal significance level of 5% for relevance test with 60% censored instrument observations.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.04	.	.	.	.
	(0.35,0.15)	0.84	0.83	0.84	.	.
	(0.72,0.30)	0.98	0.98	0.97	0.96	0.96
	(1.0,0.40)	0.98	0.99	0.98	0.98	0.97
T = 500	(0,0)	0.04	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

Table A.14: Relative rejection frequencies at nominal significance level of 10% for exogeneity test with 90% censored instrument observations.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2) (0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.08	.	.	.	.
	(0.35,0.15)	0.11	0.11	0.20	.	.
	(0.72,0.30)	0.17	0.16	0.25	0.37	0.56
	(1.0,0.40)	0.17	0.17	0.27	0.37	0.51
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	0.07	0.09	0.41	.	.
	(0.72,0.30)	0.08	0.10	0.37	0.70	0.93
	(1.0,0.40)	0.09	0.10	0.35	0.65	0.88

Table A.15: Relative rejection frequencies at nominal significance level of 5% for relevance test with 90% censored instrument observations.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2) (0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.05	.	.	.	.
	(0.35,0.15)	0.61	0.59	0.59	.	.
	(0.72,0.30)	0.96	0.95	0.94	0.93	0.92
	(1.0,0.40)	0.97	0.97	0.97	0.97	0.96
T = 500	(0,0)	0.05	.	.	.	.
	(0.35,0.15)	0.95	0.96	0.94	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.4 Two instruments for one shock

If there are q instruments, w_t and ν_t in

$$w_t = \beta\varepsilon_t + \eta\nu_t, \tag{A.1}$$

are $q \times 1$ column vectors, where ν_t are independent potentially heteroskedastic measurement errors, and β and η matrices with dimension $q \times q$. In the simulation study, we consider the case of $q = 2$ for the identification of one shock. As for $q = 1$, we order the structural shock of interest first.

Validity of the q instruments requires the following two conditions:

$$\beta_{\cdot i} = 0_{q\times 1} \quad \forall\ i = 2, \ldots, K, \tag{A.2}$$
$$\beta_{\cdot 1} \neq 0_{q\times 1}, \tag{A.3}$$

where $\beta_{\cdot i}$ indicates the i-th column of β. Equation (A.2) implies exogeneity and (A.3) relevance. If both are met, the instruments are valid.

The augmented VAR is

$$z_t = \delta + \Gamma(L)z_{t-1} + e_t, \tag{A.4}$$

where $z_t = [y_t', w_t]'$ is a $((K + q) \times 1)$-vector of observable variables, $\Gamma(L)$ is a (potentially restricted) lag matrix polynomial capturing the autoregressive component of the model, δ is a $((K + q) \times 1)$-vector of constant terms, and e_t are $(K + q)$-dimensional serially uncorrelated residuals. The latter are related to the structural innovations μ_t as

$$\begin{aligned} e_t &= D\mu_t \\ &= \begin{bmatrix} B_{(K\times K)} & 0_{(K\times q)} \\ \beta_{(q\times K)} & \eta_{(q\times q)} \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}. \end{aligned} \tag{A.5}$$

Using (A.5), the VAR in structural form is

$$z_t = \delta + \Gamma(L)z_{t-1} + D\mu_t. \tag{A.6}$$

Estimation and identification of (A.5) remains as described in Section 2. Testing the validity of the q instruments can also be done with LR-tests. To test the exogeneity condition (A.2), we compare the likelihood of a model with $\beta = (\beta_{\cdot 1}, 0_{q\times 1}, \ldots, 0_{q\times 1})$, against a model with β fully unrestricted. Formally, we test

$$H_0 : \beta_{\cdot 2} = \cdots = \beta_{\cdot K} = 0_{q\times 1}$$
$$H_1 : \exists j \in \{2, \ldots, K\} \text{ s.t. } \beta_{\cdot \mathrm{j}} \neq 0_{\mathrm{q}\times 1}.$$

This is a joint test of exogeneity of both instruments. Rejecting the null indicates the endogeneity of at least one instrument.

To test the relevance condition (A.3), we compare a restricted model version with $\beta_{\cdot 1} = 0_{q\times 1}$ to a model with $\beta_{\cdot 1}$ unrestricted. Under both the null and the alternative hypothesis $\beta_{\cdot 2} = \cdots = \beta_{\cdot K} = 0_{q\times 1}$. Formally, we test

$$\begin{aligned} &H_0 : \beta_{\cdot 1} = 0_{q\times 1} \\ &H_1 : \beta_{\cdot 1} \neq 0_{q\times 1}. \end{aligned}$$

This is a joint test of relevance of both instruments. Rejecting the null indicates the relevance of at least one instrument. If the instruments are both exogenous and relevant, they are valid. Then, we set $\beta = (\beta_{\cdot 1}, 0_{q\times 1}, \ldots, 0_{q\times 1})$.

Table A.16: Relative rejection frequencies at nominal significance level of 10% of LR-tests for exogeneity with second (strong and exogenous) instrument ($\rho_1 = 0.3, \rho_2 = 0$).

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0.0,0.0)	0.13	.	.	.	.
	(0.35,0.15)	0.15	0.18	0.50	.	.
	(0.72,0.30)	0.14	0.17	0.42	0.75	0.95
	(1.0,0.40)	0.13	0.15	0.39	0.67	0.90
T = 500	(0.0,0.0)	0.10	.	.	.	.
	(0.35,0.15)	0.10	0.18	0.82	.	.
	(0.72,0.30)	0.09	0.16	0.78	0.99	1.00
	(1.0,0.40)	0.10	0.16	0.73	0.98	1.00

Table A.17: Relative rejection frequencies at nominal significance level of 5% of LR-tests for relevance with second (strong and exogenous) instrument ($\rho_1 = 0.3, \rho_2 = 0$).

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0.0,0.0)	0.98	.	.	.	.
	(0.35,0.15)	0.98	0.98	0.98	.	.
	(0.72,0.30)	0.99	0.99	1.00	1.00	1.00
	(1.00,0.40)	0.99	1.00	0.99	1.00	1.00
T = 500	(0.0,0.0)	1.00	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.00,0.40)	1.00	1.00	1.00	1.00	1.00

Table A.18: Relative rejection frequencies at nominal significance level of 10% of LR-tests for exogeneity with second (endogenous) instrument ($\rho_1 = 0.3, \rho_2 = 0.15$).

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.66	.	.	.	.
	(0.35,0.15)	0.71	0.66	0.71	.	.
	(0.72,0.30)	0.73	0.67	0.65	0.76	0.89
	(1.0,0.40)	0.75	0.71	0.64	0.72	0.82
T = 500	(0,0)	0.97	.	.	.	.
	(0.35,0.15)	0.99	0.97	0.99	.	.
	(0.72,0.30)	0.99	0.98	0.98	0.99	1.00
	(1.0,0.40)	0.99	0.99	0.98	0.98	1.00

Table A.19: Relative rejection frequencies at nominal significance level of 5% of LR-tests for relevance with second (endogenous) instrument ($\rho_1 = 0.3, \rho_2 = 0.15$).

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.99	.	.	.	.
	(0.35,0.15)	0.99	0.98	0.99	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	1.00	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.5 DGP is MS(3)

Table A.20: Relative rejection frequencies at nominal significance level of 10% for exogeneity test and DGP with M = 3 Markov states.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.11	.	.	.	.
	(0.35,0.15)	0.12	0.13	0.46	.	.
	(0.72,0.30)	0.12	0.13	0.40	0.72	0.90
	(1.0,0.40)	0.12	0.12	0.35	0.68	0.84
T = 500	(0,0)	0.12	.	.	.	.
	(0.35,0.15)	0.10	0.22	0.85	.	.
	(0.72,0.30)	0.10	0.20	0.81	0.99	1.00
	(1.0,0.40)	0.09	0.18	0.75	0.98	1.00

Table A.21: Relative rejection frequencies at nominal significance level of 5% relevance test for DGP with M = 3 Markov states.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.08	.	.	.	.
	(0.35,0.15)	0.93	0.93	0.92	.	.
	(0.72,0.30)	1.00	1.00	1.00	0.99	0.99
	(1.0,0.40)	1.00	1.00	1.00	1.00	0.99
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.6 DGP with failure of constancy of impact effects matrix

Table A.22: Relative rejection frequencies at nominal significance level of 10% of exogeneity test for failure of constancy of impact effects.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.11	0.14	0.47	.	.
	(0.72,0.30)	0.15	0.17	0.43	0.73	0.90
	(1.0,0.40)	0.17	0.19	0.40	0.66	0.85
T = 500	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.13	0.22	0.86	.	.
	(0.72,0.30)	0.20	0.29	0.80	0.97	1.00
	(1.0,0.40)	0.25	0.33	0.73	0.94	0.99

Table A.23: Relative rejection frequencies at nominal significance level of 5% of relevance test for mild failure of constancy of impact effects.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.05	.	.	.	.
	(0.35,0.15)	0.97	0.97	0.97	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.7 DGP with smooth transition in variances

Our parameter choice in the simulations of $\gamma = -sqrt(T)/10$ and $c = 0.5T$ (Section A.4) ensures a variance change roughly in the middle of the sample. The choice implies that the variance change is largely completed over about 20% of the respective sample size.

Table A.24: Relative rejection frequencies at nominal significance level of 10% for exogeneity test and DGP with smooth transition in variances.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.12	.	.	.	.
	(0.35,0.15)	0.12	0.17	0.51	.	.
	(0.72,0.30)	0.14	0.15	0.48	0.77	0.94
	(1.0,0.40)	0.13	0.15	0.43	0.72	0.92
T = 500	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.10	0.19	0.82	.	.
	(0.72,0.30)	0.11	0.20	0.77	0.99	1.00
	(1.0,0.40)	0.10	0.18	0.71	0.97	1.00

Table A.25: Relative rejection frequencies at nominal significance level of 5% for relevance test and DGP with smooth transition in variances.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0.,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.05	.	.	.	.
	(0.35,0.15)	0.89	0.89	0.88	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.2.2.8 DGP with exogenous break in variances

For a description of the model see Section A.4.

Table A.26: Relative rejection frequencies at nominal significance level of 10% for exogeneity test and DGP with exogenous break in variances at $0.5T$.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2) (0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.11	.	.	.	.
	(0.35,0.15)	0.11	0.16	0.53	.	.
	(0.72,0.30)	0.12	0.15	0.51	0.79	0.96
	(1.0,0.40)	0.12	0.14	0.46	0.75	0.93
T = 500	(0,0)	0.10	.	.	.	.
	(0.35,0.15)	0.10	0.20	0.85	.	.
	(0.72,0.30)	0.10	0.19	0.80	0.99	1.00
	(1.0,0.40)	0.10	0.18	0.75	0.98	1.00

Table A.27: Relative rejection frequencies at nominal significance level of 5% for relevance test and DGP with exogenous break in variances at $0.5T$.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2) (0,0)	(0.05,0.03)	(0.17,0.10)	(0.27,0.15)	(0.37,0.20)
T = 200	(0,0)	0.06	.	.	.	.
	(0.35,0.15)	0.92	0.92	0.91	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00
T = 500	(0,0)	0.04	.	.	.	.
	(0.35,0.15)	1.00	1.00	1.00	.	.
	(0.72,0.30)	1.00	1.00	1.00	1.00	1.00
	(1.0,0.40)	1.00	1.00	1.00	1.00	1.00

A.3 Additional results of baseline model

This section contains additional results for the baseline model.

Table A.28: Estimates and standard errors of matrix of transition probabilities.

$$\hat{P} = \begin{bmatrix} 0.951 \ (0.029) & 0.302 \ (.) \\ 0.049 \ (0.165) & 0.698 \ (.) \end{bmatrix}$$

Notes: Estimates of transition matrix P for baseline heteroskedastic proxy-SVAR(6) with M = 2 states with $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The standard errors (in parentheses) are obtained from the inverse of the negative Hessian evaluated at the optimum of the structural model. Standard errors are available only for $K-1$ columns of matrix P as each row need to sum up to 1 and hence the K^{th}-element of each row is not estimated. The respective entries for the standard errors are marked with (.).

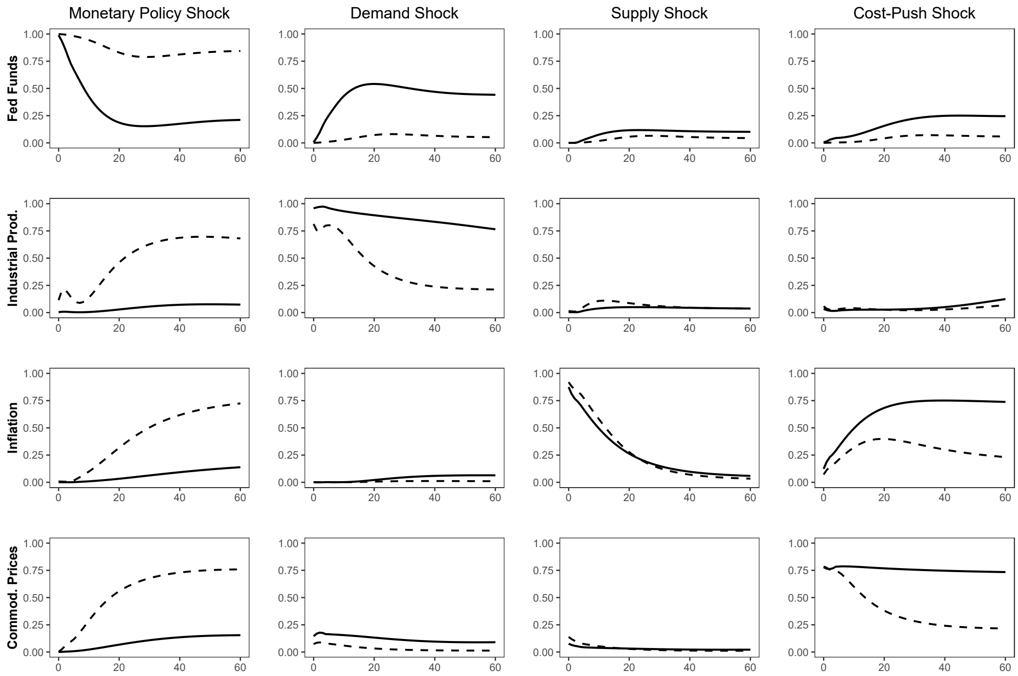
Table A.29: Estimates and standard errors of matrix of instantaneous impact effect matrix D.

Estimates				
-0.001	0.020	-0.004	0.001	0
-0.039	0.003	0.003	0.001	0
0.002	0.002	0.001	0.004	0
0.001	0.038	0.087	-0.027	0
-0.137	0	0	0	0.375
S.E.				
0.001	0.001	0.005	0.003	.
0.002	0.003	0.003	0.004	.
0.001	0.001	0.001	0.002	.
0.004	0.022	0.013	0.027	.
0.039	.	.	.	0.024

Notes: Estimates of instantaneous impact effect matrix D in the upper panel for baseline heteroskedastic proxy-SVAR(6) with M = 2 states with $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The standard errors in the lower panel are obtained from the inverse of the negative Hessian evaluated at the optimum of the structural model.

The regime-specific forecast error variance decompositions in Figure A.1 quantify the economic importance of monetary policy shocks for output and credit spread fluctuations. It shows that the contribution of monetary shocks to the variability of the endogenous variables is highly state-dependent. In the high volatility regime, monetary shocks account for up more than half of the variance of production and prices at longer horizons. In the low volatility regime, they each explain between 10% and 25%. The monetary shocks also account for a much larger share of the variance in the federal funds rate in the high volatility regime than in the low volatility regime.

Figure A.1: Variance decompositions for heteroskedastic proxy-SVAR.



Notes: The figure shows the regime-specific forecast error variance decompositions (solid line - state 1; dashed line - state 2) for the structural shocks in columns on the endogenous variables in rows for the baseline Markov switching heteroskedastic proxy-SVAR(6) model with $M = 2$ states and $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The sample is 1980M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004).

A.4 Sensitivity analysis of baseline model

In this section we show various sensitivity results for the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states, $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$ and sample 1980M1-2007M6. The model is described in detail in Section 4 of the paper. More information is in the captions and notes of the figures and tables.

A.4.1 Alternative volatility models

Various approaches have been used in the SVAR literature to identify the structural parameters by time-varying volatility besides the Markov switching in covariances, which we apply in our baseline specification. The following expositions are taken, apart from minor notational adjustments, from Lütkepohl and Netšunajev (2017) and Lütkepohl and Schlaak (2018) and explain the different approaches briefly. For a more detailed overview we refer to the respective papers.

Exogenous volatility changes

One alternative to model time-varying volatility is to assume that the changes of the covariance occur at prespecified break dates,

$$\mathbb{E}(e_t e_t') = \tilde{\Sigma}_t = \tilde{\Sigma}_m \quad \text{for} \quad t \in \mathcal{T}_m, \quad m = 1, \ldots, M, \tag{A.7}$$

where $\mathcal{T}_m = \{T_{m-1} + 1, \ldots, T_m\}$ $(m = 1, \ldots, M)$ are M given volatility regimes of consecutive time periods. The T_m, for $m = 1, \ldots, M-1$, represent the time periods of volatility changes with $T_0 = 0$ and $T_M = T$. The change points T_m may be predetermined by some statistical procedure, for example, Chow breakpoint tests. Lanne, Lütkepohl and Maciejowska (2010) state the identification conditions for this model. Assuming Gaussian residuals u_t, the log-likelihood function is

$$\log l(\boldsymbol{\xi}, \boldsymbol{\sigma}) = -\frac{KT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log \det(\tilde{\Sigma}_t) - \frac{1}{2} \sum_{t=1}^{T} e_t' \tilde{\Sigma}_t^{-1} e_t, \tag{A.8}$$

where $\boldsymbol{\xi} = \text{vec}[\nu, A_1, \ldots, A_p]$ and $\boldsymbol{\sigma}$ contains all unknown covariance parameters.

Smooth transition in variances

Alternatively, one may model the change in the residual covariance matrix as a smooth transition from a volatility regime with a positive definite covariance matrix $\tilde{\Sigma}_1$ to a regime with $\tilde{\Sigma}_2$ such that

$$\mathbb{E}(e_t e_t') = \tilde{\Sigma}_t = (1 - G(\gamma, c, w_t))\tilde{\Sigma}_1 + G(\gamma, c, w_t)\tilde{\Sigma}_1, \tag{A.9}$$

where the transition function $G(\gamma, c, w_t) = (1+\exp[-\exp(\gamma)(w_t - c)])^{-1}$ is a logistic function that depends on the smoothness parameter γ, the location parameter c and a transition variable w_t. In this setup, small values of the smoothness parameter γ imply a slow, gradual transition from one volatility regime to the other. When the smoothness parameter becomes very large, however, the transition resembles a step function with a discrete change between volatility states. A locally unique decomposition of the reduced form covariance matrices is obtained if $\tilde{\Sigma}_1 = DD'$ and $\tilde{\Sigma}_2 = D\Lambda D'$, where the diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_K)$ has distinct, strictly positive values λ_k $(k = 1, \ldots, K)$. This model was proposed and used by Lütkepohl and Netšunajev (2014) in the context of SVAR analysis.

For Gaussian e_t, the log-likelihood can be written as in (A.8). It is now a function of the transition parameters as well. Lütkepohl and Netšunajev (2014) propose an iterative procedure for estimation. Since the range of the smoothness and threshold parameters $\{\gamma, c\}$ can be bounded, a grid search can be performed over the relevant range of these two parameters.

Multivariate GARCH

Multivariate GARCH processes offer yet another possibility to model time-varying volatility. In the context of SVAR analysis the GO-GARCH model proposed by van der Weide (2002) is typically used. It specifies the volatility changes as

$$\mathbb{E}(e_t e_t' | \mathcal{F}_{t-1}) = \tilde{\Sigma}_{t|t-1} = D\Lambda_{t|t-1} D'. \tag{A.10}$$

Here $\mathcal{F}_t$ denotes the information available at time t,

$$\Lambda_{t|t-1} = \operatorname{diag}(\sigma^2_{1,t|t-1}, \ldots, \sigma^2_{K,t|t-1})$$

is a diagonal matrix with univariate GARCH(1,1) diagonal elements,

$$\sigma^2_{k,t|t-1} = (1 - \gamma_k - g_k) + \gamma_k \varepsilon^2_{k,t-1} + g_k \sigma^2_{k,t-1|t-2}, \quad k = 1, \ldots, K, \tag{A.11}$$

where $\varepsilon_{k,t} = b_k^* e_t$ and b_k^* is the k^{th} row of B^{-1} $(k = 1, \ldots, K)$. Moreover, $g_k \geq 0$, $\gamma_k > 0$, $g_k + \gamma_k < 1$ $(k = 1, \ldots, K)$. The model has been proposed and used for SVAR analysis by Normandin and Phaneuf (2004) and Bouakez and Normandin (2010), for example. Identification conditions for uniqueness of B are stated in Sentana and Fiorentini (2001) and Milunovich and Yang (2013).

The setup of the model implies an unconditional residual covariance matrix $\mathbb{E}(e_t e_t') = \tilde{\Sigma} = DD'$. Under Gaussian assumptions for the $\varepsilon_{k,t}$ the log-likelihood of the model is $\log l = \sum_{t=1}^{T} \log f_{t|t-1}(y_t)$, where the conditional densities have the form

$$f_{t|t-1}(y_t) = (2\pi)^{-K/2} \det(\tilde{\Sigma}_{t|t-1})^{-1/2} \exp\left(-\frac{1}{2} e_t' \tilde{\Sigma}^{-1}_{t|t-1} e_t\right). \tag{A.12}$$

The log-likelihood function is highly nonlinear which makes the maximization and, thus, ML estimation of the GARCH parameters and the impact effects matrix D numerically challenging. We follow the existing literature and apply the two-step algorithm for ML estimation described by Lanne and Saikkonen (2007). In the first step, the estimation procedure is broken down in univariate GARCH estimations to get initial estimates of the parameters of the volatility model and in the second step, a full, joint ML estimation of the parameters is performed starting from the initial estimates obtained in the first step.

A.4.2 Smooth transition in variances

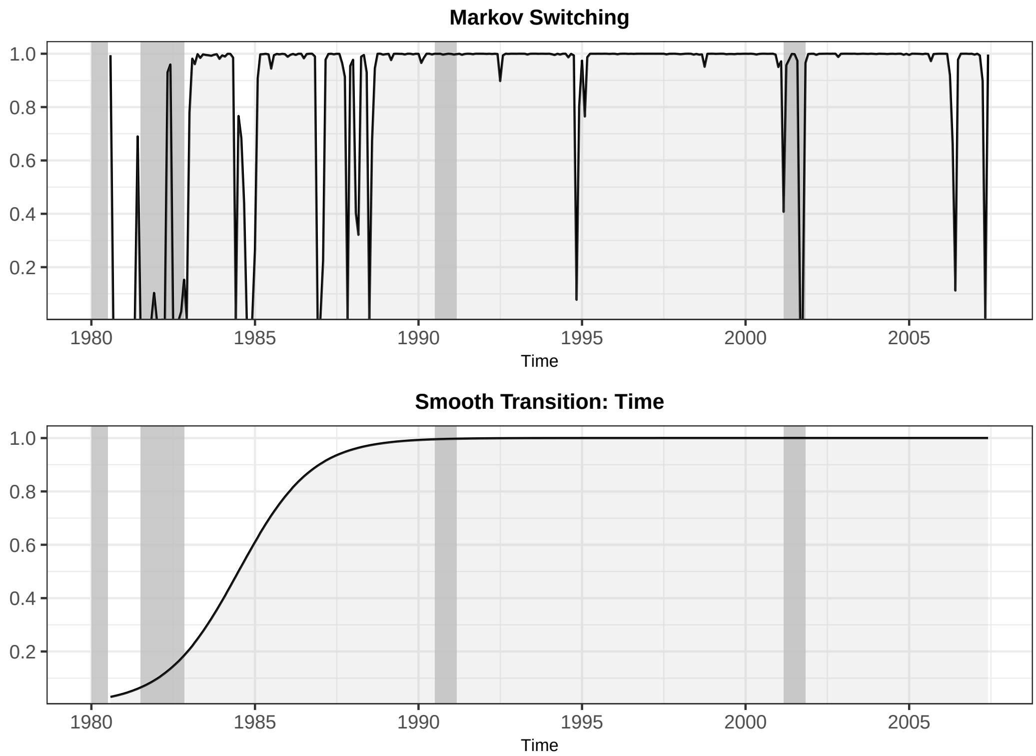
In this subsection, we assess whether our results are sensitive to a smooth transition in variances model with time as transition variable. Figure A.2 compares the volatility states across models. The smooth transition model estimates roughly similar states as the MS model. It matches the transition of the Fed chairmanship from Volcker to Greenspan. Accordingly, Table A.30 shows that the instrument is also valid when using this volatility model. Finally, Figure A.3 illustrates that the estimated effects of monetary policy shocks are similar to those from the baseline MS model as well.

Table A.30: Instrument validity for smooth transition in variance model.

	Exogeneity	Relevance
LR-statistic	2.285	60.829
p-value	0.515	0.000
Restrictions	3	1

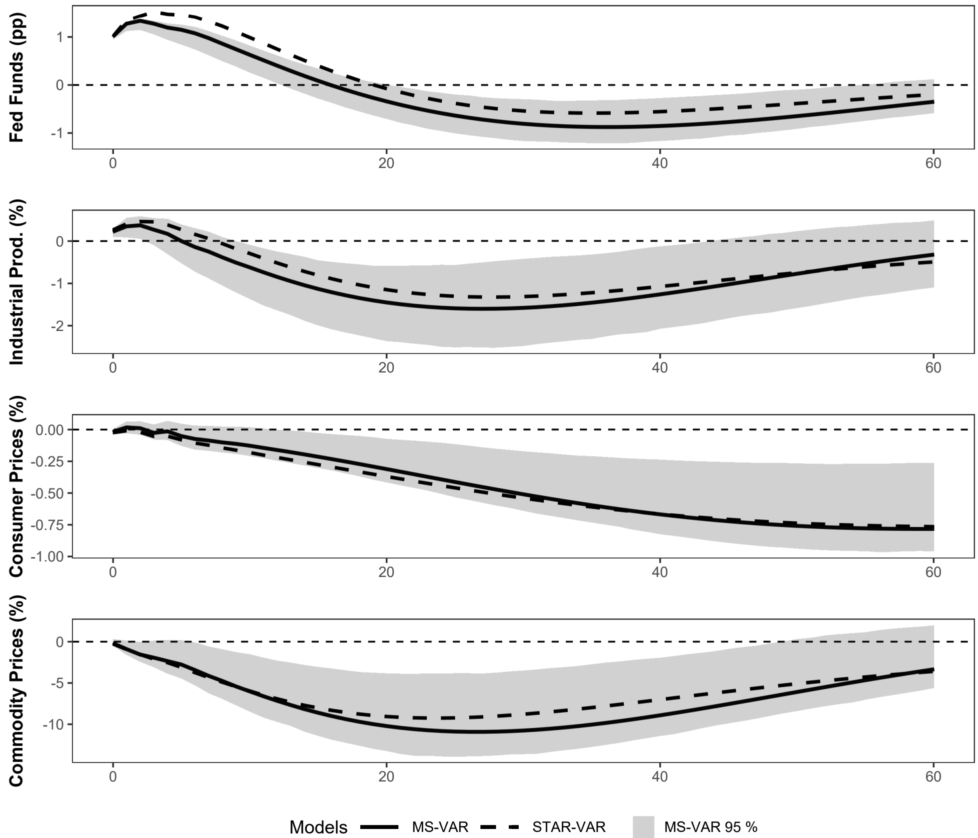
Notes: The table shows the LR-statistic, the p-value and the number of restrictions for the tests of instrument exogeneity ($H_0: \beta_2 = \cdots = \beta_K = 0$, $H_1: \beta$ unrestricted) and instrument relevance ($H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$). The instrument is the narrative-based measure of monetary surprises of Romer and Romer (2004).

Figure A.2: Volatility states 2 of Markov switching and smooth transition model.



Notes: The figure shows the state 2 probability of heteroskedastic proxy-SVARs, using the baseline Markov switching heteroskedastic proxy-SVAR(6) model with m = 2 states (upper panel) and smooth transition in variances SVAR(6) based on time as transition variable (lower panel). The model is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The shaded vertical bars mark recession periods defined by the NBER.

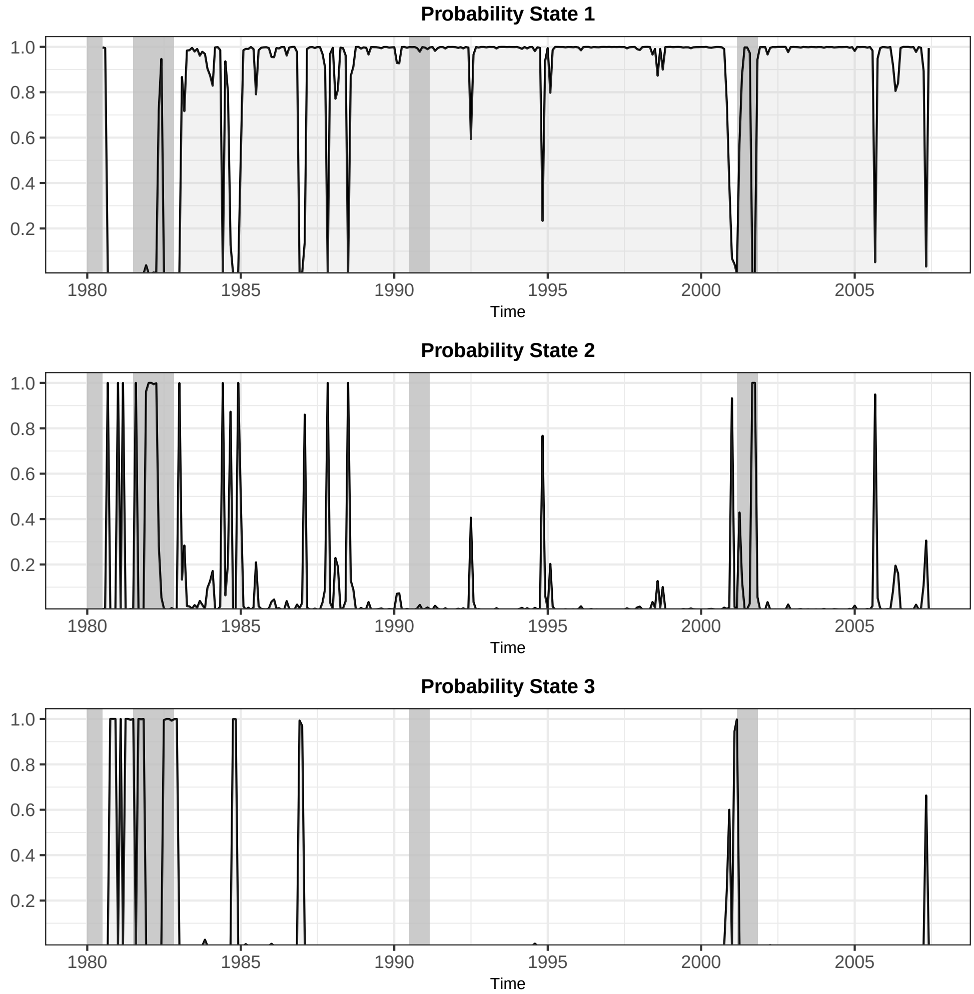
Figure A.3: Comparison of smooth transition with baseline Markov switching model.



Notes: The figure shows impulse responses to a 100 basis points monetary policy shock of heteroskedastic proxy-SVAR models using the baseline Markov Switching heteroskedastic proxy-SVAR(6) model with $M = 2$ states (shaded areas and dash dotted lines) and a smooth transition in variances with time as transition variable (dashed line). The shaded area denotes 95% pointwise confidence intervals based on 5,000 bootstrap replications of the baseline Markov switching proxy-SVAR(6).

A.4.3 Robustness of Markov switching proxy-SVAR

Figure A.4: Smoothed state probabilities of Markov switching proxy-SVAR(6) model with $M = 3$ states.



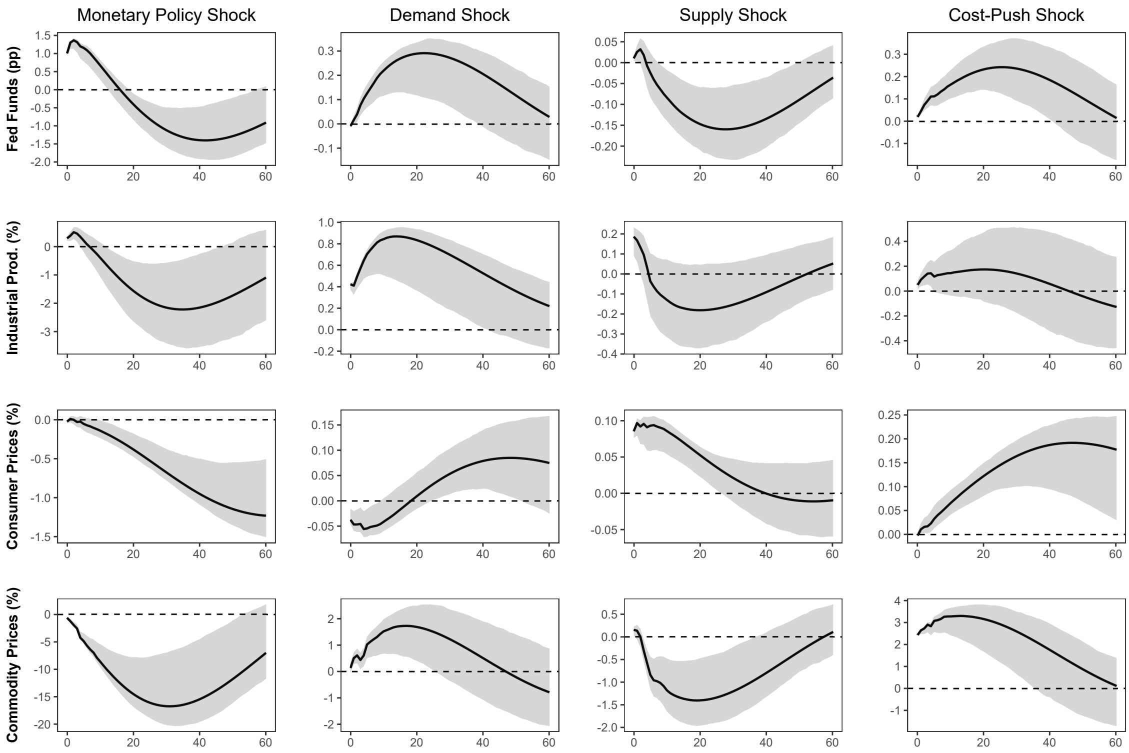
Notes: The figure shows the estimated smoothed state probabilities $\hat{\xi}_{mt|T}$ for $m = 1$ in the upper panel, for $m = 2$ in the middle panel, and for $m = 3$ in the lower panel, where $t = 1, \dots, T$, of the Markov switching proxy-SVAR(6) model with $M = 3$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The shaded vertical bars mark recession periods as defined by the NBER.

Table A.31: Instrument validity for MS(3)-VAR(6) model.

	Exogeneity	Relevance
LR-statistic	1.405	33.547
p-value	0.704	0.000
Restrictions	3	1

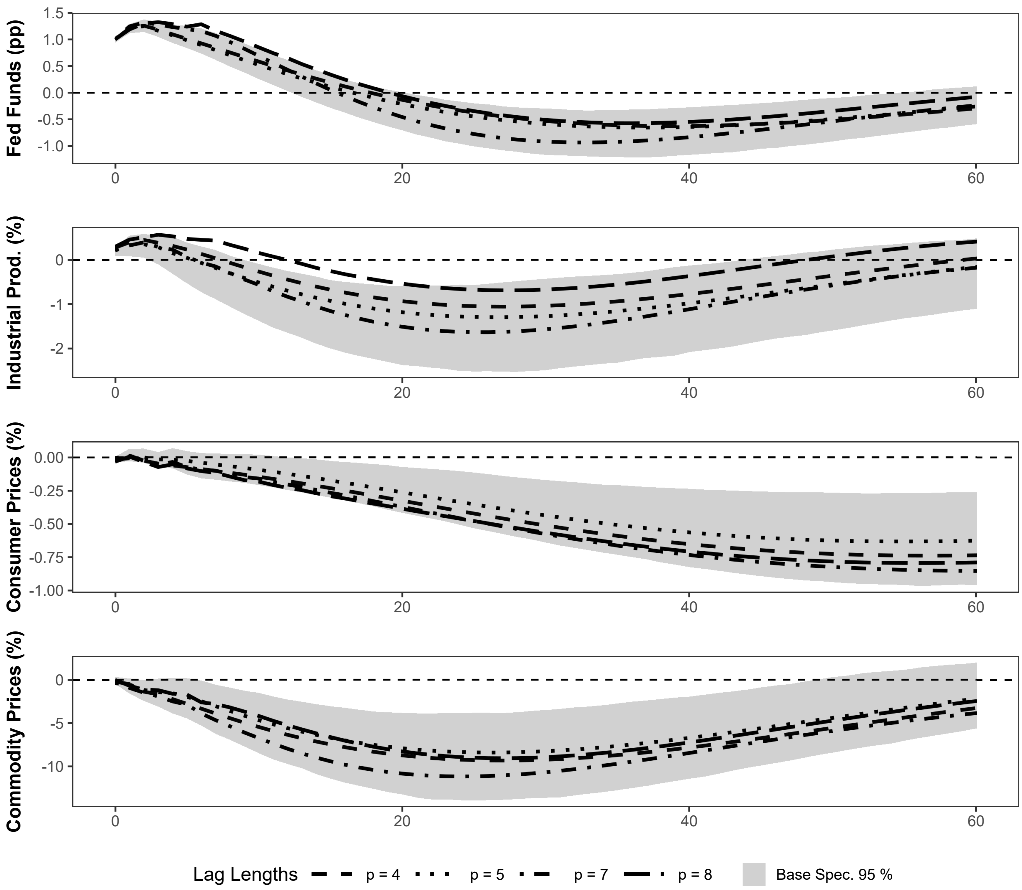
Notes: The table shows the LR statistic, p-value and number of restrictions of the test for instrument exogeneity (H_0 : $\beta_2 = \cdots = \beta_K = 0$, H_1 : β unrestricted) and for instrument relevance (H_0 : $\beta_1 = 0$, H_1 : $\beta_1 \neq 0$) for a Markov switching proxy-SVAR(6) model with $M = 3$ states. The instrument is the narrative-based measure of monetary surprises of Romer and Romer (2004).

Figure A.5: Impulse responses for Markov switching proxy-SVAR(6) model with $M = 3$ states.



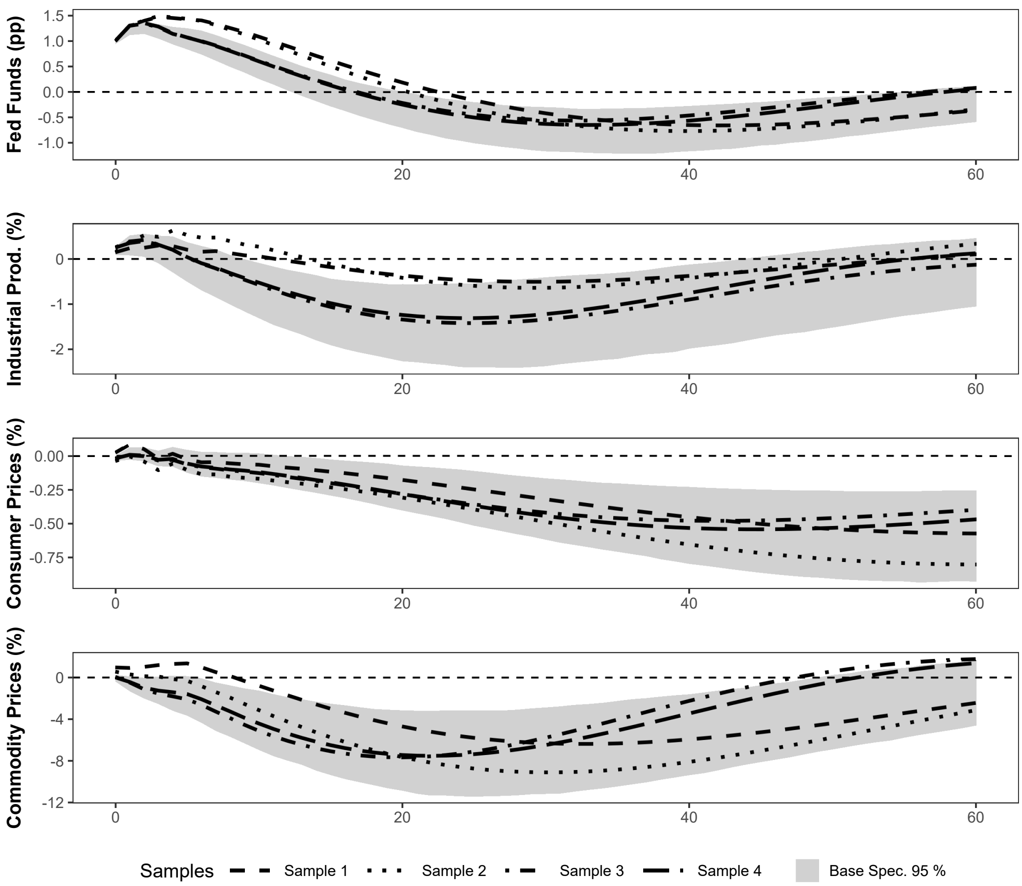
Notes: The figure shows the impulse responses to one standard deviation shocks in state $m = 1$ of the Markov switching proxy-SVAR(6) model with $M = 3$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The sample is 1980M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications.

Figure A.6: Sensitivity analysis of baseline model using different lag lengths.



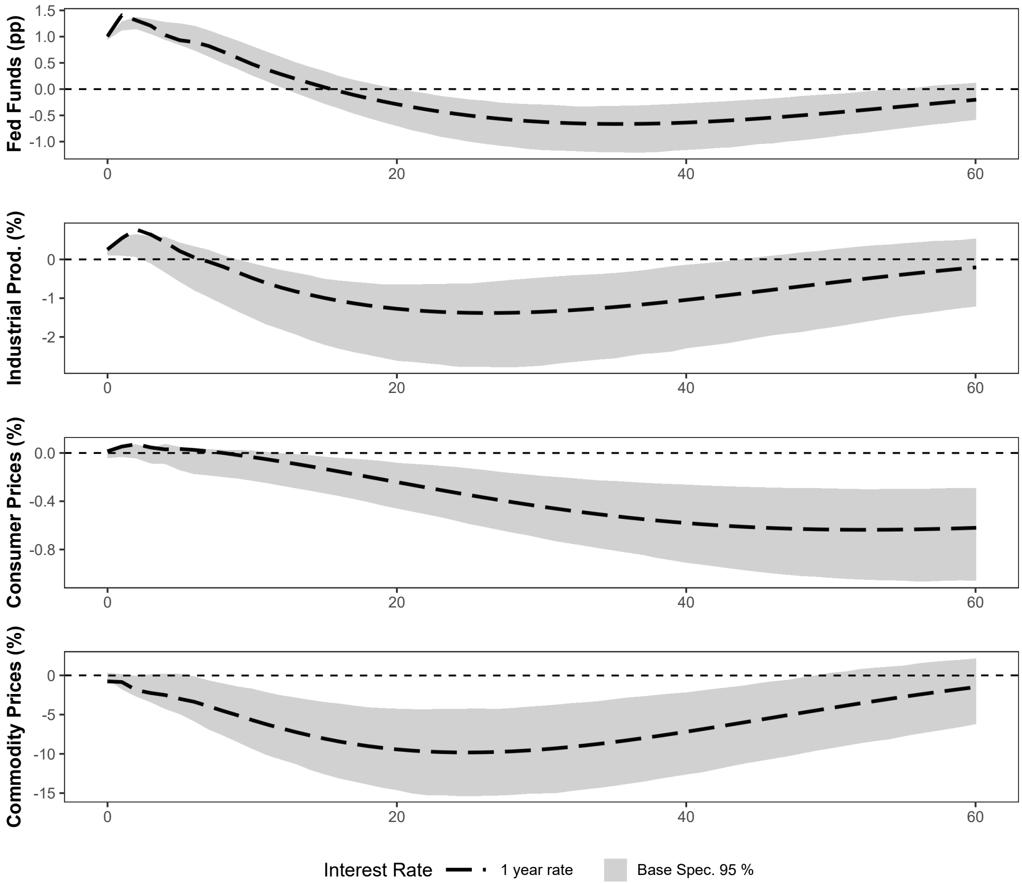
Notes: The figure shows the impulse responses to a monetary policy shock of 100 basis points in state $m = 1$ of the heteroskedastic proxy-SVAR(p) with $M = 2$ states with $p = 4, 5, 7, 8$. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The sample is 1980M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states.

Figure A.7: Sensitivity analysis of baseline model using different sample periods.



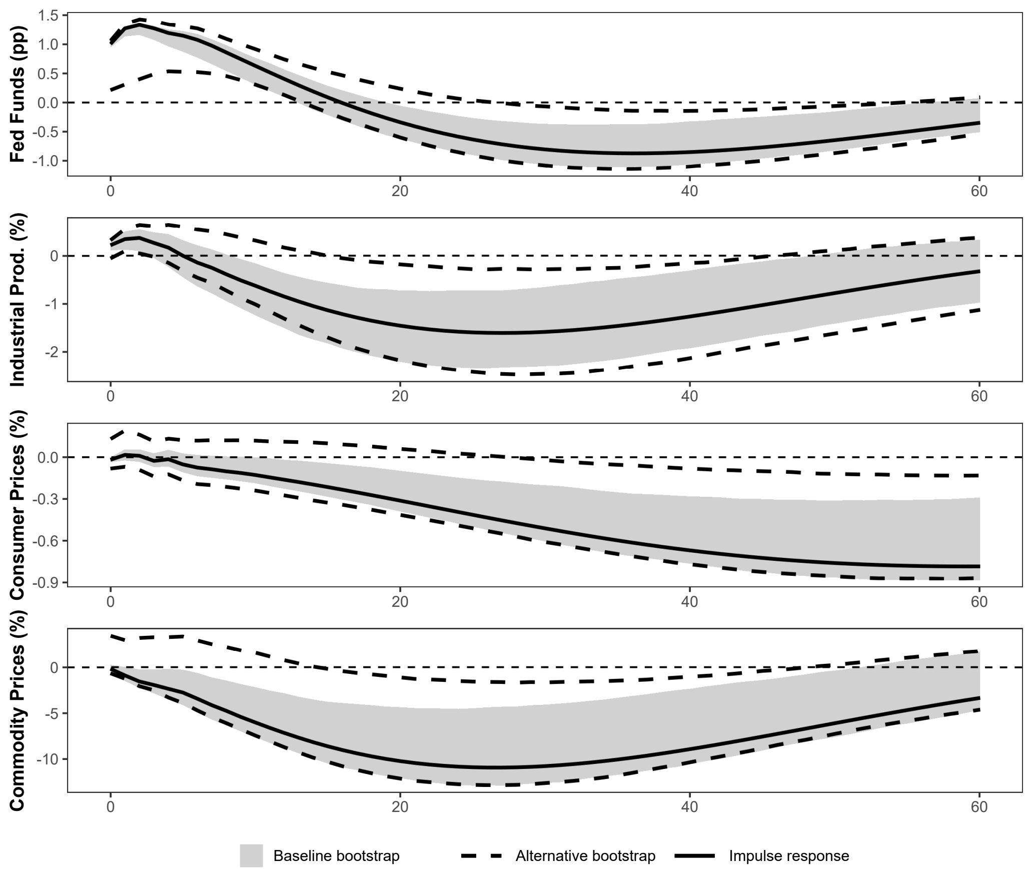
Notes: The figure shows the impulse responses to a monetary policy shock of 25 basis points in state $m = 1$ of the heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$ for different sample specifications. Sample 1 refers to 1982M1-2007M6, Sample 2 refers to 1981M1-2007M6, Sample 3 refers to 1980M1-2005M6, Sample 4 refers to 1980M1-2006M6. The instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states for the sample 1980M1-2007M6.

Figure A.8: Sensitivity analysis of baseline model using alternative indicator for monetary policy.



Notes: The figure shows the impulse responses to a monetary policy shock of 25 basis points in state $m = 1$ of the heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$, where $1yr_t$ refers to the US government bond yield with one year maturity. The sample is 1980M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states and $z_t = [ip_t, ff_t, pce_t, prm_t, rr_t]'$.

Figure A.9: Sensitivity analysis of baseline model using a residual-based moving block bootstrap method.

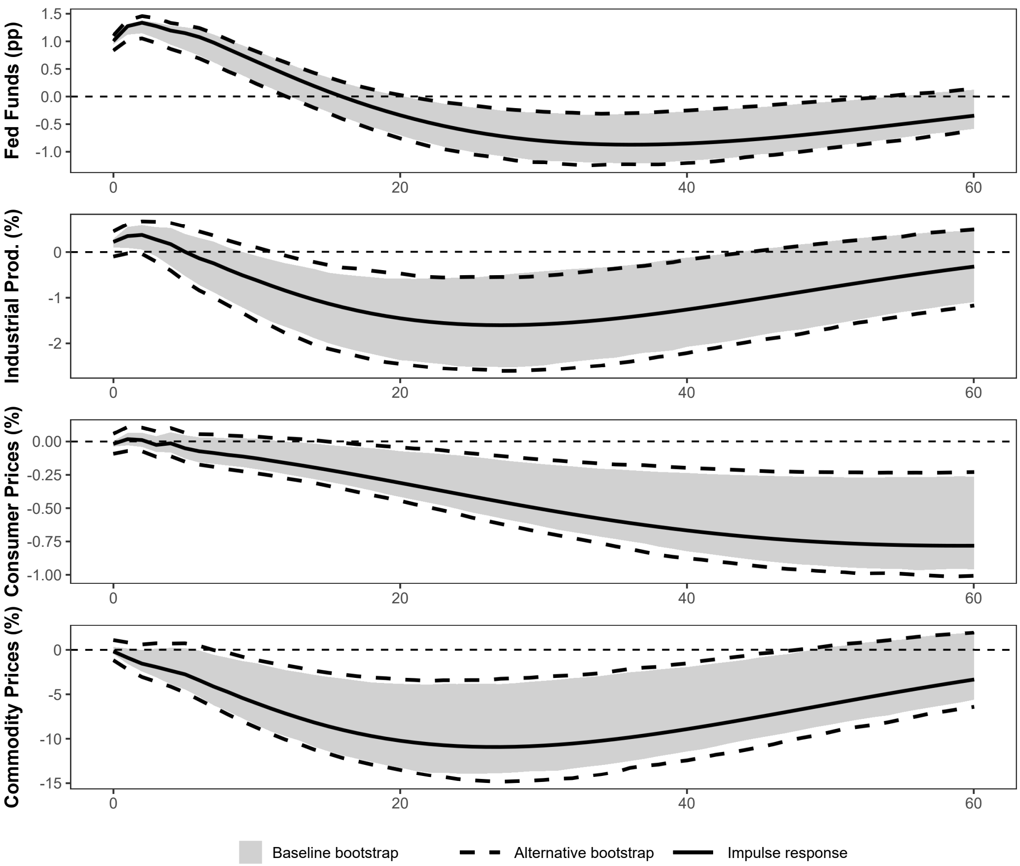


Notes: The figure shows the impulse responses to a monetary policy shock of 100 basis points in state $m = 1$ of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dashed lines denote the 95% pointwise confidence intervals based on 1,000 bootstrap replications of a VAR-residual moving block bootstrap. Bootstrapped samples are constructed by sampling blocks of the estimated residuals $\hat{e}_t$ of model (8) of the main paper. Using block length of $l = 50$ defines $n = T/l$ as the number of nonoverlapping blocks, where $ln \geq T$. The blocks of length l of the estimated residuals $\hat{e}_t$ are arranged in the form of the matrix

$$\begin{bmatrix} \widehat{e}_1 & \widehat{e}_2 & \dots & \widehat{e}_l \\ \widehat{e}_2 & \widehat{e}_3 & \dots & \widehat{e}_{1+l} \\ \vdots & \vdots & & \vdots \\ \widehat{e}_{T-l+1} & \widehat{e}_{T-l+2} & \dots & \widehat{e}_T \end{bmatrix}.$$

The bootstrap residuals are recentered by removing the columnwise mean to ensure that the bootstrap residuals have mean zero, i.e., $\widetilde{e}_{jl+i} = \widehat{e}_{jl+i} - \frac{1}{T-l+1}\sum_{r=0}^{T-l}\widehat{e}_{i+r}$ for $i = 1, 2, \dots, l$ and $j = 0, 1, \dots, n-1$. Bootstrap residuals are generated by drawing n times with replacement from the recentered rows of the matrix. These draws are combined in a time series of bootstrap residuals, $[e_1^*, \dots, e_T^*]$, by joining them end-to-end and retaining the first T bootstrap residuals. Each bootstrap samples starts with identical pre-sample values from the original data set as initial values, i.e., $z_{-p+1}^* = z_{-p+1}, \dots, z_0^* = z_0$ and is then generated recursively for $t = 1, \dots, T$ as $z_t^* = \hat{\delta} + \hat{\Gamma}(L)z_{t-1} + \hat{e}_t$. In this bootstrap, also the relative variances and transition probabilities are re-estimated in each bootstrap repetition. The ordering of the relative variances and, hence, of columns of D in each bootstrap repetition is done identically to the procedure described in the main paper.

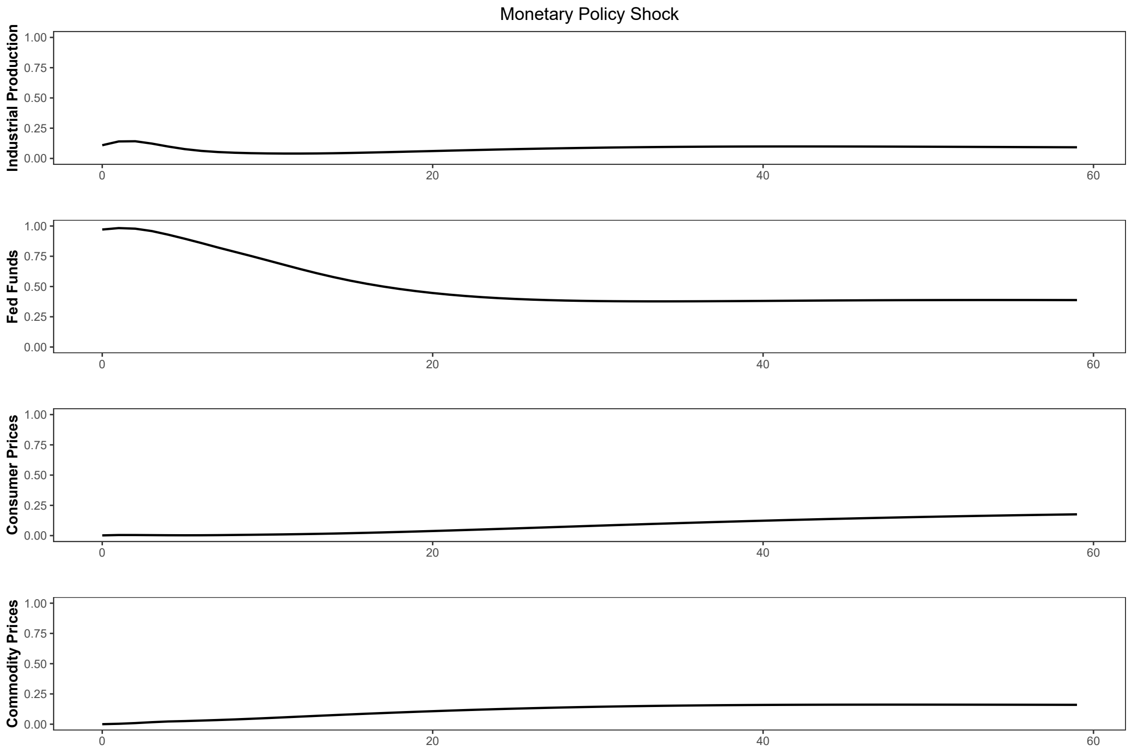
Figure A.10: Sensitivity analysis of baseline model using an recursive design wild bootstraps with draws from standard normal distribution.



Notes: The figure shows the impulse responses to a monetary policy shock of 100 basis points in state $m = 1$ of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dataset is $z_t = [ff_t, ip_t, pce_t, prm_t, rr_t]'$. The shaded bands denote 95% pointwise confidence intervals based on 1,000 bootstrap replications of the baseline heteroskedastic proxy-SVAR(6) model with $M = 2$ states. The dashed lines denote the 95% pointwise confidence intervals based on 5,000 bootstrap replications of an alternative recursive design wild bootstrap. Bootstrapped samples are contructed as $z_t^* = \hat{\delta} + \hat{\Gamma}(L)z_{t-1} + \tilde{\varphi}_t\hat{e}_t$, where $\hat{e}_t$ are the estimated residuals, $\hat{\delta}$ and $\hat{\Gamma}(L)$ are estimated counterparts of the coefficients the model (8) of the main paper, and $\tilde{\varphi}_t$ is an independent random variable drawn from a standard normal distribution, i.e., $\tilde{\varphi} \sim N(0, 1)$. Each of the 1,000 generated bootstrap samples is based on identical pre-sample values from the original data set as initial values, i.e., $z_{-p+1}^* = z_{-p+1}, \ldots, z_0^* = z_0$. The bootstrap is conducted conditionally on estimated parameters for the relative variances and transition probabilities.

A.4.4 FEVD for standard proxy-SVAR

Figure A.11: Forecast error variance decomposition for standard proxy-SVAR



References

Bouakez, H. and Normandin, M. (2010). Fluctuations in the foreign exchange market: How important are monetary policy shocks?, *Journal of International Economics* **81**(1): 139 – 153.

Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**(1): 104–116.

Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**(2): 121–131.

Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**: 61–75.

Lütkepohl, H. and Netšunajev, A. (2014). Structural vector autoregressions with smooth transition in variances: The interaction between U.S. monetary policy and the stock market, *Discussion Paper 1388*, DIW Berlin.

Lütkepohl, H. and Netšunajev, A. (2017). Structural vector autoregressions with heteroskedasticy: A review of different volatility models, *Econometrics and Statistics* **1**: 2–18.

Lütkepohl, H. and Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Oxford Bulletin of Economics and Statistics* **80**(4): 715–735.

Milunovich, G. and Yang, M. (2013). On Identifying Structural VAR Models via ARCH Effects, *Journal of Time Series Econometrics* **5**(2): 117–131.

Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**(6): 1217–1243.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications, *American Economic Review* **94**(4): 1055–1084.

Sentana, E. and Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics* **102**(2): 143 – 164.

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal garch model, *Journal of Applied Econometrics* **17**(5): 549–564.