

Herwartz, Helmut; Rohloff, Hannes; Wang, Shu

Working Paper

Proxy SVAR identification of monetary policy shocks: MonteCarlo evidence and insights for the US

cege Discussion Papers, No. 404

Provided in Cooperation with:

Georg August University of Göttingen, Department of Economics

Suggested Citation: Herwartz, Helmut; Rohloff, Hannes; Wang, Shu (2020) : Proxy SVAR identification of monetary policy shocks: MonteCarlo evidence and insights for the US, cege Discussion Papers, No. 404, University of Göttingen, Center for European, Governance and Economic Development Research (cege), Göttingen

This Version is available at:

<https://hdl.handle.net/10419/225264>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**PROXY SVAR IDENTIFICATION OF
MONETARY POLICY SHOCKS - MONTE
CARLO EVIDENCE AND INSIGHTS FOR
THE US**

Helmut Herwartz
Hannes Rohloff
Shu Wang

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Proxy SVAR identification of monetary policy shocks - Monte Carlo evidence and insights for the US

Helmut Herwartz*

Hannes Rohloff*

Shu Wang*

October 6, 2020

Abstract

In empirical macroeconomics, proxy structural vector autoregressive models (SVARs) have become a prominent path towards detecting monetary policy (MP) shocks. However, in practice, the merits of proxy SVARs depend on the relevance and exogeneity of the instrumental information employed. Our Monte Carlo analysis sheds light on the performance of proxy SVARs under realistic scenarios of low relative signal strength attached to MP shocks and alternative assumptions on instrument accuracy. In an empirical application with US data we argue in favor of the specific informational content of instruments based on the dynamic stochastic general equilibrium model of Smets and Wouters (2007). A joint assessment of the benchmark proxy SVAR and the outcomes of a structural covariance change model imply that from 1973 until 1979 monetary policy contributed on average between 2.2 and 2.4 units of inflation in the GDP deflator. For the so-called Volcker disinflation starting in 1979Q4, the benchmark structural model shows that the Fed's policy measures effectively reduced the GDP deflator within three years (i.e. by -3.06 units until 1982Q3). While the empirical analysis largely conditions on a small-dimensional trinity SVAR, the benchmark proxy SVAR shocks remain remarkably robust within a six-dimensional factor-augmented model comprising rich information from Michael McCracken's database (FRED-QD).

Keywords: Structural vector autoregression, external instruments, proxy SVAR, heteroskedasticity, monetary policy shocks.

JEL Classification: C15, C32, C36, E47

*Chair of Econometrics, University of Goettingen, Humboldtallee 3, D-37073 Goettingen, Germany. Corresponding author: Shu Wang, e-mail: shu.wang@uni-goettingen.de. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (HE 2188/8-2). We thank Lutz Kilian for commenting on a previous version of this manuscript.

1 Introduction

Various theory- or data-based approaches have been developed to recover the structural form of vector autoregressive models (VARs, SVARs), which extend beyond the traditional imposition of short- (Sims 1980) or long-run restrictions (Blanchard and Quah 1989).¹ While theory-based identification through (agnostic) sign restrictions (Faust 1998, Canova and De Nicolo 2002, Uhlig 2005) guarantees economic content of the identified shocks, such approaches leave no room for the data to object against the structural model, and they risk intermingling model assumptions and conclusions. Unlike theory-based identification, data-based identification techniques (i.e. identification through heteroskedasticity or independent component analysis (ICA); see, e.g. Rigobon 2003, Lanne and Lütkepohl 2008, Lanne et al. 2017, Gouriéroux et al. 2017) offer diagnostic tools for testing the otherwise just identifying structural model specifications. However, data-based identification as such is silent about the economic interpretation of the resulting structural shocks. Proxy SVARs (Stock and Watson 2012, Mertens and Ravn 2013) develop from an eventual prospect of measuring a shock of interest directly by means of external information. As long as such instruments are relevant and valid, proxy SVARs promise to unite the merits of theory- and data-based identification schemes. On the one hand, the construction of the proxy variable requires economic foundation and theoretical assumptions regarding the orthogonality to other structural shocks. On the other hand, the relevance of the instrument can be tested and the estimation of the structural model proceeds under relatively weak assumptions. For instance, regarding monetary policy (MP) analysis, the ‘high-frequency identification’ scheme of Gertler and Karadi (2015), or narrative evidence as in Romer and Romer (2004) could deliver fruitful external evidence. It currently seems that proxy SVARs have become the first-choice identification approach for MP analysis (Christiano et al. 1999, Ramey 2016).² Unsurprisingly, this has led to a race for most reliable instruments, whereby today a plethora of potential MP shock series are available to researchers (Romer and Romer 2004, Gürkaynak et al. 2005, Sims and Zha 2006, Smets and Wouters 2007, Barakchian and Crowe 2013, Gertler and Karadi 2015, Caldara and Herbst 2019). Against this background, it holds utmost importance for applied macroeconomists to have a sound understanding of both the relative performance of specific instruments as well as the merits and perils of proxy SVARs within a space of alternative or complementary theory- and data-based identification schemes. Following such

¹For a comprehensive and up-to-date collection of identification strategies in SVARs, the reader may consult Kilian and Lütkepohl (2017).

²The appeal of instrument-based identification methods for MP analysis might derive from the coincidence of two aspects. First, there is ample availability of candidates for shock series, based on the very nature of MP conduct, i.e. narrative evidence obtained from Federal Open Market Committee (FOMC) meetings (Romer and Romer 2004), and high-frequency series based on surprise movements surrounding FOMC meetings in federal funds futures (Gürkaynak et al. 2005, Gertler and Karadi 2015). Second, one might perceive a particularly pronounced shortcoming of both theory- and data-based identification schemes, which often yield results that are at odds with basic consensual transmission characteristics (Uhlig 2005, Carlstrom et al. 2009), such as the occurrence of ‘price’ or ‘output puzzles’ (see Rusnak et al. (2013) for meta evidence on price puzzles).

a research agenda, Mumtaz et al. (2018) provide a role study for the identification of credit supply shocks that combines Monte Carlo evidence from several dynamic stochastic general equilibrium (DSGE) models and delivers an empirical application to US data. In particular, the authors find that impulse responses obtained from proxy SVARs best match those from underlying DSGE models, while identification based on heteroskedasticity performs satisfactorily under sufficiently strong variance shifts.

The purposes of this study are threefold. Providing a natural complement to Mumtaz et al. (2018), our first aim is to analyse the relative performance of proxy SVARs in MP analysis in two important directions that are frequently encountered in practice but only marginally addressed or neglected in Mumtaz et al. (2018). On the one hand, we provide a perspective on proxy SVAR identification performance within an extended variety of identification schemes (sign restrictions, identification through heteroskedasticity or ICA). As such, we intend to address the relative performance of proxy SVARs under specifications of data generation that (i) are in line with a stylized three-variable (output, inflation, interest rates) DSGE model for monetary analysis (Carlstrom et al. 2009, Castelnuovo 2012) and (ii) feature empirical properties that have attracted the interest of several studies (e.g. breaks in shock volatilities (Sims and Zha 2006, Lanne and Lütkepohl 2008), weak contributions of MP shocks to overall variance (Paustian 2007, Wolf 2020), non-Gaussian shocks (Gouriéroux et al. 2017)).³ On the other hand, we acknowledge that the core condition of instrument exogeneity rarely holds in practice in a strict sense.⁴ Accordingly, we investigate possible consequences (i.e. misidentifications or puzzles) under realistic scenarios of weak instrumentation or distinct levels of endogenous shock contamination. As a second purpose - and similar to Miranda-Agrippino and Rey (2018) (for high-frequency modeling) or Stock and Watson (2012) - this study aims to conduct a comparative analysis of instrumental information that has been suggested for MP analysis. In this regard - and extending beyond the reference studies - we evaluate a variety of available and relevant MP instruments in two directions. On the one hand, this comparison covers several model characteristics as, e.g. economic labels of identified shocks, correlations among proxy-identified MP shocks and non-MP shocks retrieved from theory- or data-based identification, tests on exogeneity à la Podstawski et al. (2018) or correlations among identified MP shocks. On the other hand, we evaluate the extent to which alternative MP shocks are in line with prominent narratives put forth by Antolín-Díaz and Rubio-Ramírez (2018). In this context, we also distinguish the scope of the empirical SVARs to soundly characterize the structural contributions of MP to the *Great Inflation* of the 1970s and the process of disinflation initiated after the appointment of Paul Volcker as

³We are aware of a hand full of related Monte Carlo investigations of identification schemes, albeit (almost) all of which lack coverage of proxy SVARs, namely e.g. Castelnuovo (2012), Carriero et al. (2015), Herwartz and Plödt (2016), Lütkepohl and Schlaak (2018), Podstawski et al. (2018) and Herwartz et al. (2019). See Section 3 for more details.

⁴For instance, Stock and Watson (2012) show that none of their eighteen instruments considered accords with the exogeneity assumption. In Mumtaz et al. (2018), one out of two relevant external credit shock series is subject to contamination with other (non-credit) shocks.

the new Chairman of the Fed in August 1979. A third purpose of this study emanates from the observation that direct measures of shocks of interest often lack strict exogeneity. While unobserved adjustments among variables of interest, omitted variables or instrumental information are a pertinent threat for exogeneity irrespective of time resolution, the detection and procurement of instrument contamination holds particular concern if empirical models build upon low-frequency information (e.g. quarterly data). On the one hand several directions of MP analysis are limited to low-frequency models (MP and credit, MP and inequality). On the other hand, low-frequency data might be subject to latent patterns of temporal aggregation. Against this background - and similar to Bridge Proxy SVARs (Gazzani and Vicondoa 2019) - the third purpose of this study is to combine insights from proxy SVARs and other identification schemes. In this regard, our analysis aims to provide both a diagnostic step towards unraveling endogenous contaminations of MP instruments and an improvement of relevant MP instruments in terms of ‘model-implied’ exogeneity.

Our Monte Carlo exercises show that data-driven identification techniques promise a most satisfying performance overall. Proxy SVARs are capable of detecting the true impulse responses with satisfactory accuracy under reasonable signal-to-noise ratios. Moreover, we show that the related results of Paustian (2007) and Wolf (2020) carry over to the case of proxy SVARs conditioning on endogenous instruments, i.e. identifying shocks with relatively low signal strength suffers strongly from contaminated external information, and could - for instance - result in diagnosing marked price puzzles in MP analysis. As it turns out, both aspects (endogenous instrument contamination and low signal strength) are pertinent for the identification of MP surprises in the US. Revisiting the informational content of several instruments that have been put forth for the analysis of the Fed’s MP, we obtain a set of three instruments that hold sufficient relevance to identify quarterly MP shocks. The discussion of instrument exogeneity extends beyond formal testing, and benefits from complementing proxy SVARs with data-based identification and narrative evidence. Conditional on a trivariate macro-model of output (gap), inflation and the one-year government bond rate, the innovations to the estimated MP reaction function in the medium-scale DSGE model of Smets and Wouters (2007) are (i) most suitable for the extraction of MP shocks by means of proxy SVARs, and (ii) closely related to MP shocks that are derived in a data-based manner from the supposition of a (co)variance decline that coincides with the *Great Moderation*. Alternative low-frequency instrumental series (i.e. the Romer and Romer (2004) narrative MP shock series (RR), and the best-fitting SVAR shock from Sims and Zha (2006) (SZ)) give rise to sizable price puzzles and fail to explain the important disinflation period starting in late 1979. From an explicit test, we find no evidence that calls into question the fundamentalness of identified shocks. Moreover, the benchmark proxy MP shocks remain remarkably robust in information-rich models, i.e. in VARs augmented with (up to three) factors summarizing more than 200 time series variables drawn from the database of Michael McCracken (FRED-QD). A joint assessment of the benchmark proxy SVAR and the outcomes of a structural covariance change model imply that MP was a persistent source of high inflation in the 1970s.

Specifically, monetary policy contributed on average between 2.2 and 2.4 units of inflation in the GDP deflator from 1973 to 1979. For the Volcker disinflation starting in 1979Q4, the benchmark structural model shows that the Fed’s policy measures effectively reduced the GDP deflator within three years (i.e. by -3.06 units until 1982Q3). More prevalent inflationary pressures originated on the supply (marginal deflator increase of +3.56 units) rather than the demand side (+1.18).

Section 2 provides a brief outline of the SVAR model and the proxy SVARs as an attempt to solve the identification issue. Section 3 comprises an in-depth Monte Carlo investigation, while Section 4 discusses the identification of MP shocks and provides a critical assessment of the informational content of established instrumental variables. Section 5 complements the structural analysis with narrative analysis largely focusing on the Volcker disinflation, before Section 6 concludes. The appendices attached to this work collect (i) further details on theory- and data-based identification schemes (Appendix A), (ii) a sketch of the log-linearization of the trinity DSGE model that we use for data generation in Monte Carlo experiments (Appendix B), (iii) simulation results beyond those discussed in the main text (Appendix C), (iv) detailed information about external MP shock series that we collected for empirical analysis (Appendix D), and (v) a discussion of informational frictions that are specific to MP analysis (or forward guidance) and might give rise to instrument contamination (Appendix E).

2 VAR model and proxy SVARs

This section sketches the identification problem in SVARs, and subsequently outlines proxy SVAR identification.

2.1 Identification problem

Let y_t be a $K \times 1$ vector of observable variables. The VAR model of order p has the reduced and structural form, respectively:

$$y_t = \nu_t + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (1)$$

$$= \nu_t + A_1 y_{t-1} + \dots + A_p y_{t-p} + B \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where $A_j, j = 1, 2, \dots, p$ are $K \times K$ coefficient matrices, and ν_t is a $K \times 1$ vector of non-stochastic terms (intercept and deterministic trends). The residual vector u_t in the reduced-form (1) is free from serial correlation with a zero mean and positive definite (non-diagonal) covariance matrix Σ_u . The structural shocks ε_t in equation (2) have a diagonal covariance matrix Σ_ε , and without loss of generality, the structural shocks are often normalized to have unit variance, such that $u_t \sim (0, \Sigma_u)$ and $\varepsilon_t \sim (0, I_K)$. For simplicity, we assume that the process is causal and $\det A(z) = \det \left(I_K - \sum_{j=1}^p A_j z^j \right) \neq 0$ for $|z| \leq 1$. This ensures that

the process has a Wold moving average (MA) representation, i.e.

$$\begin{aligned} y_t &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i B \varepsilon_{t-i} = \mu + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i}, \end{aligned} \quad (3)$$

where $\mu = A(1)^{-1}\nu$, $\Phi_0 = I_K$ and $\Phi_i = \sum_{j=1}^i A_j \Phi_{i-j}$ with $A_j = 0$ for $j > p$. The MA representation in (3) holds particular importance for structural impulse response functions (IRFs) and forecast error variance decompositions. While reduced-form parameters (A_j , $j = 1, \dots, p$, and thus Φ_i , $i = 0, 1, \dots$) and the covariance matrix of residuals (Σ_u) can be estimated consistently, the structural parameters collected in the non-singular matrix B are not identified without further theoretical or statistical assumptions. Consequently, the structural MA coefficients cannot be recovered, since they require knowledge of B , i.e. $\Theta_i = \Phi_i B$. One possible solution is to restrict matrix B as the lower-triangular Choleski factor imposing a recursive causal structure among the model variables. This approach has a tradition in the identification of MP shocks, where the policy indicator is ordered at or near the bottom of the system (Bernanke and Blinder 1992). However, it has been criticized that the Choleski scheme is inconsistent with a large class of DSGE models, which typically feature contemporaneous responses of macroeconomic variables to the policy shock (Smets and Wouters 2007, Castelnuovo 2012).

Apart from imposing zero restrictions, it has been argued that theoretical patterns of effect directions or statistical properties of the data (non-Gaussianity and heteroskedasticity) provide fruitful identifying information (see Appendix A for brief summaries of theory- and data-based identification schemes). In the following, we briefly outline the proxy SVARs identification (Stock and Watson 2012, Mertens and Ravn 2013) which relies on external information rather than the imposition of restrictions on the structural relationships among the variables included in the VAR.

2.2 Identification by external instruments

As an alternative to exclusion restrictions - which have an interpretation of ‘internal’ instrumentation (Stock and Watson 2012) - one can consider a measure of a structural shock of interest based on information from outside the VAR, and treat this measure as an external instrument for identification (for a recent review, see Stock and Watson 2018). While the use of external instruments has a fruitful history in microeconomic research, its adaptation to SVAR identification has been put forth more recently by Stock and Watson (2012) and Mertens and Ravn (2013).⁵ Proxy SVARs could be considered to reconcile the merits

⁵In principle, it seems tempting to augment the VAR model by including such an instrument as an endogenous variable. Subsequently, identification can be achieved by imposing a recursive scheme with corresponding variable ordering (e.g. Kilian 2009, Auerbach and Gorodnichenko 2012). For simulation-based evidence pointing to a performance lead of proxy SVARs over augmented (S)VARs, the reader may consult Mumtaz et al. (2018).

of theory- and data-based identification techniques. While the choice of specific external information is typically guided by theoretical considerations or institutional knowledge, identification is based on empirical correlations and some of the underlying assumptions can be tested, similar to over-identifying assumptions within statistical identification schemes.

Stock and Watson (2012) have pioneered proxy SVARs to investigate macroeconomic dynamics. Specifically, they use eighteen instruments to identify six structural shocks within a factor-augmented VAR framework (oil, MP, productivity, uncertainty, liquidity and financial risk, and fiscal policy) that have received considerable attention in the discussion of the *Great Recession* and the subsequent recovery. Mertens and Ravn (2013) utilize narratively-identified exogenous shocks to average tax rates as instruments, and analyze the dynamic effects of structural tax shocks on the US economy. Employing a mortgage rate-based instrument, Mian et al. (2017) investigate the distinct role of household credit supply shocks on output growth compared with corporate credit supply shocks. Nonetheless, MP analysis remains the main field of application of proxy SVARs. Gertler and Karadi (2015) have constructed several instruments for MP shocks by means of high-frequency surprises in federal funds futures around FOMC announcements to unravel the role of a credit channel for the transmission of MP shocks (including forward guidance). Miranda-Agrippino and Ricco (2017) have further advanced the high-frequency identification approach by accounting for anticipatory effects that are present e.g. in the Gertler and Karadi (2015) instruments.

For a formal exposition of proxy SVAR identification, let z_t denote an external instrument for a structural shock of interest ε_{kt} , $k \in \{1, \dots, K\}$. In analogy to the common assumptions in the instrumental variable estimation literature, the instrument must satisfy the following conditions:

- (i) z_t is *relevant* for the underlying structural shock ε_{kt}

$$\mathbb{E}(\varepsilon_{kt}z_t) = \phi \neq 0, \quad (4)$$

- (ii) z_t is *exogenous* from other structural shocks in the system

$$\mathbb{E}(\varepsilon_{lt}z_t) = 0, \quad \forall l \in \{1, \dots, K\} \setminus \{k\}. \quad (5)$$

From both conditions in (4) and (5) it follows that up to scale ϕ , the population covariance between the instrument and VAR residuals obtains the k -th column of B , denoted by B_k , i.e.

$$\mathbb{E}(u_t z_t) = B_k \mathbb{E}(\varepsilon_{kt} z_t) = \phi B_k. \quad (6)$$

Let Π denote the $1 \times K$ coefficient vector from the regression of z_t on the residual vector u_t . From (4), (5) and (6), along with the unit-variance normalization of the structural shocks, the predictor for z_t conditional on u_t gives the shock ε_{kt} up to scale ϕ :

$$\mathbb{E}(z_t | u_t) = \Pi u_t = \mathbb{E}(z_t u_t') \Sigma_u^{-1} u_t = \phi B_k' [B B']^{-1} u_t = \phi e_k' \varepsilon_t = \phi \varepsilon_{kt}, \quad (7)$$

where e_k is the k -th column of I_K . Combining (7) and (4), the squared covariance between z_t and ε_{kt} can be recovered as

$$\Pi\mathbb{E}(u_t z_t) = \phi \left[B^{-1} B_k \right]' B^{-1} B_k \phi = \phi^2. \quad (8)$$

Dividing (6) by the square root of (8) produces a unique (up to sign) solution for parameter vector B_k . Lunsford (2015) discusses the estimation of moments in (6) and (8) and the asymptotic properties of these estimators.

Observing the potential prevalence of weak instrumentation, Lunsford (2015) analyses the asymptotic bias caused by weakly relevant proxies, and suggests a test of the null hypothesis of zero coefficients Π in the regression of instruments on VAR residuals (see (8)).⁶ Regarding inference about structural IRFs, Mertens and Ravn (2013) and Gertler and Karadi (2015) use a recursive wild bootstrap, where the residuals and instruments are resampled by means of the ‘Rademacher’ distribution. However, Jentsch and Lunsford (2019) show that (Rademacher) wild bootstrap schemes are generally inconsistent for inference on the structural impulse responses. As a consistent alternative, they propose a modified moving block bootstrap (MBB) procedure that is capable of replicating the joint fourth-order moments of the VAR residuals and the instruments.

3 Monte Carlo study

The simulation experiments documented in this section shed light on the performance of proxy SVARs in comparison with alternative identification schemes. In the following, we first encounter some related Monte Carlo investigations of theory- or data-based identification approaches. Subsequently, we describe our simulation design, i.e. the data-generating process (DGP), stochastic characteristics of the structural shocks and the criteria to assess the identification outcomes. We are also explicit about the implementation of theory- and data-based identification techniques. For further details on these identification methods, we refer the reader to Appendix B. Subsequent to the discussion of simulation results, finally, we provide a stylized theoretical note on a pitfall of proxy SVAR identification that evolves if (i) an instrument under scrutiny lacks strict exogeneity, and (ii) the shock of interest suffers from relatively weak signal strength.

3.1 Monte Carlo-based evidence on identification schemes

Several related Monte Carlo investigations have analyzed the performance characteristics of recent theory- and data-based identification techniques, and they provide guidance for the matching of empirical VARs and suitable identification schemes. Based on a stylized three-equation DSGE model, Castelnovo (2012) investigates the performance of imposing agnostic sign restrictions (Uhlig 2005), finding ambiguous effects of a monetary tightening on output with data simulated from structural models that are at odds with monetary neutrality.

⁶Montiel Olea et al. (2018) discuss weak instrument robust inference for proxy SVARs. See also Cesa-Bianchi et al. (2015) for a multiple instrument approach to alleviate weak instrumentation issues.

Herwartz and Plödt (2016) find that the exploitation of volatility breaks promises a lead over identification by means of sign restrictions. Based on bivariate DGPs, Lütkepohl and Schlaak (2018) exam alternative time-varying (co)variance change (henceforth denoted CV) models and discuss model selection in this context. The large-scale simulation study of Herwartz et al. (2019) provides evidence in favor of identification by means of ICA. Specifically, as suggested in Matteson and Tsay (2017), the minimization of the distance covariance statistic proves to obtain robustly useful identification outcomes under various DGPs. Simulation-based evidence on the relative performance of proxy SVARs is scant. Podstawski et al. (2018) show that - in a mean squared error sense - proxy SVARs could lead to serious distortions of structural impulse response estimates if one does not properly take account of heteroskedasticity.

Tools of data-based identification and proxy SVARs are typically conditional on underlying assumptions and hence prone to statistical nuisance under model misspecification and eventual finite sample biases. Therefore, a thorough understanding of the robustness of structural analysis under realistic properties of real-life data is in the core interest of empirical analysts. Providing particular motivation to the Monte Carlo study at hand, the subsequent simulation experiments highlight the scope of alternative identification schemes in ascertaining (unique) MP shocks and their impact effects under practically realistic notions of (i) misspecification of the empirical model regarding core identifying assumptions, (ii) low signal-to-noise ratios or invalid instrumental information, and (iii) low relative signal strength of a specific shock of interest.

3.2 Data-generating processes

Time series data are drawn from a variety of statistical models. Henceforth, we provide in turn (i) a causal VAR representation of a trinity DSGE model, (ii) alternative covariance regimes for sampling the structural shocks, (iii) and alternative distributions of the scale-free components of the structural shocks.

3.2.1 Log-linearized DSGE model in VAR form

We employ a stylized 3-equation DSGE model comprising output gap (x_t), inflation (π_t) and the nominal interest rate (r_t). The so-called trinity model has been widely used for understanding the role of MP for the evolution of other major economic aggregates (Carlstrom et al. 2009, Castelnovo 2012, Herwartz and Plödt 2016). After log-linearization the structural model reads in VAR form as

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + B \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (9)$$

where $y_t = (x_t, \pi_t, r_t)'$. As implied by typical calibration assumptions, the B-matrix is:⁷

$$B = \begin{bmatrix} 2.32 & -0.48 & -0.41 \\ 0.72 & 2.32 & -0.22 \\ 0.98 & 1.57 & 0.76 \end{bmatrix}. \quad (10)$$

⁷For further details on the calibration of the underlying DSGE model, see Appendix B.

Given that the first shock (ε_1) has on impact a positive effect on all three variables, it can be interpreted as a demand shock. The second shock (ε_2) raises the price level and the interest rate but invokes a negative response of output, and hence it can be labeled as a supply shock. The third shock (ε_3) reflects a contractionary MP action, which raises the interest rate, fights inflation and dampens economic activities.

3.2.2 Covariance regimes

The structural shocks in (9) are written as:

$$\varepsilon_t = \Lambda_t \varsigma_t, \quad \varsigma_{kt} \stackrel{iid}{\sim} \psi^{(\bullet)}, \quad k \in \{1, \dots, K\}, \quad (11)$$

where $\psi^{(\bullet)}$ is a (centered and standardized) parametric distribution such that $\mathbb{E}(\varsigma_t) = \mathbf{0}$ and $\mathbb{E}(\varsigma_t \varsigma_t') = I_K$. Moreover, $\Lambda_t = \text{diag}(\lambda_{1t}, \lambda_{2t}, \dots, \lambda_{Kt})$ is a K -dimensional time-dependent diagonal matrix formalizing distinct variance scenarios. We consider two variance breaks occurring at time points $T_{B_1} = \frac{1}{3}T$ and $T_{B_2} = \frac{2}{3}T$. By construction, ε_t has the covariance structure $\mathbb{E}(\varepsilon_t \varepsilon_t') = \text{diag}(\lambda_{1t}^2, \lambda_{2t}^2, \dots, \lambda_{Kt}^2)$, where

$$\Lambda_t = \begin{cases} \Lambda^{(1)} = I_K & \text{for } t = 1, \dots, T_{B_1} - 1 \\ \Lambda^{(2)} = \text{diag}(1.2, 1.1, 1) & \text{for } t = T_{B_1}, \dots, T_{B_2} - 1 \\ \Lambda^{(3)} = \text{diag}(3.6, 1.1, 0.5) & \text{for } t = T_{B_2}, \dots, T. \end{cases} \quad (12)$$

Hence, the shocks are normalized to have unit variance in the pre-break period. The second volatility break is much stronger than the first one in terms of both magnitude and direction. With this covariance design, we try to mimic a scenario where the data exhibit two breaks while the econometrician applies identification based only on a single break-point occurring at T_{B_2} , since she might find stronger statistical evidence of structural change conditional on this time instance.⁸ It is worth noting that the MP shock - in particular after the second break-point - contributes least to the variation of the system. This is consistent with empirical evidence about the role of monetary interventions in the post-World War II US economy (Faust 1998, Uhlig 2005, Smets and Wouters 2007). Hence, the particular covariance design may help to evaluate the performance of alternative identification schemes in a scenario where the signal strengths from the underlying shocks considerably differ.

3.2.3 Distributional frameworks

Innovations in SVAR models may well deviate either symmetrically or asymmetrically from Gaussianity. In order to analyze the sensitivity of the identification performance with respect to distributional characteristics of the data, the shocks ς_{kt} in (11) are independently and identically drawn from the following distributions:

⁸For example, Lanne and Lütkepohl (2008) analyze monthly US data for the 1965M1-1996M12 period. They examine two break-points that have been considered in the literature: 1979M10 and 1984M2 (Bernanke and Mihov 1998, Christiano et al. 1999, Perez-Quiros and McConnell 2000, Bacchiocchi and Fanelli 2015). Based on LR tests they find much stronger significance for the second volatility break-point 1984M2 in comparison with 1979M10.

- (i) $\psi^{(G)} \sim N(0, 1)$,
- (ii) $\psi^{(\text{Chi})} \sim \left(\chi_{(v)}^2 - v\right) \frac{1}{\sqrt{2v}}$,
- (iii) $\psi^{(t)} \sim t_{(v)} \frac{\sqrt{v-2}}{\sqrt{v}}$.

With this distribution design, we imitate non-Gaussianity of the data in terms of second- (shift in covariances), third- (skewness) and fourth-order moments (leptokurtosis). Overall, five scenarios are considered and summarized in Table 1, where \mathcal{L}_I is taken as the baseline scenario. We set $v = 5$, or alternatively $v = 10$, to distinguish between stronger or weaker deviations from multivariate Gaussianity (\mathcal{L}_{II}).

	\mathcal{L}_I	$\mathcal{L}_{II}^{(a)}$	$\mathcal{L}_{II}^{(b)}$	$\mathcal{L}_{III}^{(a)}$	$\mathcal{L}_{III}^{(b)}$
ν	-	10	5	5	5
ς_{1t}	$\psi^{(G)}$	$\psi^{(t)}$	$\psi^{(t)}$	$\psi^{(t)}$	$\psi^{(G)}$
ς_{2t}	$\psi^{(G)}$	$\psi^{(t)}$	$\psi^{(t)}$	$\psi^{(t)}$	$\psi^{(t)}$
ς_{3t}	$\psi^{(G)}$	$\psi^{(t)}$	$\psi^{(t)}$	$\psi^{(\text{Chi})}$	$\psi^{(\text{Chi})}$

Table 1: Overview of distributional settings for the simulations.

Similar to Herwartz (2018), we also simulate systems, in which structural innovations are drawn from heterogeneous families of distributions (\mathcal{L}_{III}) to investigate possible consequences of the presence of skewed ($\mathcal{L}_{III}^{(a)}$) or Gaussian ($\mathcal{L}_{III}^{(b)}$) components in the mixture.

3.3 Instrumental variables

Proxy SVAR identification relies on external instruments. For a sound understanding of proxy SVARs, it is worth noting that its performance could face similar risks as the gathering of external information in GMM estimation. Taking into account uncertainty about instrument selection, it is interesting to evaluate the performance of proxy SVARs for the case that the instrumental relevance is weak or the exogeneity condition (5) is violated. Therefore, our Monte Carlo analysis covers scenarios where instrumental information is (i) relevant and exogenous, (ii) weak or (iii) correlated with other structural shocks (i.e. endogenous). Conditional on the samples of the structural shocks outlined in (11), we simulate three distinct instruments for each shock accordingly:

$$(i) \quad z_{kt}^{(1)} = \varepsilon_{kt} + \xi_{kt}, \quad (13)$$

$$(ii) \quad z_{kt}^{(2)} = \frac{\varepsilon_{kt}}{\sqrt{T}} + \xi_{kt}, \quad (14)$$

$$(iii) \quad z_{kt}^{(3)} = \gamma_k \varepsilon_{kt} + \sum_{l \neq k} \delta_l \varepsilon_{lt} + \xi_{kt}, \quad \delta_l < \gamma_k \quad \forall l, k \in \{1, \dots, K\}. \quad (15)$$

Given that external instruments $z_{kt}^{(m)}$, $m = 1, 2, 3$, are considered as proxies for the shock ε_{kt} , it is reasonable to assume some degree of noise. This is captured by a Gaussian white noise measurement error ξ_{kt} with zero mean and variance $\text{Var}[\xi_{kt}] = \text{Var}[\varepsilon_{kt}] = \lambda_{kt}^2$. Accordingly, the respective signal-to-noise ratio is time-invariant. While the scenario in (13) formalizes

the ideal case of proxy SVAR analysis, such a situation might hold limited representation in empirical macroeconomics. On the one hand weak instruments are frequently encountered in practice. For instance, of the five instruments considered in Mumtaz et al. (2018), only two prove to be strong proxies. Our simulations cover scenarios of weak instrumentation (14) with a local-to-zero assumption (Lunsford 2015). Specifically, the covariance between the instrument and its underlying structural shock vanishes with the sample size, i.e.

$$\mathbb{E}(u_t z_{kt}^{(2)}) = B_k \mathbb{E}(\varepsilon_{kt} z_{kt}^{(2)}) = \frac{1}{\sqrt{T}} B_k. \quad (16)$$

If the instruments for distinct shocks are valid, proxy SVARs should provide uncorrelated shocks. However, in practice, shocks identified by instruments from specific categories are often contemporaneously correlated with each other, at least to some degree. For instance, Stock and Watson (2012) find that the MP shocks identified by various instruments strongly correlate with fiscal policy shocks or oil shocks. We generate instruments that contemporaneously correlate with multiple shocks in the system (15). Nevertheless, one may expect that the analyst is able to construct an instrument such that its signals for other shocks do not exceed the signal for the main shock of interest. With these considerations we formalize a scenario of instrument contamination as

$$\begin{aligned} z_{1t}^{(3)} &= 0.40\varepsilon_{1t} + 0.30\varepsilon_{2t} + 0.30\varepsilon_{3t} + \xi_{1t}, \\ z_{2t}^{(3)} &= 0.25\varepsilon_{1t} + 0.50\varepsilon_{2t} + 0.25\varepsilon_{3t} + \xi_{2t}, \\ z_{3t}^{(3)} &= 0.20\varepsilon_{1t} + 0.20\varepsilon_{2t} + 0.60\varepsilon_{3t} + \xi_{3t}. \end{aligned} \quad (17)$$

3.4 Further comments on the simulation design

Overall, we perform $S = 1000$ replications of each Monte Carlo experiment. Three alternative sample sizes, $T = 120, 240$ and $T = 480$ shall represent macroeconomic data measured at the quarterly or monthly frequency, respectively. The length of simulated time series is set to $T + 1000$ and the first 1000 realizations of each generation are discarded to immunize simulation outcomes against initial values.

Throughout, we assume that conditional on the sample information, the analyst is able to specify the correct VAR order and consistently estimate the reduced-form model. Conditional on the outcome of a given identification scheme (indicated by ‘•’) the analyst determines the impact multipliers and the implied IRFs (up to horizon h) which we denote henceforth as \hat{B}^\bullet and $\hat{\Theta}_i^\bullet, i = 0, 1, \dots, h$, respectively.

For the application of sign restrictions, Table 2 summarizes the agnostic sign pattern imposed on the instantaneous responses of the system variables.⁹ Henceforth, the structural covariance

⁹We refrain from restricting the output reaction to monetary surprises and let the data decide on this issue. While Christiano et al. (1999), Uhlig (2005), Smets and Wouters (2007) and Ahmadi and Uhlig (2015) provide evidence of a rather modest (or even neutral) effect of MP on output, Romer and Romer (2004) and Gertler and Karadi (2015) show that a contractionary MP is quite effective in muting economic activities (see Ramey (2016) for a comprehensive discussion).

Variable	Shock		
	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$
x	+	−	?
π	+	+	−
r	+	+	+

Table 2: Imposed sign restrictions. Structural shocks ε_k refer to a demand shock ($k = 1$), a supply shock ($k = 2$) and an MP shock ($k = 3$).

decomposition matrix and IRFs identified by means of sign restrictions are denoted as \hat{B}^{SR} and $\hat{\Theta}_i^{SR}$, respectively. Identification by means of volatility shifts requires knowledge of the time point T_B at which the shift occurs. In empirical practice, this can be either known to the econometrician or she could resort to historical information about relevant events that have led to a permanent institutional or behavioral change (Bacchiocchi and Fanelli 2015). Alternatively, she could apply suitable consistent tests for this purpose. As previously mentioned, given the covariance design in (12), the econometrician is expected to apply identification based on the second break-point T_{B_2} . Accordingly, break-point ‘estimates’ are drawn as in Herwartz and Plödt (2016)

$$T_B^* = \left(\frac{2}{3} + \frac{2}{3} \frac{f}{\sqrt{T}} \right) T, \quad f \sim N(0, 1), \quad (18)$$

such that the estimated break-point converges to the true one, i.e. $T_B^* \xrightarrow{p} T_{B_2} = \frac{2}{3}T$. Structural covariance decomposition matrices and IRFs gathered by changes in volatility are denoted as \hat{B}^{CV} and $\hat{\Theta}_i^{CV}$, respectively. For outcomes from the extraction of independent components, we use $\bullet = dCov$ and $\bullet = CvM$ if criterion minimization refers to the distance covariance and the Cramér-von-Mises distance, respectively (to obtain \hat{B}^{dcov} , \hat{B}^{CvM} , $\hat{\Theta}_i^{dCov}$ and $\hat{\Theta}_i^{CvM}$). Similarly, \hat{B}^{IV} , \hat{B}^{IV-W} , \hat{B}^{IV-E} and $\hat{\Theta}_i^{IV}$, $\hat{\Theta}_i^{IV-W}$, $\hat{\Theta}_i^{IV-E}$ are used to indicate identified proxy SVARs that rely on the instruments (13), (14) and (15), respectively.

3.5 Performance evaluation

One particular purpose of performing SVAR analysis is to trace the dynamic responses of macroeconomic variables to the structural shocks of interest. Therefore, our main evaluation criterion measures whether alternative identification schemes can (approximately) recover the true underlying structural IRFs in (3). However, assessing their performance simply based on the (squared) distances between the resulting dynamic IRFs and the true ones might be subject to criticism, since the performance of alternative identification schemes might depend on the model calibration as implied by the DGP. Herwartz and Plödt (2016) construct ‘areas of acceptance’ (AoA) around the IRFs implied by the chosen calibration and evaluate various approaches based on the (scale-free) frequencies of identified IRFs located in these AoA. First, following Canova and Paustian (2011), we specify for each DSGE model parameter an interval that is chosen to be sufficiently large, to contain theoretically reasonable

values, and sufficiently tight to obtain unique sign patterns at least for the impact multiplier matrix.¹⁰ Second, we randomly draw a set of parameter values from uniform distributions on the parameter intervals, and compute IRFs arising from such calibrations up to horizon $h = 15$. We define point-wise 5% and 95% quantiles from these IRFs based on 10000 draws. Subsequently, we check in each experiment whether the impulse response point estimates gathered by alternative identification approaches are located within these AoA, and we finally report the respective acceptance frequencies.¹¹

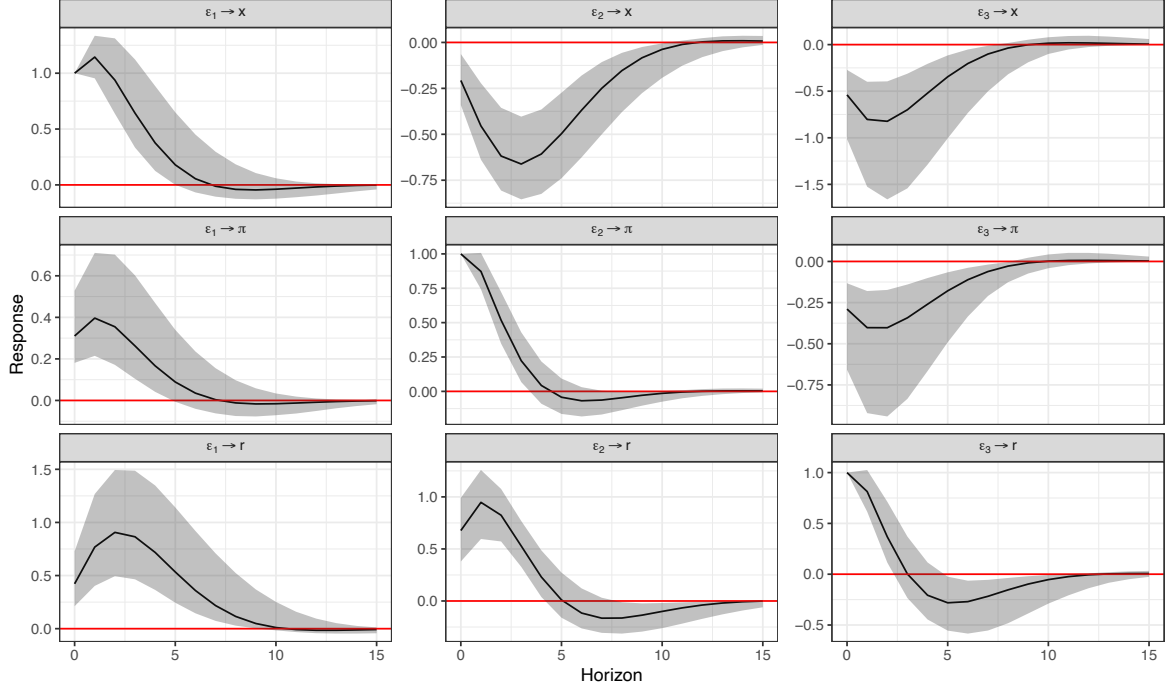


Figure 1: Areas of acceptance. The solid line represents the true IRFs conditional on the chosen parameterization (10). The shaded area represents 5% and 95% quantiles of IRFs resulting from 10000 draws of DSGE parameters from uniform distributions with corresponding intervals summarized in Table A1. Variables included in the system are output gap x , inflation π and the nominal interest rate r . Structural shocks are demand shock ε_1 , supply shock ε_2 and MP shock ε_3 .

Given the parameterization in (10), the true impulse responses and the AoA are shown in Figure 1. Note that all structural shocks are normalized to have a positive unit impact effect along the diagonal. Hence, the acceptance frequencies of the respective IRF estimates are equal to 1 by construction for all identification schemes.¹²

¹⁰Herwartz and Plödt (2016) show that the 5% and 95% quantiles of posterior densities from a typical Bayesian estimation of the DSGE model do not produce a clear-cut sign pattern for the B matrix and hence are not suitable for AoA construction. Parameter intervals used in this study are reported in Appendix B.

¹¹Since the column orderings of \hat{B}^\bullet , $\bullet \in \{CV, dCov, CvM\}$ are not unique, we select from six alternative column permutations the one that yields the largest overall frequency estimate.

¹²This is achieved by dividing each column in matrices \hat{B}^\bullet by the respective main diagonal element of this column. The normalization of the impact multiplier is applied only for the construction of the AoA and the

We consider two further loss functions with a specific focus on the purposes of MP analysis. In this respect, it is first documented how often the identification scheme employed obtains a shock that triggers the effects of an MP shock uniquely (henceforth denoted UMPS). Accordingly, the loss function is defined as:

$$g^\bullet = \frac{1}{S} \sum_{s=1}^S \mathbb{1} \left(\text{UMPS}^{(s)} \right), \quad (19)$$

where $\mathbb{1} \left(\text{UMPS}^{(s)} \right) = 1$ if in experiment s there exists only one k such that $\hat{b}_{2,k}^{(s)} \hat{b}_{3,k}^{(s)} < 0$. Since sign restrictions achieve a full score by construction, we do not report g^{SR} .

Second, conditional on an UMPS, we consider a relative root mean squared error (RMSE) of effect estimates defined as:

$$RMSE_k^\bullet = \frac{1}{Sg^\bullet} \sum_{s=1}^S \sqrt{\left(\frac{\hat{b}_{k,3}^{(s)} - b_{k,3}^{(s)}}{b_{k,3}^{(s)}} \right)^2 \mathbb{1} \left(\text{UMPS}^{(s)} \right)}, \quad k = 1, 2, 3, \quad (20)$$

where $Sg^\bullet = \sum_{s=1}^S \mathbb{1} \left(\text{UMPS}^{(s)} \right)$ is the number of simulation experiments obtaining a UMPS. We additionally report the average relative RMSE of the estimated impacts for each identification scheme $\overline{RMSE}^\bullet = \frac{1}{K} \sum_{k=1}^K RMSE_k^\bullet$.

3.6 Monte Carlo results

This section presents the simulation results. We place a particular focus on identifying structural shocks with relatively low signal strength, in our case the MP shock. With such a focus on partial identification, our simulation study is in a similar vein to the study by Mumtaz et al. (2018) on detecting credit shocks. However, unlike this benchmark study, the present Monte Carlo analysis places more weight on the realistic situation that empirical data are at odds with stylized modeling assumptions and hence the performance of alternative identification schemes is subject to nuisance (parameters).

We begin by assessing the performance of SVARs based on simulated time series with length $T = 240$ under the baseline distributional scenario \mathcal{L}_I (Gaussian heteroskedastic shocks) and display the acceptance frequencies of estimated IRFs in Figure 2. Simulation results for a variety of distributional scenarios and alternative sample sizes are subsequently discussed. Figure 3 displays the frequencies of detecting a UMPS and the average relative RMSEs of impact estimates, which are documented in further detail in Tables A4 and A5, respectively.¹³

3.6.1 Theory- and data-based identification of baseline DGPs

The performance of identification by agnostic sign restrictions considerably varies across the dynamic relationships. Responses of model variables (except for inflation π) to the identified calculation of their coverage frequencies, but not for other evaluation criteria like g^\bullet in (19) or $RMSE_k^\bullet$ in (20).

¹³Some detailed simulation results are documented in Appendix C as follows: Tables A2 and A3 show the average acceptance frequencies of responses to each structural innovations from a short- ($i = 0, \dots, 4$) and long-run ($i = 5, \dots, 15$) perspective, respectively.

demand shock can be relatively well captured by this approach. However, dynamic effects of the MP shock are only poorly detected. The identified short-run dynamics triggered by MP shocks are located within the AoA in less than 20% of all replications. More specifically - and in line with Castelnuovo (2012) - about 80% of admissible models imply that a contractionary MP shock boosts output, which does not comply with the DGP employed. Apart from a lack of directional or sign accuracy, the quantification of MP effects is seriously questioned by the sizable relative RMSEs of the associated structural parameters in the estimated decomposition matrices. We consider this deficiency of agnostic sign restrictions as a result of the relatively low signal strength of the structural shock of interest (see also the respective discussion in Section 3.7).

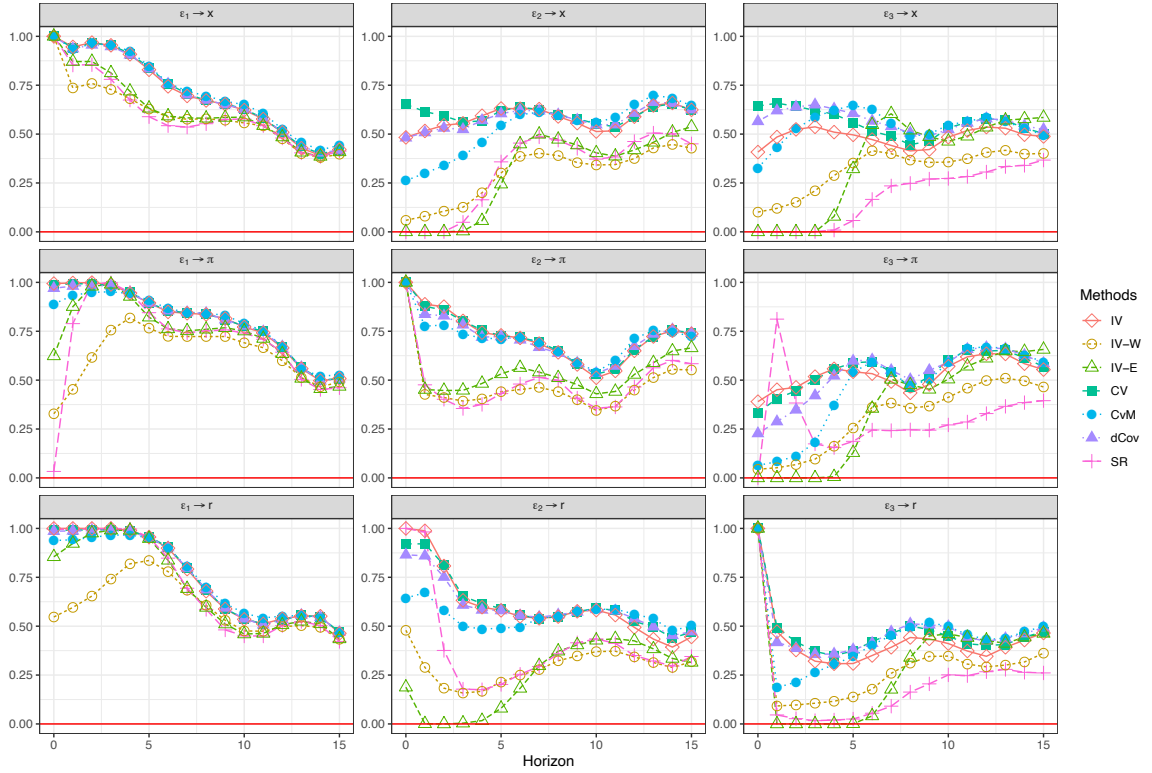


Figure 2: Distribution scenario \mathcal{L}_I . Results are based on $S = 1000$ Monte Carlo experiments with a sample size $T = 240$. The identification approaches employed are: *CV*: Identification by heteroskedasticity; *CvM*: Identification by minimizing Cramér-von-Mises distance; *dCov*: Identification by minimizing distance covariance; *SR*: Identification by sign restrictions; *IV*, *IV-W* and *IV-E*: Proxy SVAR based on instruments (13), (14) and (15), respectively.

By contrast, dynamic IRFs gathered by statistical identification techniques are considerably more common within the defined AoA. For example, in almost two-thirds of all replications, the statistically-identified short-run responses of output gap to the MP shock (except for identification by minimizing the CvM distance) are located within the defined acceptance range. Confirming the Monte Carlo outcomes of Herwartz and Plödt (2016), identification based on (co)variance shifts outperforms all other competitors in detecting the true impulse

response coefficients with average acceptance frequencies of 75.3% and 59.2% for periods $i = 0, \dots, 4$ and $i = 5, \dots, 15$, respectively. Turning to ICA-based identification, the minimization of the distance covariance performs close to the exploitation of (co)variance shifts. Among the alternative independence measures, we find that minimization of the distance covariance is preferable to minimization of the CvM distance, as it allows for a more frequent labeling of UMPS and produces more accurate estimates for the respective effects.

3.6.2 Proxy SVARs and baseline DGPs

If the econometrician manages to construct external measures capturing the dynamics of the underlying shocks with reasonable signal-to-noise ratios, using these instruments for identification would produce structural decompositions and IRFs that are quite close to the true ones. Given that the instruments defined in (13) are valid (exogenous) and strong (with average F -statistics of 77.1), identification by means of proxy SVARs succeeds in detecting a UMPS with the largest frequency (76.1%), and it provides the most accurate assessment of the instantaneous reactions to the UMPS. However, this methodology may severely suffer from the weakness of instruments (14). On average, the identified short-run IRFs $\hat{\Theta}^{IV-W}$ are outside the acceptance areas in more than 60% of the experiments. In only 18.4% of all replications, \hat{B}^{IV-W} provides a UMPS with an average relative RMSE of about 2.25, which is the largest aside from \overline{RMSE}^{SR} . Finally, if instruments are contaminated in such a way that they simultaneously correlate with multiple shocks in the model, the performance of proxy SVARs strongly depends on the relative signal of the underlying structural shock. If the contribution of the shock to the system variation is relatively small (e.g. here the contribution of the MP shock), it is almost impossible for these instruments to pin down the respective structural parameters. As a particular result, \hat{B}^{IV-E} fails to deliver a UMPS in all Monte Carlo experiments (see also the theoretical discussion below in Section 3.7).

3.6.3 Alternative distributional scenarios and sample sizes

For a given sample size of $T = 240$ and under alternative distributional scenarios, we do not observe systematic improvements or deteriorations in identification performance for proxy SVARs in terms of qualitative (sign) or quantitative accuracy (see Tables A2, A3 and A5 from top to bottom panels). This also holds true for identification by means of sign restrictions and (co)variance shifts. By contrast, the displays in Figure 3 (from left to right) underline that skewed and leptokurtic shocks seem to enhance the informational content of independence diagnostics (see also Herwartz et al. 2019). Regarding alternative sample sizes, simulation results show that the performance of both data-driven approaches and proxy SVARs based on valid instruments remarkably improve with increasing sample information (see Figure 3). Moreover, estimation biases of using weak or endogenous instruments fully manifest in large samples. In particular, endogenous proxies fail to detect shocks featuring opposite impact effects on interest rates and inflation in all experiments, regardless of the sample size and

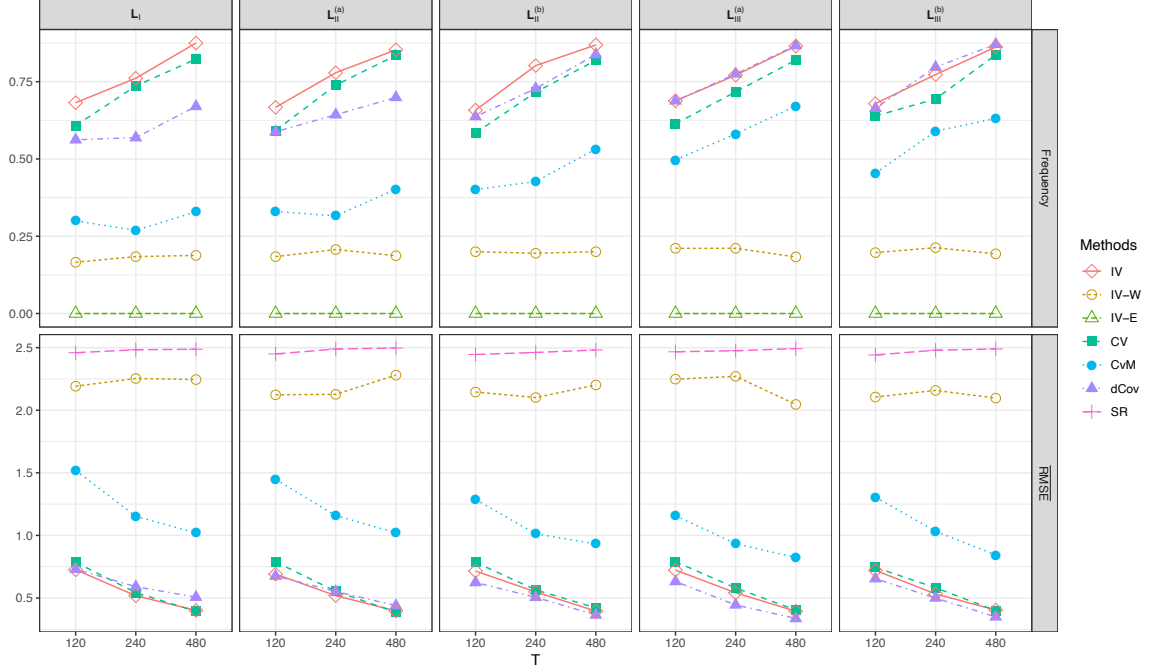


Figure 3: Frequency of detecting an UMPS and average relative RMSEs of impact estimates (see also the definitions in (19) and (20)).

shock distribution (see Table A4).¹⁴

3.7 Role of the relative signal strength of the shocks

From the Monte Carlo results discussed in the previous section, we note that agnostic sign restrictions and proxy SVARs based on endogenous instruments are particularly sensitive to the relative signal strength of the underlying shock.¹⁵ Paustian (2007) analyzes this issue for sign restrictions à la Uhlig (2005) and finds that the standard deviation of the MP shock needs to be extremely large for this identification scheme to deliver a clear-cut sign pattern regarding the unrestricted output response.

Along similar lines as in Paustian (2007), we illustrate this problem by means of a simple generic example. Consider a two-dimensional VAR, where the linear mapping between the residuals and the underlying structural shocks is

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ + & - \\ b_{21} & b_{22} \\ - & + \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \quad (21)$$

Unlike the outline in Section 2, we allow structural shocks to have distinct variances $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon = \text{diag}(\sigma_1^2, \sigma_2^2)$, such that $\sigma \equiv \sigma_2^2/\sigma_1^2$ is the variance of the second shock relative to that of the first one. Turning to identification, the agnostic sign restrictions on the covariance factor

¹⁴In analogy to Figure 2, graphical displays of detailed simulation outcomes for all distributional scenarios and alternative sample sizes are available from the authors upon request.

¹⁵This can be easily seen by comparing the accuracy of the resulting responses to the demand shock with those to the MP shock (see Table A2).

B involve the sign pattern $\begin{bmatrix} m_{11} & m_{12} \\ + & ? \\ m_{21} & m_{22} \\ - & + \end{bmatrix}$. Aside from the imposed sign restrictions, all accepted candidates for matrix B - say M - must fulfill the moment condition $MM' = \Sigma_u = B\Sigma_\varepsilon B'$, i.e.

$$\begin{bmatrix} m_{11}^2 + m_{12}^2 & m_{11}m_{21} + m_{12}m_{22} \\ * & m_{21}^2 + m_{22}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 b_{11}^2 + \sigma_2^2 b_{12}^2 & \sigma_1^2 b_{11}b_{21} + \sigma_2^2 b_{12}b_{22} \\ * & \sigma_1^2 b_{21}^2 + \sigma_2^2 b_{22}^2 \end{bmatrix} \quad (22)$$

After some algebra, it is easy to see that $m_{12} = \frac{\sigma_1^2 b_{11}b_{21} + \sigma_2^2 b_{12}b_{22}}{m_{22}} - \frac{m_{11}m_{21}}{m_{22}}$, where $-\frac{m_{11}m_{21}}{m_{22}}$ is restricted to be positive. In order to obtain the correct sign, negativeness of the first term $\frac{\sigma_1^2 b_{11}b_{21} + \sigma_2^2 b_{12}b_{22}}{m_{22}}$ requires that the relative volatility σ exceeds the critical value $-\frac{b_{11}b_{21}}{b_{12}b_{22}}$ (Paustian 2007).

Proxy SVAR analysis faces a similar issue once the exogeneity condition in (5) is violated. Assume that an instrument z_{2t} for the second structural shock is available for identification of B_2 , which takes the form $z_{2t} = \gamma_1 \varepsilon_{1t} + \gamma_2 \varepsilon_{2t}$, $\gamma_1, \gamma_2 \neq 0$. Since $\gamma_1 \neq 0$, z_{2t} is correlated with both shocks in the system. Let Γ collect the population covariances between the instrument and the shocks, i.e. $\mathbb{E}(\varepsilon_t z_{2t}) = \begin{bmatrix} \gamma_1 \sigma_1^2 \\ \gamma_2 \sigma_2^2 \end{bmatrix} \equiv \Gamma$. Intuitively, the population covariance between the instrument and the reduced-form residuals is no longer pointing in the direction of B_2 , but now to a linear combination of both columns of B, i.e.

$$\mathbb{E}(u_t z_{2t}) = \gamma_1 \sigma_1^2 B_1 + \gamma_2 \sigma_2^2 B_2 = B\Gamma. \quad (23)$$

Instead of the squared covariance between the instrument and the shock of interest as shown in (8), this cross-moment now gives the variance of the instrument:

$$\Pi \mathbb{E}(u_t z_{2t}) = \Gamma' B' [BB']^{-1} B\Gamma = \Gamma' \Gamma = \gamma_1^2 \sigma_1^4 + \gamma_2^2 \sigma_2^4. \quad (24)$$

While one obtains $B\Gamma [\Gamma' \Gamma]^{-\frac{1}{2}} = B_k$ for valid instruments, in a more general case ($\gamma_1 \neq 0$) it holds that $B\Gamma [\Gamma' \Gamma]^{-\frac{1}{2}} = B\beta = \sum_{k=1}^K B_k \beta_k$, where β is a $K \times 1$ vector. The identification result can be interpreted as a weighted sum of columns in B with weight parameters collected by β . In our bivariate case:

$$\beta = \Gamma [\Gamma' \Gamma]^{-\frac{1}{2}} = \begin{bmatrix} \frac{\gamma_1 \sigma_1^2}{\sqrt{\gamma_1^2 \sigma_1^4 + \gamma_2^2 \sigma_2^4}} \\ \frac{\gamma_2 \sigma_2^2}{\sqrt{\gamma_1^2 \sigma_1^4 + \gamma_2^2 \sigma_2^4}} \end{bmatrix} = \begin{bmatrix} \frac{\gamma_1}{\sqrt{\gamma_1^2 + \gamma_2^2 \sigma^2}} \\ \frac{\gamma_2}{\sqrt{\gamma_1^2 / \sigma^2 + \gamma_2^2}} \end{bmatrix} \equiv \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}. \quad (25)$$

Despite the violation of the exogeneity condition, it can be easily shown that if σ is sufficiently large, identification based on this instrument can still pin down the structural parameters of interest. By contrast, if the relative signal strength of the shock of interest shrinks to zero, the weight attached to the corresponding column of the B matrix - in this case - β_2 will vanish. Consequently, the structural model cannot be recovered. These theoretical insights

are also interesting in the light of recent advances by Wolf (2020) who rationalized the role of relative shock volatility from a Bayesian perspective for set-identified systems. However, suggesting proxy SVARs as a viable alternative to avoid misidentification of MP shocks in such a context, Wolf (2020) assumes exogeneity. Accordingly, our results highlight that issues of relatively low signal strength might remain unresolved in realistic scenarios where proxy SVARs condition on contaminated instrumental information.

4 Structural models of US monetary policy

In this section, we consider the small-scale macroeconomic model of Section 3 as a convenient framework to identify MP shocks and address whether MP interventions have been effective in dampening inflation or affecting real economic activity in the US. Prior to a comparative assessment of the relative merits of theory- and data-based identification on the one hand and proxy SVARs on the other, we carefully discuss the suitability of the low-dimensional empirical model for identifying fundamental shocks. Extending beyond unconditional features of the structural model (impact estimates, IRFs, diagnostic tests), Section 5 provides an assessment of the structural profiles identified to capture important narratives on MP in the US, especially the structural contributions to the *Great Inflation* in the 1970s, as well as the Volcker disinflation in the early-1980s.

4.1 Reduced-form model and fundamentalness

The trivariate baseline system comprises the percentage log deviation of real GDP with respect to the estimate of potential output by the Congressional Budget Office (x_t), annualized quarter-on-quarter growth rates of the GDP deflator (π_t) and the one-year government bond rate (GBR1_t) as a policy indicator.¹⁶ Quarterly time series data cover the period from 1965Q1 until 2008Q3. The VAR model includes an intercept term and lags up to order $p = 4$, which is chosen based on the AIC and the Lagrange multiplier (LM) test for no remaining serial correlation. Throughout, we subject the same reduced-form (model) residuals to structural analysis, i.e. we refrain from re-estimating the VAR conditional on sample periods for which specific instrumental information is available.

As placed on the structural model in (2), the assumption of causality is essential for identification of the SVAR model, and it ensures that the process has a Wold moving average representation. Put differently, in its VAR form the structural model is only properly identified if reduced-form disturbances u_t can be traced back to the current and historic shocks ε_{t-j} , $j \geq 0$.¹⁷ While the stylized trinity model provides a convenient framework for Monte

¹⁶Considering that the Fed has been increasingly relying on communication strategies to manage expectations about the future path of interest rates (Ramey 2016), the one-year government bond rate promises a more comprehensive measure for intentions of the monetary authority, since it includes information about surprises in the forward guidance (Gertler and Karadi 2015). One-year government bond rates have been downloaded from Thomson Reuters Datastream (code: `USTRCN1.`), while series for the output gap and inflation have been obtained from Herwartz and Plödt (2016).

¹⁷The case of non-causality - i.e. the dependence of y_t on future shocks ε_{t+j} , $j > 0$ - holds particular

Carlo analysis, it could be prone to risks of variable omission and non-invertibility in empirical practice. Hence, it is unclear whether such a model is suitable for identifying strictly fundamental shock, and in particular fundamental MP shocks. However, we believe that adverse effects of omitted information are likely of minor concern for the detection of MP shocks, since their opposite impact on prices and interest rates remains unique in many higher-dimensional systems. For instance, Plagborg-Møller and Wolf (2019) show by means of simulated data from the medium-scale (i.e. seven shock) DSGE model of Smets and Wouters (2007) that the MP shock is near-invertible within a trivariate model. Regarding prediction, Herbst and Schorfheide (2012) underpin that the performance of a larger-scale model in the vein of Smets and Wouters (2007) is not necessarily superior to that of a related small-scale model. In this work, we address the issue of fundamentalness from two complementary angles. First, we explicitly document results from testing the reduced-form model for fundamentalness. Hamidi Sahneh (2016) has recently proposed an approach to test whether reduced-form residuals u_t accord with a multivariate martingale difference scheme that implies the fundamentalness of structural shocks. We perform the test based on a generalized spectral density and do not obtain evidence against the martingale difference null hypothesis. Specifically, the estimated test statistic implies a p -value of 0.90.¹⁸ Second, we augment the small-dimensional (S)VAR by means of (up to three) factors that we extract from data collected by Michael McCracken (FRED-QD), and inquire whether core structural outcomes from the small-scale model remain robust within the information-rich environment. Instead of providing proxy SVAR outcomes from FAVARs in full detail, we report results from information-rich models below to complement the empirical evidence characterizing our main results for identified benchmark MP (proxy) shocks.¹⁹

4.2 Theory- and data-based identification

We impose the same (agnostic) sign pattern on the impact multipliers as documented in Table 2 for theory-based identification. The leftmost column of Figure 4 shows structural responses to a contractionary MP shock. Throughout, we normalize the size of the MP shock such that it raises GBR1 on impact by 25 bps. The results look quite similar to those in Uhlig (2005), although a smaller model is employed here, and we place restrictions

interest, if agents receive information or form expectations about future innovations (for non-causal models, see Lanne and Saikkonen (2013)).

¹⁸The test outcomes are robust to kernel choice (Parzen, Daniell, Quadratic spectral). We thank Mehdi Hamidi Sahneh for providing us with easy-to-use MATLAB codes.

¹⁹Specifically, we select from FRED-QD 220 series that are available for the time period covered by our baseline VAR. For a detailed description of the series and transformation codes to obtain stationarity, the reader may visit the website of the Federal Reserve Bank of St. Louis (<https://research.stlouisfed.org/econ/mccracken/fred-databases>) and consult McCracken and Ng (2020). First, we extract r common factors (denoted \tilde{F}_t) from these variables by means of principal component estimation, where $1 \leq r \leq 3$. Subsequently, we use linear projections to extract from \tilde{F}_t the part that is orthogonal to the variables y_t in our system and denote these factors F_t . Finally, we recover the structural form of VARs for $(y'_t, F'_t)'$ by means of alternative identification schemes (for a lead study in the field of FAVARs see Bernanke et al. (2005)). Detailed results from structural FAVARs are available from the authors upon request.

only on the impact effects. About 71% ($\approx 2/3$ in Uhlig 2005) of all admitted models obtain positive instantaneous responses of the output gap to an unexpected monetary tightening. The reaction of output is ambiguous for all horizons. Based on this finding, Uhlig (2005) resumes that monetary neutrality is not inconsistent with the data. While such a conclusion casts doubts on conventional views of MP transmission, the uncertain response of output might be due to the lack of identification power of this particular approach induced by the fact that the MP shock has a relatively small signal strength. Similar to the results in Uhlig (2005), the identified MP shock only accounts for a small fraction of the forecast error variance for the output gap.²⁰

break-point	Test statistic	95% critical value	p -value
1979Q3	$\lambda_{BP} =$	580.80	628.60
	$\lambda_{SS} =$	115.60	128.20
1984Q1	$\lambda_{BP} =$	586.86	562.99
	$\lambda_{SS} =$	89.41	86.04

Table 3: Break-point (λ_{BP}) and sample-split (λ_{SS}) tests for structural change (see Chapter 3 of Lütkepohl and Krätzig 2004). Both critical values and p -values are based on 2000 bootstrap replications.

Observing some consensus concerning regime switches of MP in the US (Christiano et al. 1999, Bernanke and Mihov 1998, Sims and Zha 2006), it is worth exploiting CV for identification purposes. Reflecting the beginning and accomplishment of the Volcker disinflation (see also the narrative analysis in Section 5), two dates have been considered as potential (co)variance break-points, namely 1979Q3 and 1984Q1. While Bacchiocchi and Fanelli (2015) and Herwartz and Plödt (2016) perform their analysis based on 1979Q3, Perez-Quiros and McConnell (2000) and Lanne and Lütkepohl (2008) focus on 1984Q1.

The null hypothesis of covariance stability in 1979Q3 cannot be rejected at common significance levels, while the evidence of a structural break in 1984Q1 is very strong (see Table 3).²¹

Conditional on $T_B^* = 1984Q1$, the estimated decomposition matrices are (with standard errors in parentheses):

$$\hat{B}^{CV} = \begin{bmatrix} 0.734 & -0.227 & -0.556 \\ (0.166) & (0.520) & (0.178) \\ 0.051 & 1.359 & -0.515 \\ (0.824) & (0.151) & (0.372) \\ 0.689 & 0.328 & 0.531 \\ (0.217) & (0.489) & (0.180) \end{bmatrix} \text{ and } \hat{\Psi} = \begin{bmatrix} 0.296 & 0 & 0 \\ (0.065) & & \\ 0 & 0.233 & 0 \\ & (0.051) & \\ 0 & 0 & 0.115 \\ & & (0.025) \end{bmatrix}. \quad (26)$$

²⁰ Aside from the first quarter following the shock, the identified MP shock contributes less than 10% of the forecast error variance of the output gap.

²¹ Regarding regime switches in US MP, several analysts do not find sufficient evidence against the stability of the reduced-form system (Bernanke and Mihov 1998, Sims and Zha 2006). It is therefore reasonable to assume that regime change is specific to the variance of the structural shocks.

The sign pattern of \hat{B}^{CV} matches the typical instantaneous effects of demand, supply and MP shocks (see Table 2). Accordingly, we label the third shock as the MP shock. The test results documented in Table 4 highlight the (partial) identifiability of the MP shock, and indicate that the period starting with 1984Q1 features a lower variation of all shocks and especially the MP shock. Notably, this result corresponds to stylized findings for the period of the *Great Moderation*.

H_0	Test statistic	p -value
$\varphi_1 = \varphi_2$	0.58	0.45
$\varphi_1 = \varphi_3$	6.74	0.01
$\varphi_2 = \varphi_3$	4.27	0.04

Table 4: Pairwise Wald tests for equality of diagonal elements in $\hat{\Psi}$.

Heteroskedasticity-based IRFs are shown in the middle column of Figure 4. A contractionary MP shock is followed by an immediate reduction of the output gap by more than 25 bps. The output reaction remains negative for ten quarters (significant up to 2 years) and becomes insignificantly positive thereafter. Inflation responds negatively for all horizons, with significant effects emerging after a propagation period of approximately two quarters.

Based on ICA, the uniqueness of a structural model requires that at most one shock in the system is Gaussian (Comon 1994). For a respective diagnostic analysis, we apply fourth-order blind identification, as suggested by Nordhausen et al. (2017a). Indicating strong deviations from Gaussianity, the null hypothesis stating that the last (first) $K - k_0$ (k_0) shocks are Gaussian (non-Gaussian), can be rejected for both $k_0 = 2$ and $k_0 = 1$ with 1% significance.²² For ICA modeling, we employ the distance covariance, which showed a superior identification performance in the simulation study of Section 3. The estimated structural impact multiplier (with bootstrap standard errors in parentheses) aligns with the stylized sign pattern of Table 2 and reads as

$$\hat{B}^{dCov} = \begin{bmatrix} 0.355 & -0.249 & -0.565 \\ (0.244) & (0.231) & (0.157) \\ 0.157 & 1.036 & -0.312 \\ (0.232) & (0.147) & (0.365) \\ 0.682 & 0.098 & 0.154 \\ (0.089) & (0.148) & (0.265) \end{bmatrix}.$$

Hence, the third shock allows for labeling as an MP shock. As displayed in the right-most column of Figure 4, an unexpected monetary tightening leads to a strongly negative response of the output gap up to nearly -130 bps after two quarters. The output reduction remains significant for two years following the MP shock. The profile of the inflation response is qualitatively similar to that obtained by means of CV, although it is larger in magnitude.

²²Specifically, we employ the function `F0BIboot` from the R package `ICtest` (Nordhausen et al. 2017b) with 2000 bootstrap replications. We also apply univariate Jarque-Bera (JB) tests to each VAR residual series and obtain equivalent test decisions.

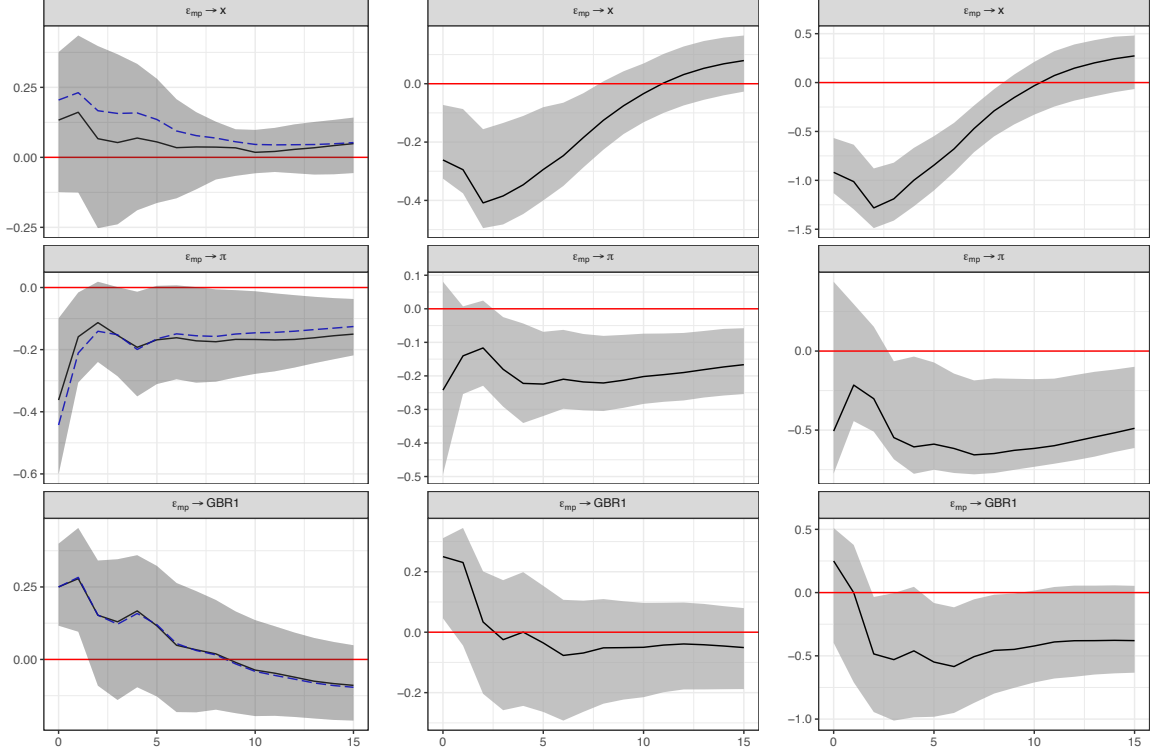


Figure 4: Responses to an MP shock up to quarter $h = 15$. From left to right: *First column*: Identification by agnostic sign restrictions. Median (solid) and MT (dashed) IRFs with 16% and 84% quantiles from 1000 accepted structural models. *Second (Third) column*: Identification by changes in volatility (minimizing the distance covariance). IRFs with 16% and 84% quantiles from wild bootstrap with 2000 replicates. Variables included in the system are output gap x , inflation π and one-year government bond rate GBR1. The size of MP shock is normalized to impact on the one-year rate by 25 bps.

4.3 Proxy SVARs

We consider a wide range of potential external shock series that have been constructed under distinct frameworks (narrative identification, high-frequency identification, SVARs and DSGE models), and have already attracted a lively discussion in the literature (e.g. Stock and Watson 2012, Ramey 2016, Miranda-Agrippino and Ricco 2017). In order to qualify for structural analysis, instruments must be (i) reasonably relevant for the estimated reduced-form system, and (ii) free of endogenous contamination. We first test for relevance of instruments and then overview the identified shocks in terms of their impulse responses. With these insights, we are in a position to subsequently address exogeneity issues from complementary angles, i.e. (i) insights from the related literature, (ii) correlations between MP shocks obtained by means of proxy SVARs and CV, and (iii) LR tests developed by Podstawski et al. (2018).

4.3.1 Instrument relevance

Table A6 in the Appendix D documents in detail the instruments considered and inferential outcomes of testing the null hypothesis of weak instrumentation (Lunsford 2015). Only

Instrument	SW	RR	SZ	RR*	SZ*
x_t	-0.166 (0.105)	0.096 (0.126)	0.081 (0.088)	-0.375 (0.171)	-0.383 (0.125)
π_t	-0.299 (0.135)	0.245 (0.106)	0.139 (0.102)	-0.347 (0.144)	-0.355 (0.142)
GBR1 $_t$	0.525 (0.128)	0.683 (0.112)	0.702 (0.103)	0.358 (0.293)	0.366 (0.253)
F -statistic	19.81	32.435	39.336	10.929	16.579

Table 5: Contemporaneous responses to an MP shock identified by means of external instruments (with MBB standard errors in parentheses) and tests for weak instrumentation. The critical values at which one can reject the null hypothesis with threshold signal-to-noise ratios corresponding to a 5% asymptotic bias at 5% (10%) significance level are 13.28 (11.20) for $K = 3$ (Lunsford 2015).

three proxies proved to be strong instruments with at least 5% significance, namely (i) innovations to the estimated MP reaction function in the medium-scale DSGE model of Smets and Wouters (2007) (henceforth denoted as SW with availability 1959Q1-2004Q4); (ii) intended Federal Funds Rate (FFR) changes during FOMC meetings, which are adjusted to be exogenous from the current rate and to the Fed’s Greenbook forecasts as constructed by Romer and Romer (2004) (RR, 1969Q2-2007Q4); and (iii) MP shocks from the best-fitting SVAR model in Sims and Zha (2006) (SZ, 1960Q2-2003Q1). By contrast - and consistent with the results of Miranda-Agrippino and Rey (2018) - instrumental information based on high-frequency data suffers from a lack of relevance, which might be due to limited sample sizes or issues of time aggregation.²³ We therefore focus on the instruments SW, RR and SZ in the following.²⁴

4.3.2 Impulse responses

Having reduced the set of informative instruments to three alternatives, we present the estimated structural parameters associated with the respective MP shocks in columns 1 to 3 of Table 5, while impulse responses are depicted in columns 1 to 3 of Figure 5. Depending on the instrument employed, the dynamic IRFs considerably differ. Proxy shocks from SW are unique in providing effect directions that are in line with the theoretical sign pattern of an MP shock. The response of inflation to a contractionary MP shock remains significantly below zero for almost all horizons. The instruments RR and SZ appear to share a similar informational content, and they are associated with a marked and persistent price puzzle, i.e.

²³Notably, high-frequency shock series are only available since 1988, the year when federal funds futures contracts were introduced by the Chicago Board of Trade. Hence, they do not cover the regime switch detected above at $T_B = 1984Q1$.

²⁴Standard Ljung-Box tests indicate that - unlike SW - the instruments RR and SZ are subject to serial correlation with common significance (see also Miranda-Agrippino and Ricco 2017). Accordingly, we adjust both series by regressing them on their lagged realizations (lag 1 for RR and lag 2 for SZ) and take the residuals for subsequent analysis. The empirical results for unadjusted instruments are qualitatively identical and quantitatively very similar to those documented in this work.

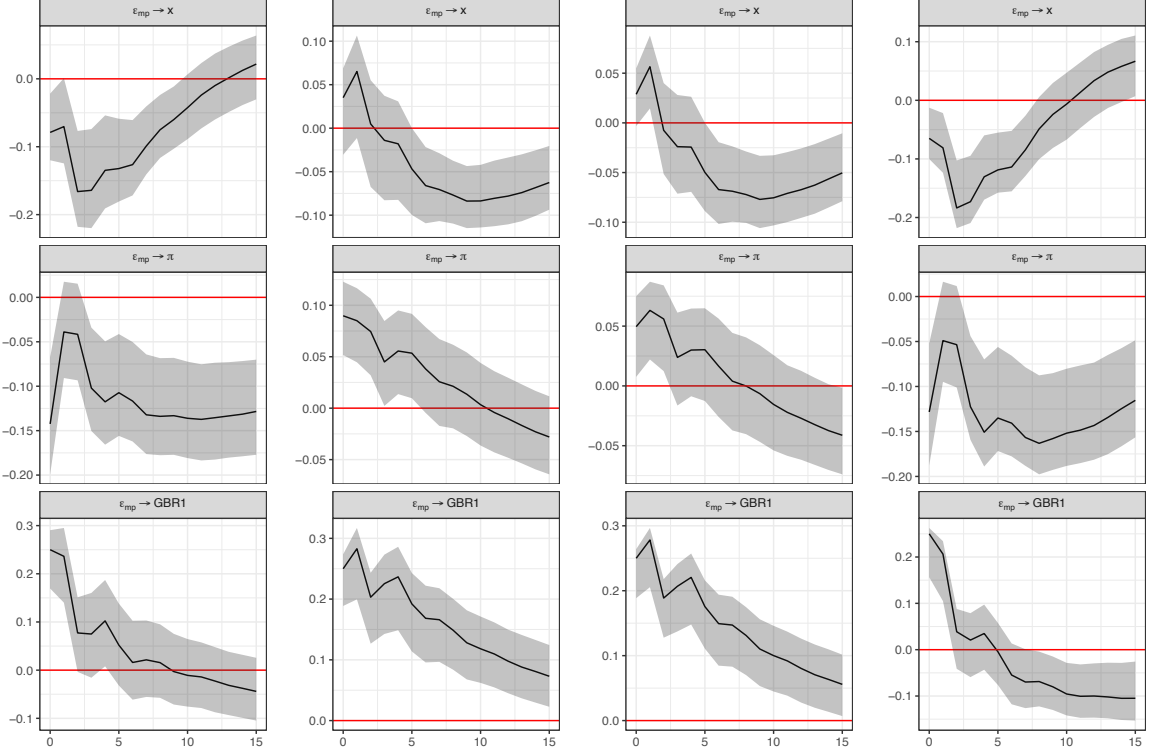


Figure 5: Responses to an MP shock up to quarter $h = 15$. The shaded area represents 16% and 84% quantiles from MBB with 2000 replicates. Proxy SVARs are based on instruments, from left to right: *First column*: SW (1959Q1 - 2004Q4); *Second column*: RR (1969Q2 - 2007Q4); *Third column*: SZ (1960Q2 - 2003Q1); *Fourth column*: SW in a data-rich environment (VAR augmented with $r = 3$ latent factors, the results based on $r = 1, 2$ factors are almost equivalent). The residual-based MBB is performed based on a block length of 15 for sample sizes varying from 149 (SZ) to 156 (SW) (for details see Jentsch and Lunsford 2019). Variables included are output gap x , inflation π and one-year government bond rate GBR1.

a co-movement of interest rates and inflation lasting up to 2.5 years after impact.

In order to understand the detected price puzzles for MP shocks conditioning on RR and SZ, it is instructive to recall the discussion in Section 3.7. Under instrument contamination, the proxy estimate of \hat{B}_3 is best understood as a weighted linear combination of all columns in the true matrix B . From (25), one can see that the bias (i.e. the false weighting) depends on (i) the signal from the instrument for the wrong shock relative to that for the shock of interest, and (ii) the relative signal strength of the shock of interest. Since data-based estimates obtained from the supposition of covariance change (see the result in (26)) imply that an aggregate supply shock leads to a stronger reaction of prices in comparison with an aggregate demand shock, one might expect that on average the contamination with the supply shock invokes a stronger upward bias for the price reaction than the contamination with the demand shock. Regarding the first origin of bias (i), our discussion of instrument exogeneity below (see also Table 6) suggests that the misidentification could indeed be the

result of the contemporaneous mixing of instruments with the supply shock. Unlike SW, both the instruments RR and SZ strongly correlate with the aggregate supply shock identified through covariance change. The resulting bias is aggravated by the fact that MP shocks contribute least to the system variation (ii).

The IRFs displayed in the right-most column of Figure 5 are obtained from proxy SW shocks conditioning on an information-rich framework (FAVAR, $r = 3$). As a strong underpinning of their fundamentality, it is striking to see that SW proxy MP shocks govern almost identical impact and dynamic responses irrespective of using the small-scale (Figure 5, column 1) or the factor-augmented model (column 4). Owing to their theory-conforming response profiles, we henceforth consider the MP shocks derived from SW (proxy SVARs) and CV (data-based) as benchmark MP shocks (see column 2 of Figure 4 and column 1 of Figure 5). In the subsequent discussions of exogeneity, we shed some light on potential reasons for the pronounced differences among proxy MP shocks conditioning on SW on the one hand and RR or SZ on the other.

4.3.3 Instrument exogeneity

In order to establish exogeneity of instrumental information, two general concerns warrant particular in the context of MP shock identification. First, asymmetries among the monetary authority and economic agents can lead to a mixing of genuine (exogenous) MP shocks with (endogenous) reactions of the monetary authority to information revealed about future (and possibly current) fundamentals (along with further references, see Noh 2018, Miranda-Agrippino and Ricco 2017, Miranda-Agrippino 2016). As a result, MP instruments frequently violate both the lag and the contemporaneous exogeneity condition (see Appendix E for a more detailed discussion of this issue). While lag endogeneity might be as serious a problem as contemporaneous endogeneity, our prime focus is on the latter for two reasons: first, as a result of temporal aggregation, contemporaneous effects appear more relevant for quarterly (i.e. low-frequency) empirical modeling; and second, the high-frequency instruments considered lack relevance.²⁵ Based on the findings in Jarociński and Karadi (2020), one could argue that under pronounced endogeneity proxy-identified SVARs would show ‘upward’ biases of the responses of inflation and output to contractionary MP shocks. Hence, observing instead a pronounced decline in inflation and output provides an informal indication against a sizable contamination with central bank information shocks. Unlike SW, proxy MP shocks conditioning on RR and SZ suffer from a marked price puzzle and hence might be subject to endogenous contamination (see columns 2 and 3 of Figure 5). Moreover, the latent MP shocks typically provide only limited contributions to observable system variances, which complicates the recovery of structural relations by means of instru-

²⁵Misidentification due to lag endogeneity can be alleviated by enhancing the info set of the VAR, while this is not applicable for coping with contemporaneous endogeneity (see also the respective discussion in Appendix E). In this regard, it is worth highlighting that the benchmark MP shocks retrieved from SW are remarkably robust after factor augmentation of the VAR model.

mental identification if the exogeneity condition (5) does not strictly hold. Hence, evaluating whether contemporaneous endogeneity is present appears especially important for MP analysis. In the following, we discuss further potential origins or indications of contemporaneous contamination of proxy-identified MP shocks.

Insights from the literature The set of shocks identified through instrumental information in Stock and Watson (2012) suggests that the considered shock series of SW, RR and SZ are indeed subject to contemporaneous contamination with non-MP shocks. Most prominently, RR- and SZ-based MP shocks exhibit sizable correlations with fiscal policy and oil shocks, whereas SW-based MP shocks only correlate sizeably with liquidity shocks. While the shock series obtained from SZ and SW have not been discussed extensively in the context of proxy SVAR identification, the RR narrative shock series has been shown to be autocorrelated and predictable by common macro/finance variables (Miranda-Agrippino and Ricco 2017). Hence, issues related to information shock contamination likely apply to this specific series, although it passes the exogeneity tests in Podstawski et al. (2018) with monthly time resolution. Turning to findings regarding the correlation among MP proxy shocks, Stock and Watson (2012) document that the SW proxy shock shows only a meaningful correlation with the GSS high-frequency based shock, which we excluded in our analysis due to a lack of relevance (see Appendix D). The other two proxy shocks SZ and RR are strongly related with each other (with linear correlation of 0.93, Stock and Watson 2012). Assuming exogeneity, this indicates that they (consistently) identify the same shock. Unfortunately, given that they show similar mixing characteristics regarding non-MP shocks, it seems more plausible that their close (linear) relation is due to some common contamination with other non-MP disturbances.

Instrument(●)	SW	RR	SZ	SW	RR	SZ
$\hat{\rho}_{\varepsilon_d^{CV}, \varepsilon_{mp}^{(\bullet)}}$	0.341	0.705	0.689	0.353	0.707	0.698
$\hat{\rho}_{\varepsilon_s^{CV}, \varepsilon_{mp}^{(\bullet)}}$	0.018	0.424	0.345	0.002	0.438	0.324
R^2	0.116	0.663	0.579	0.125	0.668	0.575
F -statistic	10.117	149.544	100.890	9.486	133.519	90.038
$\pm 2/\sqrt{T}$	± 0.160	± 0.161	± 0.164	± 0.172		

Table 6: Contemporaneous correlation between MP shocks identified by proxy SVARs and non-MP shocks identified by means of CV. The results in the left- (right-)hand side panel are based on all available observations (synchronized samples of size $T = 135$). The last row documents rule-of-thumb critical values for non-zero correlations with 5% significance. R^2 and F -statistics obtained from regressing the proxy-identified MP shocks on the CV-based non-MP shocks.

Correlation patterns of identified shocks It is tempting to further examine the endogeneity issue by exploiting the information inherent in structural shocks identified in a data-based manner. Summarizing the shocks identified in Stock and Watson (2012) as supply- (e.g. oil

shocks) and demand-type shocks (e.g. liquidity and fiscal shocks) to accommodate for our smaller system, we report correlations between the proxy-identified MP shocks and demand and supply shocks from the covariance change model in Table 6.²⁶ One might expect a-priori that an MP shock - whose characteristics regarding the responses of output and prices renders it a representative of the class of demand shocks - bears a stronger risk of mixing with other demand shocks than with supply shocks. Hence, when assessing the contemporaneous mixing with other shocks, seeing correlations with supply shocks might be a more meaningful indicator of potential endogeneity than correlations with demand shocks. As it turns out, the MP shocks identified via RR and SZ exhibit sizable correlations with both the demand and supply shock identified through (co)variance breaks, whereas the SW-based shock is only (and least strongly) correlated with the CV demand shock.

LR test for exogeneity Podstawski et al. (2018) suggest translating the exogeneity condition into testable over-identifying restrictions within the framework of a pre-specified heteroskedastic SVAR model. Conditional on the (co)variance break model, the respective test outcomes are displayed in Table 7. It appears that all three instruments considered are fairly exogenous. The strongest evidence (i.e. largest p -values) of exogeneity is obtained for SW. However, in light of the documented strong correlations among CV-based non-MP shocks and MP shocks from proxy SVARs, one could raise the concern that the exogeneity test is subject to low power.²⁷

Instrument	SW	RR	SZ
LR statistic	0.356	1.032	1.057
p -value	0.837	0.597	0.589

Table 7: LR tests for exogeneity of the instruments. Identification is conditional on the CV supposed at $T_B^* = 1984Q1$. See Podstawski et al. (2018) for more details.

4.4 Alternative MP shocks

The upper panel of Figure 6 provides time series illustrations of the identified MP shocks. Irrespective of the identification approach adopted, the implied shocks show less variation starting in 1984, which provides further justification to exploit this particular statistical property for purposes of identification. Directly contrasting SW proxy shocks obtained from the trinity VAR and the FAVAR ($r = 3$, see the lower panel of Figure 6) reveals that information enrichment leaves the proxy shocks almost unchanged. Hence, the benchmark MP shocks are unlikely to suffer from a lack of fundamentalness.

In order to further unravel similarities and differences among MP shocks obtained from

²⁶While it is still debated whether (global) oil shocks behave as domestic supply or demand shocks (Baumeister and Kilian 2016), the proxy oil shocks studied in Stock and Watson (2012) feature impulse response patterns that comply with a supply shock characterization.

²⁷Monte Carlo exercises in Podstawski et al. (2018) show that for sample sizes of $T \approx 150$ (as in the present examples), the correct detection of endogeneity might suffer from power weakness.

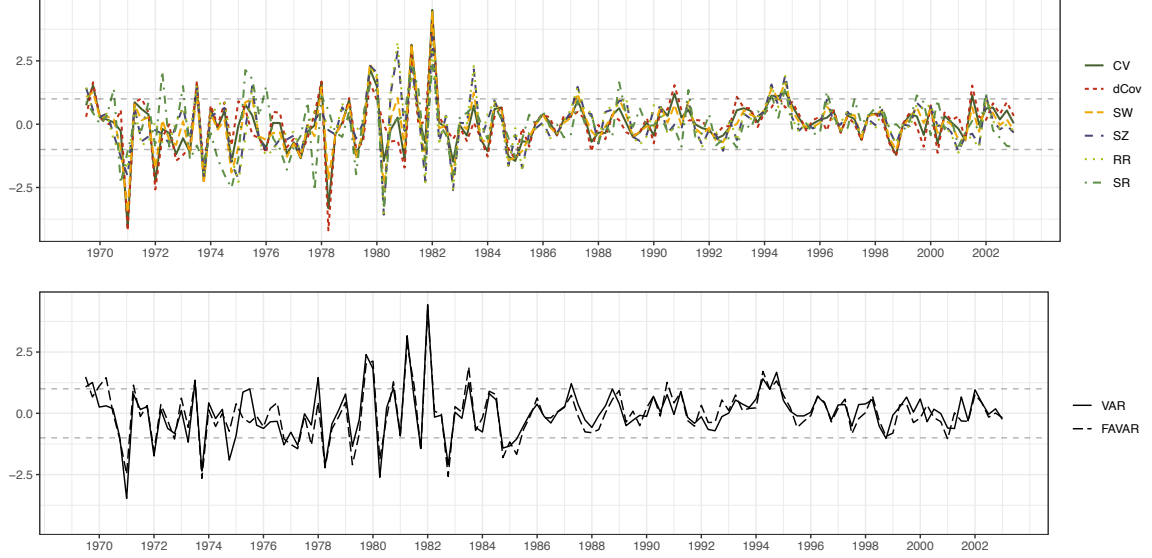


Figure 6: Identified MP shocks. *Upper panel*: All shocks except proxy shocks from FAVAR; *Lower panel*: proxy shocks conditioning on SW from VAR and FAVAR ($r = 3$). Dashed horizontal lines indicate locations of ± 1 . All shock series are standardized to have unit standard deviation.

theory- and data-based identification and proxy SVARs, Table 8 documents respective correlation statistics. Imposing opposite impact responses of interest rates and inflation by means of sign restrictions obtains structural shocks featuring surprisingly weak correlations with all remaining MP shock candidates (the five documented correlations are between 0.2 and 0.69). Raising some criticism about the imposition of (agnostic) sign restrictions, both benchmark MP shocks (CV and SW) exhibit incomplete correlations with sign-restricted shocks, i.e. 0.48 (data-based identification, CV) and 0.69 (proxy SVARs, SW). Moreover, the set-identified MP shocks exhibit a strong positive (moderate negative) correlation with the CV demand (supply) shocks (not documented in Table 8).²⁸ Interestingly, this correlation pattern fits perfectly to the results in Wolf (2020), highlighting the case of ‘masked’ contractionary MP shocks that can be traced back to a weighted linear combination of demand, supply and MP shocks.

While the proxy shocks identified via SZ and RR are closely related with each other (correlation of 0.99), both fail to fully reflect the information inherent in the benchmark MP shocks (correlations with SW are 0.85 for SZ and 0.78 for RR). Proxy MP shocks retrieved from FAVAR models are very closely correlated with their counterparts obtained from the trivariate SVAR. Augmenting the baseline model with three factors (FAVAR, $r = 3$), the correlations between proxy MP shocks from VAR and FAVAR models are 0.898 (SW), 0.918 (RR) and 0.934 (SZ). For the case of our preferred instrumental variable SW, similar cor-

²⁸The correlation coefficients for the median target SR MP shock are: $\text{Cor}(\hat{\varepsilon}_{mp}^{SR}, \hat{\varepsilon}_d^{CV}) = 0.694$, $\text{Cor}(\hat{\varepsilon}_{mp}^{SR}, \hat{\varepsilon}_s^{CV}) = -0.557$, ($2/\sqrt{T} = 0.153$, $R^2 = 0.797$, $F = 331.4$).

relations are obtained when the model augmentation comprises one ($r = 1$, correlation of MP shocks is 0.909) and two factors ($r = 2$, 0.931). Within the three-dimensional and the information-rich model, proxy shocks conditioning on RR and SZ exhibit an almost complete correlation of about 0.99, such that the contamination of both shocks is unlikely due to informational issues. Similarly, correlation results from the large dimensional dynamic factor model of Stock and Watson (2012) indicate that both shocks are subject to common sources of contamination. Specifically, proxy MP shocks conditioning on RR and SZ exhibit a correlation of 0.93 in Stock and Watson (2012).²⁹

While an unconditional diagnostic assessment of alternative MP shocks provides guidance e.g. in terms of statistical model selection, they hold limited value to assess the informational content of the structural findings for the understanding of concrete historic episodes. In the next section, we highlight the informational value of alternative MP shocks (i) to approximate important narratives about MP measures taken in the US, and we shed light on (ii) the monetary policy in the 1970s, and (iii) the macroeconomic forces behind the Volcker disinflation in the early-1980s.

	SR	CV	dCov	SW	SZ	RR	RR*	SZ*
SR	1.000							
CV	0.483	1.000						
dCov	0.201	0.934	1.000					
SW	0.687	0.946	0.769	1.000				
SZ	0.623	0.677	0.396	0.853	1.000			
RR	0.532	0.603	0.326	0.784	0.991	1.000		
SZ*	0.447	0.999	0.937	0.942	0.656	0.578	1.000	
RR*	0.496	0.999	0.946	0.934	0.659	0.587	0.998	1.000

Table 8: Correlations among identified MP shocks. Shock series identified by means of sign restrictions (SR) are calculated based on median target estimates.

5 Narrative evidence and monetary policy in the 1970s and early-1980s

With diagnostic and descriptive results for a variety of MP shocks at hand, this section provides insights into the scope of the respective SVARs to align with decisions of the Fed featuring a (wide) consensus on their intended macroeconomic effects. More specifically, we (i) characterize the monetary policy in the 1970s as a source of the high and persistent inflation, (ii) assess the scope of MP shocks identified in explaining the Volcker disinflation in terms of their marginal contribution to realized reductions of the GDP deflator, (iii) characterize the aggregate demand and supply side of the US economy during the early-1980s and take

²⁹Moreover, undocumented results show that the instruments RR and SZ give rise to marked price puzzles irrespective of the (FA)VAR dimension. Respective IRFs are available from the authors upon request.

full advantage of combining proxy SVAR results with outcomes from identification through covariance change, and (iV) use the complementary approach to provide purged variants of the contaminated MP instruments (RR and SZ) featuring ‘model-implied’ exogeneity.

5.1 Evaluating the shock series against consensual shocks

Although identification through (agnostic) sign restrictions ensures economically meaningful structural features, the set of admitted model structures is often large and might call for a sharpening of model-implied dynamic structures (e.g. IRFs). In this context, Antolín-Díaz and Rubio-Ramírez (2018) have suggested casting implied shocks into the context of consensual narrative information that economists can agree upon. For the purposes of this study, it is worth highlighting that Antolín-Díaz and Rubio-Ramírez (2018) qualify the monetary contraction of October 1979 as ‘the clearest and most uncontroversial example of an MP shock’ in the US. Interestingly, using the SW instrument for proxy SVAR analysis obtains the strongest contractionary signal in 1979Q4 among all identification schemes. Regarding a more comprehensive encountering of four contractionary (April 1974, October 1979, December 1988 and February 1994) and four expansionary monetary signals (December 1990, October 1998, April 2001, and November 2002) in Antolín-Díaz and Rubio-Ramírez (2018), the quarterly shocks displayed in Table 9 show that strongest signals are obtained in three out of eight events when using the benchmark proxy SVAR. In terms of directional accuracy, MP shocks conditioning on the instruments RR and SZ are unique in matching the narrative information for all eight events encountered. However, given the evidence of marked price puzzles ‘identified’ by means of these instruments (see columns 2 and 3 of Figure 5), it is important to recall that these shocks actually affect prices in the wrong direction, i.e. contractionary (expansionary) shocks initiate a price increase (reduction) on impact.

MP shocks	Contractionary				Expansionary				Freq.
	1974Q2	1979Q4	1988Q4	1994Q1	1990Q4	1998Q4	2001Q2	2002Q4	
SR	-1.415	1.612	1.719	0.385	-0.151	0.006	-0.339	-0.895	6
CV	-0.103	1.547	0.448	0.393	0.854	-0.859	-0.457	0.387	5
dCov	-0.065	1.704	0.164	0.464	1.579	-1.291	-0.602	0.893	5
SW	-0.224	2.521	1.042	0.612	0.801	-1.079	-0.653	0.193	5
RR	0.784	2.336	0.806	0.502	-0.142	-0.816	-0.446	-0.006	8
SZ	0.490	2.366	0.934	0.521	-0.018	-0.807	-0.478	-0.051	8
RR*	-0.099	2.260	0.597	0.576	1.306	-1.301	-0.675	0.629	5
SZ*	-0.237	2.210	0.674	0.565	1.258	-1.223	-0.660	0.540	5

Table 9: Comparing MP shocks identified with eight events discussed in the literature (Antolín-Díaz and Rubio-Ramírez 2018). *Bold*: the largest quantification for each event with correct sign. Since underlying events are provided with monthly resolution (contractionary: 1974M4, 1979M10, 1988M12, 1994M2; expansionary: 1990M12, 1998M10, 2001M4, 2002M11), the documented quarterly shocks are subject to temporal aggregation.

5.2 1970s inflation

We start by investigating the scope of MP shocks in explaining the observed high and persistent inflation in the 1970s. Following Barsky and Kilian (2001), MP was characterized by changes of motives and commitment of the federal government, which can be traced back to the *Great Depression* and the belief that a tight MP was responsible for the high unemployment during that episode. In response to the government’s pursuit of full employment (e.g. among others, the Employment Act of 1946), the Fed started to gradually loosen its commitment to price stability and attempted to stimulate employment by exploiting the trade-off between unemployment and inflation, as expressed by the Phillips curve. With the collapse of Bretton Woods, MP lost its anchor (Kilian 2010). Barsky and Kilian (2001) document unusually large monetary expansions for many OECD countries during the 1970s. Without traditional constraints, MP was conducted more experimentally, while the public was only slowly beginning to understand the changes in MP, i.e. hesistantly adjusted (long-term) inflation expectations, which translated to sluggishness in inflation. Barsky and Kilian (2001) show that if inflation is inherently sluggish, an unexpected expansionary policy will temporally stimulate output. Inflation initially increases slowly but continues to rise after output has reached its maximum. Consequently, a stagflation occurs.

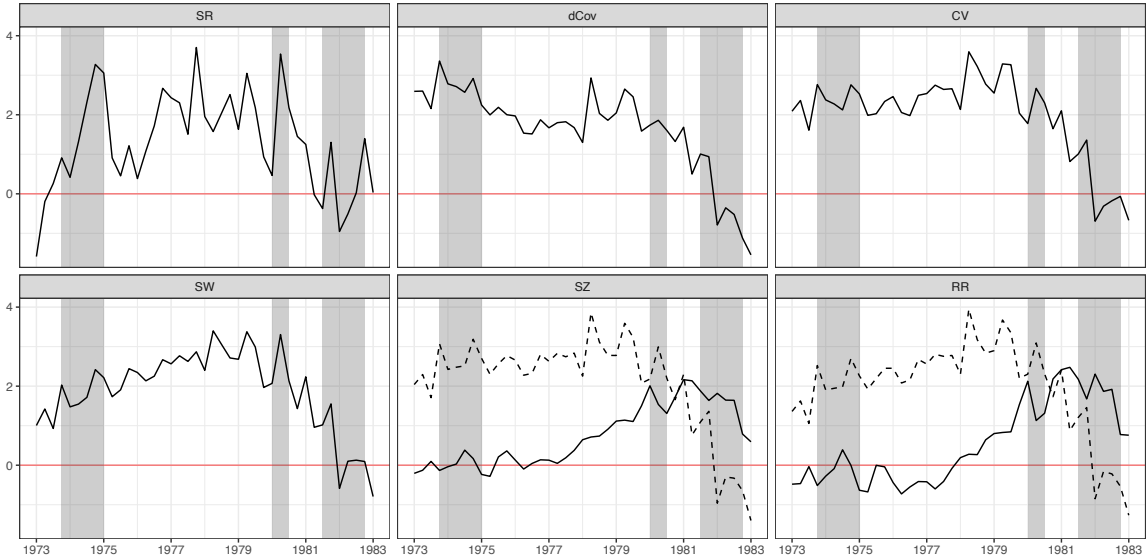


Figure 7: Cumulative effects of MP shock on the inflation in the GDP deflator over time (selected period). For a textbook definition of historical decompositions, we refer the reader to Chapter 4.3 of Kilian and Lütkepohl (2017). The annualized average change of the deflator amounts to 7.46 during this period. Shaded areas represent episodes of economic recession as defined by the National Bureau of Economic Research (NBER). Dashed lines indicate results based on ‘purged’ instruments.

In 1973, the Fed attempted to bring inflation under control by tightening its policy. However, at end of the year, when facing an energy shortage and dramatically increased costs for

industrial raw materials such as crude oil, the central bank had to ease its policy to mitigate the negative effects of the recession prior to the full containment of inflation.³⁰ In April 1974, shortly after the OPEC embargo was lifted, the Fed aggressively raised the FFR to 11% and held it between 10% and 13% throughout the recession. Meanwhile, inflation had reached 12% according to our measure. Figure 7 displays the contribution of MP shocks identified to the evolution of the GDP deflator during the selected period from the early-1970s until the 1980s. As the MP exhibited a go-and-stop pattern, the Fed found it difficult to credibly lower inflation expectations, and persistently high inflation expectations raised the costs of potential disinflation. Structural models recovered by SR, dCov, CV and SW indicate that the MP did not effectively bring down the sustainably high inflation in the 1970s and early-1980s. In fact, MP shocks identified by these approaches were persistent sources of inflation during that time, until the anchor was eventually reestablished in 1982 as a result of Volcker’s policy. Specifically, the benchmark models (i.e. covariance change model and proxy SVAR conditioning on SW) imply that during the episode from 1973 to 1979 the inflationary pressure from MP was on average between 2.2 to 2.4 units. By contrast, structural results obtained by both the SZ and RR instruments imply that MP played a minor role in the *Great Inflation*. Interestingly, MP shocks identified by these instruments became slowly inflationary after Paul Volcker took office and announced the fight against inflation in late 1979. We address this issue in the next subsection.

5.3 Volcker disinflation

Volcker disinflation was initiated by the MP shock issued late in 1979. When Paul Volcker became Chairman of the Fed in August 1979, inflation had reached two-digit levels. In October 1979, after an unscheduled FOMC meeting, Volcker announced his intention to bring inflation quickly under control and focus on non-borrowed reserve targeting (for a discussion of implicit policy instrument regimes of the Fed, see e.g., Thornton 2006). Characterizing this exceptional period of US MP history, the FFR approached $\approx 20\%$ in June 1981, which sent the economy into a recession in 1981Q3. The Fed did not ease the tight money supply, despite enormous political pressures. Arguably as a result of Volcker’s credible contractionary MP, inflation rates as analyzed in this study plummeted from 8.22% in 1979Q4 to 5.88% (4.25%)

³⁰ The oil price increased by a factor of four in January 1974, after Arab OPEC countries installed an embargo on the US and started a series of production cuts in October 1973. Providing a different perspective, Baumeister and Kilian (2016) interpret the decline in Arab oil production and hence the oil price spike as endogenous responses to global macroeconomic conditions - among others - to the accelerated (consumption or flow) demand for oil reflecting a worldwide economic expansion since early 1972. According to these authors, the demand channel of oil price shocks also played an essential role for the next oil crisis in 1979, which was traditionally considered as being caused by distortions of oil production in Iran following the Iranian Revolution. Baumeister and Kilian (2016) and Kilian and Murphy (2014) argue that the revolution has caused an upward shift in the (speculative) inventory demand reflecting that the agents had anticipated future oil shortages, rather than a shock to the (flow) supply of oil, since Barsky and Kilian (2001) find little evidence of an overall decrease of OPEC oil output. Against this perspective, Baumeister and Hamilton (2019) find that speculation and inventory demand shocks have played a much lesser role in the oil price fluctuations.

in 1982Q3 (1983Q3). Hence, it seems natural to address the fight against inflation to a large extent to (contractionary) MP measures. The bars with solid lines in Figure 8 indicate the cumulated inflation responses to model-implied MP shocks issued during both reference periods.³¹

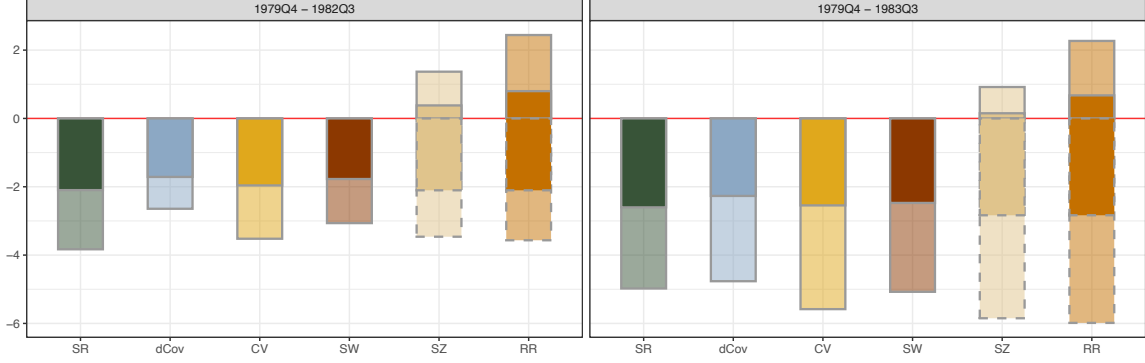


Figure 8: Contribution of MP shocks issued during the disinflation periods to (dis)inflation (see footnote 31). Conditional on our data, inflation decreases during the 1979Q4-1982Q3 period (1979Q4-1983Q3) by 2.34 % (3.97 %). Dashed lines indicate results based on ‘purged’ instruments (i.e. the residuals obtained after regressing original instruments on aggregate demand and supply shocks as identified by means of the covariance change model). The area with low transparency indicates the disinflationary effect attributed to the particular shock occurring in 1979Q4.

As it emerges for both reference periods, joint with median target shocks identified by means of sign restrictions and results from ICA identification, the benchmark shocks retrieved from CV or proxy SVARs conditional on SW imply the largest contributions to the inflation decline. For instance, proxy shocks conditioning on SW triggered an inflation decline of -3.06 units of the GDP deflator during 1979Q4 and 1982Q3. It is interesting to note that the aggregate effect realized within the reference period can be attributed to a large extent (i.e. to about 60%) to the shock issued in 1979Q4. Putting the marginal effect on (aggregated) deflator change of -3.06 into perspective, it is worth commenting on effects of MP shocks during other periods of equivalent length. Specifically for the twelve quarters from 1976Q4 until 1979Q3 (1982Q4 until 1985Q3), corresponding effects amount to deflator changes of +2.72

³¹Let $\hat{\varepsilon}_{3,t} = \hat{\varepsilon}_{mp,t}$ denote identified MP shocks within a time period of interest, $T_1 \leq t \leq T_2$, and $\hat{\theta}_{k3,i}$ the estimated structural responses showing up in model variable k at horizon i . Then, the cumulated effects of these shocks on the k -th variable are given by

$$\Xi_k(T_1, T_2) = \sum_{t=T_1}^{T_2} \left(\hat{\varepsilon}_{3,t} \sum_{i=0}^{T_2-t} \hat{\theta}_{k3,i} \right). \quad (27)$$

Outcomes $\Xi_2(T_1, T_2)$ for all identification schemes are displayed in Figure 8. Owing to the definition of y_{2t} as quarter-on-quarter deflator changes, $\Xi_2(T_1, T_2)$ is formally an integrated contribution to the change of the deflator within the reference period. Economically, we refer to this quantity as a ‘contribution to (dis)inflation’.

(+1.72) units. Hence, interpreting the period of disinflation as having laid the ground for the *Great Moderation*, our results point to the effectiveness of the new MP strategy and thereby support the so-called ‘good policy’ hypothesis (along with further references, see Giannone et al. 2008). While these reference statistics underpin the marked policy change initialized by the switch to monetary tightening in 1979Q4, it holds further interest to explore which non-MP effects might have shaped inflation during the period of disinflation policies. In this regard, we obtain from the CV model a prime role of supply-sided effects. While the CV model obtains a similar contribution of MP shocks to disinflation (see Figure 8), cumulative inflationary effects associated with identified demand and supply shocks amount to deflator units of 1.18 and 3.56 (1.21 and 3.99) for the period 1979Q4 until 1982Q3 (1979Q4 until 1983Q3), respectively. Hence, in managing the disinflation, the Fed had to keep more of an eye on supply-sided factors rather than effects originating on the demand side of the economy.³²

Unlike the benchmark models, proxy SVARs conditioning on instruments RR or SZ deliver structural shocks and biased IRFs that jointly imply expansionary total impacts on inflation for both reference periods. Conditional on these instruments, proxy shocks induced an increase of the GDP deflator by 1.37 (SZ) and 2.44 units (RR) during the period from 1979Q4 until 1982Q3. Intuitively, this failure to accord with the narrative of regime change in the Fed’s monetary strategy mirrors the marked price puzzles obtained from conditioning on these instruments (see columns 2 and 3 of Figure 5).

5.4 Disentangling proxy MP and non-MP shocks

The complex overall economic situation in the US in the early-1980s places identifying genuine monetary measures by means of proxy SVARs at particular risks of instrument contamination. After conditioning on proxies RR and SZ, the historical decomposition based on the identified structural models fails to explain the inflationary MP in the 1970s or provide a sound directional effect of disinflationary MP under the Volcker regime. Moreover, the estimated MP shocks lack a theory-conform economic label. Unlike the structural IRFs from conditioning on SW, their counterparts based on RR and SZ are likely biased (e.g. giving rise to a price puzzle). Nevertheless, it is interesting to uncover the extent to which demand and supply shocks contribute to the proxy MP shocks conditioning on RR and SZ. In the following we address this issue from two perspectives. On the one hand, we employ the SW-identified MP shocks and the respective structural MA representation to decompose the detected contribution to disinflation regarding the endogenous components of proxy MP shocks obtained from conditioning on SZ or RR. On the other hand, we use the aggregate demand and supply shocks identified through covariance change for adjusting the instruments RR and SZ.

³²Regarding the effects on output, MP shocks were the main force that pushed the US economy into recession in the early-1980s. Tracing the cumulated effects of MP shocks on the output gap ($\Xi_1(T_1, T_2)$) within the period 1979Q4 until 1982Q3 amounted to -13.02 and -19.75 according to the benchmark proxy SVAR and the covariance change (CV) model, respectively. By contrast, the effects of supply shocks were only -7.85, according to the covariance change model.

Instrument contamination and non-MP shocks In order to decompose the contribution of SW-based MP shocks, we make use of hierarchical least squares projections of the following type:

$$\hat{\varepsilon}_{mp,t}^{SW} = \hat{a}\hat{\varepsilon}_{mp,t}^{\bullet} + \hat{\zeta}_t^{\bullet} \quad (28)$$

$$\hat{\varepsilon}_{mp,t}^{\bullet} = \hat{b}\hat{\varepsilon}_{d,t}^{CV} + \hat{c}\hat{\varepsilon}_{s,t}^{CV} + \hat{\xi}_t^{\bullet}, \bullet \in \{RR, SZ\}. \quad (29)$$

From these regressions, we obtain specifically (with standard errors in parentheses):

$$\begin{aligned} \hat{\varepsilon}_{mp,t}^{SW} &= \underset{(0.053)}{0.779}\hat{\varepsilon}_{mp,t}^{RR} + \hat{\zeta}_t^{RR} \text{ and } \hat{\varepsilon}_{mp,t}^{RR} = \underset{(0.067)}{0.918}\hat{\varepsilon}_{d,t}^{CV} + \underset{(0.071)}{0.578}\hat{\varepsilon}_{s,t}^{CV} + \hat{\xi}_t^{RR}, \\ \hat{\varepsilon}_{mp,t}^{SW} &= \underset{(0.046)}{0.876}\hat{\varepsilon}_{mp,t}^{SZ} + \hat{\zeta}_t^{SZ} \text{ and } \hat{\varepsilon}_{mp,t}^{SZ} = \underset{(0.073)}{0.883}\hat{\varepsilon}_{d,t}^{CV} + \underset{(0.077)}{0.404}\hat{\varepsilon}_{s,t}^{CV} + \hat{\xi}_t^{SZ}. \end{aligned}$$

The regression hierarchy in (28) and (29) allows for a decomposition of the contribution of SW proxy shocks (see footnote 31) to overall disinflation in a reference period $([T_1, T_2])$ into the four components encountered in Table 10. For instance, the residual estimates from (29) quantify the component of the contaminated MP shocks (conditioning on RR and SZ) that show a positive correlation with benchmark MP shocks and cannot be traced back to correlation with aggregate supply or demand shocks. Unsurprisingly, for both instruments these components contribute sizeably (i.e. more than one-third) to overall disinflation achieved by the benchmark MP measures. However, the components of the contaminated shocks that correlate with non-MP shocks also contribute markedly to explained disinflation. In this regard, it is interesting to note that the aggregate demand shocks issued during the periods under scrutiny account for a much larger contribution to disinflation (i.e. at least by a factor of 1.5) in comparison with aggregate supply shocks. Seeing that the coefficient estimates \hat{b} and \hat{c} in (29) are throughout positive, MP shocks obtained from conditioning on RR or SZ might be - to some extent - interpreted as masked demand or supply shocks. For the practical MP conduct, the hierarchical regressions offer an important insight into the (dis)functioning of contaminated proxy shocks. While it is intuitive to expect that under instrument (or shock) contamination the actual disinflation can hardly be traced back to clear-cut monetary surprises, the decompositions in Table 10 highlight the two components C_2 and C_3 (and their relative importance) for which it is not possible to distinguish whether their disinflationary contribution originates in MP measures or non-MP shocks.

Purging instruments Since the covariance change model and the proxy SVAR conditioning on SW obtain closely-related MP shocks, it is promising to employ the information on exogenous demand and supply shocks from the former model $(\varepsilon_{d,t}^{CV}, \varepsilon_{s,t}^{CV})$ to improve the RR and SZ instruments as cleaned measures of the MP shocks (see also the Bridge Proxy SVAR model of Gazzani and Vicendoa (2019) for an attempt to complement proxy SVARs with data-based, i.e. ICA, identification). In order to achieve ‘model-implied exogeneity’, we project the instrument series RR and SZ on the non-MP shocks detected through covariance change and consider the respective residuals as new instruments to which we refer as RR*

		1979Q4 - 1982Q3		1979Q4 - 1983Q3	
Component	Definition	RR	SZ	RR	SZ
C_1	$\sum_{t=T_1}^{T_2} \left(\hat{\zeta}_t \sum_{i=0}^{T_2-t} \hat{\theta}_{23,i}^{SW} \right)$	-0.259	-0.047	-0.725	-0.341
C_2	$\hat{a}\hat{b} \sum_{t=T_1}^{T_2} \left(\hat{\varepsilon}_{d,t}^{CV} \sum_{i=0}^{T_2-t} \hat{\theta}_{23,i}^{SW} \right)$	-0.972	-1.051	-1.422	-1.537
C_3	$\hat{a}\hat{c} \sum_{t=T_1}^{T_2} \left(\hat{\varepsilon}_{s,t}^{CV} \sum_{i=0}^{T_2-t} \hat{\theta}_{23,i}^{SW} \right)$	-0.647	-0.508	-0.907	-0.713
C_4	$\hat{a} \sum_{t=T_1}^{T_2} \left(\hat{\xi}_t \sum_{i=0}^{T_2-t} \hat{\theta}_{23,i}^{SW} \right)$	-1.186	-1.458	-2.020	-2.483
Σ		-3.064	-3.064	-5.074	-5.074

Table 10: Component specific contributions of SW-identified MP shocks to the disinflation during the reference periods. Components are obtained from hierarchical regressions in (28) and (29). See also the bars in Figure 8 for the total effects (Σ) and their definition in footnote 31. MA coefficients $\theta_{23,i}^{SW}$ are taken from the proxy SVAR conditioning on SW.

and SZ^* . Respective regression results are displayed in Table 11. As detailed in the notes of the table, both purged instruments remain relevant. As it emerges, RR^* and SZ^* seem to closely approach the benchmark instruments SW. Specifically, the implied proxy MP shocks exhibit a strong correlation in excess of 0.93 (without purging the respective correlations are at most 0.85, see Table 8). Historical decompositions based on models identified by these new shock measures accord with typical notions that inflationary pressures in the 1970s were primarily caused by MP conduct (see Figure 7). From the right-most bars in both panels of Figure 8, we also observe that proxy MP shocks obtained from the adjusted instruments imply theory-conform contributions to the Volcker disinflation that are similar in size to those retrieved from the benchmark instrument SW.

	RR			SZ		
	Estimate	Std. Error	p -value	Estimate	Std. Error	p -value
$\hat{\varepsilon}_d^{CV}$	0.335	0.054	0.000	1.442	0.226	0.000
$\hat{\varepsilon}_s^{CV}$	0.201	0.055	0.000	0.717	0.237	0.003

Table 11: Orthogonalizing instruments RR and SZ with respect to benchmark demand and supply shocks (from CV). F -statistics for tests of weak instrumentation are 10.93 and 16.58 for RR^* and SZ^* , respectively. The critical value at which one can reject the null hypothesis with threshold signal-to-noise ratios that correspond to a 10 % asymptotic bias with 5% (10%) significance is 7.38 (8.53) for $K = 3$ (Lunsford 2015).

6 Conclusion

Being among the most prominent fields in macroeconometric analysis, the identification of US MP shocks has taken full account of rich methodological advances in the identification of structural VAR models (SVARs). In this study, we follow similar lines as Mumtaz et al.

(2018) in their analysis of credit supply shocks. Specifically, we combine insights from alternative identification schemes applied to log-linearized DSGE models on the one hand, as well as empirical systems of US data covering the period 1965Q1 until 2008Q3 on the other. The identification schemes considered comprise identification by means of (agnostic) sign restrictions, identification through heteroskedasticity or ICA and proxy SVARs. The simulated processes are characterized by reduced-form (co)variance patterns that align with stylized facts invoked by both the *Great Moderation* and recent findings that the detection of US MP shocks generally suffers from low relative signal strength (Paustian 2007, Wolf 2020). Extending beyond the analysis in Mumtaz et al. (2018), we report on potential pitfalls of proxy SVARs that might occur when using weak or contaminated instruments. Our Monte Carlo results highlight that the case of contaminated instrumentation bears risks of biased identification outcomes. Denoting prime relevance of these insights for the empirical analysis of MP in the US, proxy SVARs could fail to detect unique MP shocks in realistic scenarios of instrument endogeneity coupled with relatively low signal strength.

Our empirical analysis of quarterly US data largely conditions on a stylized trinity VAR comprising the output gap, inflation and the one-year government bond rate. In order to address a potential lack of fundamentalness, we underpin the small-scale empirical models with results from an information-rich factor-augmented model that processes information drawn from Michael McCracken’s database (FRED-QD). Along similar lines as in Mumtaz et al. (2018), we evaluate a rich set of structural models (agnostic sign restrictions, data-based SVARs, proxy SVARs). For the construction of instrumental information, we find that many promising instrument candidates building on futures-based high-frequency information drop out as they provide only weak external information (Miranda-Agrippino and Rey 2018). Unlike their high-frequency counterparts, established low-frequency instruments are strong (i.e. relevant) but subject to endogenous contamination. All structural results point to a marked downward shift of the volatility of structural shocks starting in 1984. Hence, the explicit consideration of covariance shifts is natural for implementing SVAR identification in a data-based manner. The corresponding MP shocks show almost perfect correlations with proxy SVAR results based on instrumental information that has been retrieved from the medium-scale DSGE model of Smets and Wouters (2007). MP shocks extracted from this instrument fully accord with a sound economic labeling. Moreover, they are almost orthogonal to aggregate demand and supply shocks identified by means of covariance shifts, and thereby promise the best potential to cope with the relatively low signal strength of MP shocks. Overall, these considerations affirm related results in Stock and Watson (2012) and lend credibility to the SW instrument. Further underpinning their benchmark property in a context of narrative analysis, proxy shocks conditioning on SW quantify effectively inflationary MP in the 1970s, and contractionary policy signals issued during reference periods of the Volcker disinflation in the early-1980s (Barsky and Kilian 2001, Antolín-Díaz and Rubio-Ramírez 2018, Bernanke and Mihov 1998). For alternative low-frequency instruments (i.e. the popular Romer and Romer (2004) narrative MP shock series (RR), and the best-fitting SVAR shock from Sims

and Zha (2006) (SZ)), our analysis indicates marked price puzzles and in particular a failure to explain both the *Great Inflation* and the Volcker disinflation. While this note of caution seems important for policy analysis conditioning e.g. on the RR narratives (Miranda-Agrippino and Rey 2018, Coibion et al. 2017), the identification through CV pursued in this work provides (i) valuable guidance for tracing the endogenous contamination of RR (and SZ), and (ii) a constructive way towards an efficient purging of these instruments (see also Gazzani and Vicondoa (2019) for the Bridge Proxy SVAR model, which also combines data-based identification and proxy SVARs). Both merits are likely to further materialize, in particular in those fields of MP analysis that are limited to low-frequency empirical analysis for reasons of data availability (e.g., MP and credit, or MP and inequality). Moreover, after establishing ‘model-implied’ exogeneity, the purged variants of RR and SZ (as well as benchmark instruments) allow a direct measurement of MP shocks in higher-dimensional systems or empirical models where core preconditions for data-based identification (covariance change, non-Gaussianity) do not apply.

References

- AHMADI, P. A. AND H. UHLIG (2015): “Sign restrictions in Bayesian FAVARs with an application to monetary policy shocks,” NBER Working Papers 21738, National Bureau of Economic Research, Inc.
- ANTOLÍN-DÍAZ, J. AND J. F. RUBIO-RAMÍREZ (2018): “Narrative sign restrictions for SVARs,” *American Economic Review*, 108, 2802–29.
- AUERBACH, A. J. AND Y. GORODNICHENKO (2012): “Measuring the output responses to fiscal policy,” *American Economic Journal: Economic Policy*, 4, 1–27.
- BACCHIOCCHI, E. AND L. FANELLI (2015): “Identification in structural vector autoregressive models with structural changes, with an application to us monetary policy,” *Oxford Bulletin of Economics and Statistics*, 77, 761–779.
- BARAKCHIAN, S. M. AND C. CROWE (2013): “Monetary policy matters: Evidence from new shocks data,” *Journal of Monetary Economics*, 60, 950 – 966.
- BARSKY, R. B. AND L. KILIAN (2001): “Do we really know that oil caused the great stagflation? A monetary alternative,” NBER Working Papers 8389, National Bureau of Economic Research, Inc.
- BAUMEISTER, C. AND J. D. HAMILTON (2019): “Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks,” *American Economic Review*, 109, 1873–1910.
- BAUMEISTER, C. AND L. KILIAN (2016): “Forty years of oil price fluctuations: Why the price of oil may still surprise us,” *Journal of Economic Perspectives*, 30, 139–60.
- BERNANKE, B. AND A. BLINDER (1992): “The federal funds rate and the channels of monetary transmission,” *American Economic Review*, 82, 901–21.
- BERNANKE, B. S., J. BOIVIN, AND P. ELIASZ (2005): “Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach,” *The Quarterly Journal of Economics*, 120, 387–422.
- BERNANKE, B. S. AND I. MIHOV (1998): “Measuring monetary policy,” *The Quarterly Journal of Economics*, 113, 869–902.
- BLANCHARD, O. J. AND D. QUAH (1989): “The dynamic effects of aggregate demand and supply disturbances,” *American Economic Review*, 79, 655–673.
- CALDARA, D. AND E. HERBST (2019): “Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs,” *American Economic Journal: Macroeconomics*, 11, 157–92.

- CANOVA, F. AND G. DE NICOLO (2002): “Monetary disturbances matter for business fluctuations in the G-7,” *Journal of Monetary Economics*, 49, 1131–1159.
- CANOVA, F. AND M. PAUSTIAN (2011): “Business cycle measurement with some theory,” *Journal of Monetary Economics*, 58, 345–361.
- CARLSTROM, C., T. FUERST, AND M. PAUSTIAN (2009): “Monetary policy shocks, Choleski identification, and DNK models,” *Journal of Monetary Economics*, 56, 1014–1021.
- CARRIERO, A., H. MUMTAZ, K. THEODORIDIS, AND A. THEOPHILOPOULOU (2015): “The impact of uncertainty shocks under measurement error: A proxy SVAR approach,” *Journal of Money, Credit and Banking*, 47, 1223–1238.
- CASTELNUOVO, E. (2012): “Monetary policy neutrality: Sign restrictions go to Monte Carlo,” "Marco Fanno" Working Papers 0151, Dipartimento di Scienze Economiche "Marco Fanno".
- CESA-BIANCHI, A., L. F. CESPEDES, AND A. REBUCCI (2015): “Global liquidity, house prices, and the macroeconomy: Evidence from advanced and emerging economies,” *Journal of Money, Credit and Banking*, 47, 301–335.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1999): “Chapter 2 Monetary policy shocks: What have we learned and to what end?” Elsevier, vol. 1 of *Handbook of Macroeconomics*, 65 – 148.
- COIBION, O., Y. GORODNICHENKO, L. KUENG, AND J. SILVIA (2017): “Innocent bystanders? Monetary policy and inequality,” *Journal of Monetary Economics*, 88, 70–89.
- COMON, P. (1994): “Independent component analysis, A new concept?” *Signal Processing*, 36, 287 – 314.
- FAUST, J. (1998): “The robustness of identified VAR conclusions about money,” *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.
- FRY, R. AND A. PAGAN (2007): “Some issues in using sign restrictions for identifying structural VARs,” NCER Working Paper Series 14, National Centre for Econometric Research.
- GAZZANI, A. G. AND A. VICONDOA (2019): “Proxy-SVAR as a bridge for identification with higher frequency data,” 2019 Meeting Papers 855, Society for Economic Dynamics.
- GERTLER, M. AND P. KARADI (2015): “Monetary policy surprises, credit costs, and economic activity,” *American Economic Journal: Macroeconomics*, 7, 44–76.
- GIANNONE, D., L. REICHLIN, AND M. LENZA (2008): “Explaining the Great Moderation: It is not the shocks,” *Journal of the European Economic Association*, 6, 621–633.

- GOURIÉROUX, C., A. MONFORT, AND J.-P. RENNE (2017): “Statistical inference for independent component analysis: Application to structural VAR models,” *Journal of Econometrics*, 196, 111–126.
- GÜRKAYNAK, R. S., B. SACK, AND E. SWANSON (2005): “Do actions speak louder than words? The response of asset prices to monetary policy actions and statements,” *International Journal of Central Banking*, 1.
- HAMIDI SAHNEH, M. (2016): “Testing for non-fundamentalness,” MPRA Paper 71924, University Library of Munich, Germany.
- HERBST, E. AND F. SCHORFHEIDE (2012): “Evaluating DSGE model forecasts of comovements,” *Journal of Econometrics*, 171, 152–166.
- HERWARTZ, H. (2018): “Hodges Lehmann detection of structural shocks - An analysis of macroeconomic dynamics in the Euro Area,” *Oxford Bulletin of Economics and Statistics*, 80, 736–754.
- HERWARTZ, H., A. LANGE, AND S. MAXAND (2019): “Statistical identification in SVARs - Monte Carlo experiments and a comparative assessment of the role of economic uncertainties for the US business cycle,” CeGe Discussion Papers 375.
- HERWARTZ, H. AND M. PLÖDT (2016): “Simulation evidence on theory-based and statistical identification under volatility breaks,” *Oxford Bulletin of Economics and Statistics*, 78, 94–112.
- JAROCIŃSKI, M. AND P. KARADI (2020): “Deconstructing monetary policy surprises—the role of information shocks,” *American Economic Journal: Macroeconomics*, 12, 1–43.
- JENTSCH, C. AND K. G. LUNSFORD (2019): “The dynamic effects of personal and corporate income tax changes in the United States: Comment,” *American Economic Review*, 109, 2655–78.
- KILIAN, L. (2009): “Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market,” *American Economic Review*, 99, 1053–1069.
- (2010): “Oil price shocks, monetary policy and stagflation,” in *Inflation in an Era of Relative Price Shocks*, ed. by R. Fry, C. Jones, and C. Kent, Reserve Bank of Australia, RBA Annual Conference Volume.
- KILIAN, L. AND H. LÜTKEPOHL (2017): *Structural Vector Autoregressive Analysis*, Cambridge University Press.
- KILIAN, L. AND D. P. MURPHY (2014): “The role of inventories and speculative trading in the global market for crude oil,” *Journal of Applied Econometrics*, 29, 454–478.

- LANGE, A., B. DALHEIMER, H. HERWARTZ, S. MAXAND, AND H. RIEBL (2020): *svars: data-driven identification of svar models*, R package version 1.3.3.
- LANNE, M. AND H. LÜTKEPOHL (2008): “Identifying monetary policy shocks via changes in volatility,” *Journal of Money, Credit and Banking*, 40, 1131–1149.
- (2010): “Structural vector autoregressions with nonnormal residuals,” *Journal of Business & Economic Statistics*, 28, 159–168.
- LANNE, M., H. LÜTKEPOHL, AND K. MACIEJOWSKA (2010): “Structural vector autoregressions with Markov switching,” *Journal of Economic Dynamics and Control*, 34, 121–131.
- LANNE, M., M. MEITZ, AND P. SAIKKONEN (2017): “Identification and estimation of non-Gaussian structural vector autoregressions,” *Journal of Econometrics*, 196, 288–304.
- LANNE, M. AND P. SAIKKONEN (2013): “Noncausal vector autoregression,” *Econometric Theory*, 29, 447–481.
- LUNSFORD, K. G. (2015): “Identifying structural VARs with a proxy variable and a test for a weak proxy,” Working Paper 1528, Federal Reserve Bank of Cleveland.
- LÜTKEPOHL, H. AND M. KRÄTZIG (2004): *Applied Time Series Econometrics*, Cambridge University Press.
- LÜTKEPOHL, H. AND T. SCHLAAK (2018): “Choosing between different time-varying volatility models for structural vector autoregressive analysis,” *Oxford Bulletin of Economics and Statistics*, 80, 715–735.
- MATTESON, D. S. AND R. S. TSAY (2017): “Independent component analysis via distance covariance,” *Journal of the American Statistical Association*, 112, 623–637.
- MCCRACKEN, M. AND S. NG (2020): “FRED-QD: A quarterly database for macroeconomic research,” Working Paper 26872, National Bureau of Economic Research.
- MERTENS, K. AND M. O. RAVN (2013): “The dynamic effects of personal and corporate income tax changes in the United States,” *American Economic Review*, 103, 1212–1247.
- MIAN, A., A. SUFI, AND E. VERNER (2017): “Household debt and business cycles worldwide,” *The Quarterly Journal of Economics*, 132, 1755–1817.
- MIRANDA-AGRIPPINO, S. (2016): “Unsurprising shocks: Information, premia, and the monetary transmission,” LSE Research Online Documents on Economics 86234, London School of Economics and Political Science, LSE Library.
- MIRANDA-AGRIPPINO, S. AND H. REY (2018): “US monetary policy and the global financial cycle,” Tech. rep., National Bureau of Economic Research.

- MIRANDA-AGRIPPINO, S. AND G. RICCO (2017): “The transmission of monetary policy shocks,” Bank of England working papers 657, Bank of England.
- MONTIEL OLEA, J. L., J. H. STOCK, AND M. W. WATSON (2018): “Inference in structural vector autoregressions identified with an external instrument,” Working paper, Columbia University.
- MUMTAZ, H., G. PINTER, AND K. THEODORIDIS (2018): “What do VARs tell us about the impact of a credit supply shock?” *International Economic Review*, 59, 625–646.
- NOH, E. (2018): “Impulse-response analysis with proxy variables,” Manuscript, University of California San Diego.
- NORDHAUSEN, K., H. OJA, D. TYLER, AND J. VIRTÄ (2017a): “Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace,” *IEEE Signal Processing Letters*, PP.
- (2017b): *ICtest: estimating and testing the number of interesting components in linear dimension reduction*, R package version 0.3-2.
- PAUSTIAN, M. (2007): “Assessing sign restrictions,” *The B.E. Journal of Macroeconomics*, 7, 1–33.
- PEREZ-QUIROS, G. AND M. M. MCCONNELL (2000): “Output fluctuations in the United States: What has changed since the early 1980’s?” *American Economic Review*, 90, 1464–1476.
- PLAGBORG-MØLLER, M. AND C. K. WOLF (2019): “Instrumental variable identification of dynamic variance decompositions,” Manuscript, Princeton University.
- PODSTAWSKI, M., T. SCHLAAK, AND M. RIETH (2018): “Monetary policy, external instruments and heteroskedasticity,” Discussion Papers of DIW Berlin 1749, DIW Berlin, German Institute for Economic Research.
- RAMEY, V. A. (2016): “Macroeconomic shocks and their propagation,” Working Paper 21978, National Bureau of Economic Research.
- RAVENNA, F. (2007): “Vector autoregressions and reduced form representations of DSGE models,” *Journal of Monetary Economics*, 54, 2048–2064.
- RIGOBON, R. (2003): “Identification through heteroskedasticity,” *The Review of Economics and Statistics*, 85, 777–792.
- RISK, B. B., N. A. JAMES, AND D. S. MATTESON (2015): *steadyICA: ICA and tests of independence via multivariate distance covariance*, R package version 1.0.
- ROMER, C. D. AND D. H. ROMER (2004): “A new measure of monetary shocks: Derivation and implications,” *American Economic Review*, 94, 1055–1084.

- RUSNAK, M., T. HAVRANEK, AND R. HORVATH (2013): “How to solve the price puzzle? A meta-analysis,” *Journal of Money, Credit and Banking*, 45, 37–70.
- SIMS, C. A. (1980): “Macroeconomics and reality,” *Econometrica*, 48, 1–48.
- SIMS, C. A. AND T. ZHA (2006): “Were there regime switches in U.S. monetary policy?” *American Economic Review*, 96, 54–81.
- SMETS, F. AND R. WOUTERS (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *American Economic Review*, 97, 586–606.
- STOCK, J. H. AND M. W. WATSON (2012): “Disentangling the channels of the 2007-2009 recession,” Working Paper 18094, National Bureau of Economic Research.
- (2018): “Identification and estimation of dynamic causal effects in macroeconomics using external instruments,” *The Economic Journal*, 128, 917–948.
- THORNTON, D. L. (2006): “When did the FOMC begin targeting the federal funds rate? What the verbatim transcripts tell us,” *Journal of Money, Credit and Banking*, 2039–2071.
- UHLIG, H. (2005): “What are the effects of monetary policy on output? Results from an agnostic identification procedure,” *Journal of Monetary Economics*, 52, 381–419.
- WOLF, C. K. (2020): “SVAR (mis-)identification and the real effects of monetary policy shocks,” *American Economic Journal: Macroeconomics*, forthcoming.

A Theory- and data-based identification

A.1 Identification by means of sign restrictions

Sign restrictions ground in economic theory and are often considered to be relatively weak (and more consensual) in comparison with exclusion restrictions. Following Faust (1998), Canova and De Nicolo (2002), the approach has been widely applied for studying effectiveness of MP using e.g. the fact that an unanticipated monetary tightening triggers a simultaneous increase in interest rates and a decrease in inflation according to stylized New Keynesian DSGE models. At the implementation side, the imposition of sign restrictions proceeds by sampling from a structured space of reduced-form covariance decompositions:

$$\Sigma_u = CQQ'C' = BB, \quad (30)$$

where C is a benchmark decomposition matrix (e.g. a Cholesky factor), and Q can be any normalized rotation matrix satisfying $QQ' = I_K$. For random sampling of rotation matrices Q we employ a set of the orthonormal Givens rotation matrices. For instance, in the trivariate case, one has:

$$Q(\vartheta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\vartheta_1) & -\sin(\vartheta_1) \\ 0 & \sin(\vartheta_1) & \cos(\vartheta_1) \end{bmatrix} \begin{bmatrix} \cos(\vartheta_2) & 0 & -\sin(\vartheta_2) \\ 0 & 1 & 0 \\ \sin(\vartheta_2) & 0 & \cos(\vartheta_2) \end{bmatrix} \begin{bmatrix} \cos(\vartheta_3) & -\sin(\vartheta_3) & 0 \\ \sin(\vartheta_3) & \cos(\vartheta_3) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (31)$$

where ϑ_j , $j = 1, \dots, K(K-1)/2$ are drawn from a uniform distribution on the interval $[0, \pi]$. Each random draw of rotation angles produces a candidate $B(\vartheta) = CQ(\vartheta)$ for the matrix B . Conditional on candidate $B(\vartheta)$, orthogonalized IRFs $\Theta_i(\vartheta) = \Phi_i B(\vartheta)$, $i = 0, \dots, h$, are calculated. If the signs of the resulting responses comply with the imposed restrictions, $B(\vartheta)$ is accepted; otherwise, it will be discarded. This procedure is repeated a sufficient number of times, until a pre-specified number of successful draws (in this study, 1000) is obtained. Note that the result from this identification approach is a set (e.g. 1000) of models, rather than a unique structural model. In order to summarize core implications of these models, we use the median of the impulse responses as the ‘point estimator’, which is by far the most popular choice in the literature, and we additionally report 16% and 84% quantiles from the empirical distribution of the accepted IRFs.

Sign restrictions are typically derived from theoretical models such as DSGE models (Fry and Pagan 2007, Canova and Paustian 2011). However, if the main idea of performing VAR analysis is to let the data speak about the dynamic relationships among the variables, insights from the imposition sign restrictions might be limited since the outcomes are - to some extent - predetermined by those a-priori restrictions. For instance, one cannot observe a ‘price puzzle’ if the data are not allowed to do so. Such considerations give rise to identification techniques that exploit certain statistical properties of the data and do not fully rely on economic assumptions. We briefly discuss two different data-driven identification procedures in the following subsections.

A.2 Identification by heteroskedasticity

It has been shown that changes of volatility in the data provide additional information that can be used for solving the identification problem. Rigobon (2003) and Lanne and Lütkepohl (2008) consider exogenous jumps in the unconditional variance of the shocks at discrete points in time. Assume that there is one permanent volatility break at time point T_B , i.e.

$$\mathbb{E}(u_t u_t') = \begin{cases} \Sigma_1 & \text{for } t = 1, \dots, T_B - 1 \\ \Sigma_2 & \text{for } t = T_B, \dots, T \end{cases} \quad (32)$$

Since both covariance matrices are positive definite and $\Sigma_1 \neq \Sigma_2$, there exist simultaneous factorizations $\Sigma_1 = WW'$, $\Sigma_2 = W\Psi W'$, where W is a $K \times K$ non-singular matrix and $\Psi \neq I_K$ is a diagonal matrix with positive diagonal elements $\varphi_k > 0$, $k = 1, \dots, K$. In particular, one could view the decomposition factor W as a solution for B (Lanne and Lütkepohl 2008) i.e. $W = B$. This provides a unique matrix B up to sign and column permutations, such that the underlying structural shocks $\varepsilon_t = W^{-1}u_t$ have covariances: $\mathbb{E}(\varepsilon_t \varepsilon_t') = I_K$ for $t = 1, \dots, T_B - 1$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Psi$ for $t = T_B, \dots, T$.

This choice of decomposition relies on the assumption that the structural parameters in matrix B remain constant across volatility regimes, which seems reasonable in certain scenarios (Christiano et al. 1999).³³ Another crucial point ensuring a unique identification is that there must exist at least one regime in which the volatility shift (relative to pre-shift period) in the k -th variable differs from that in the l -th variable, i.e. $\varphi_k \neq \varphi_l$ for any $l, k \in \{1, \dots, K\}$ (Lanne et al. 2010). For estimation purposes, we employ the algorithm provided by R package `svars` (Lange et al. 2020).

A.3 Identification by non-Gaussianity

Apart from changes in the second-order moment, it has been shown that if the data are not Gaussian - which is not uncommon in empirical analysis (Lanne and Saikkonen 2013, Lanne and Lütkepohl 2010) - this distributional feature can be used for identification. An important result of Comon (1994) states that the linear relation between reduced-form VAR residuals and the structural shocks is unique if at most one of the latter exhibits a Gaussian distribution. Given the uniqueness of the structural model established by assuming independent shocks, various approaches have been developed in the literature for SVAR identification through ICA. Owing to its flexibility and superior identification precision in a heteroskedastic environment shown by Herwartz et al. (2019), in this study we adopt the non-parametric approach that does not depend on any pre-specified likelihood function. For this purpose, we consider a space of decompositions of Σ_u based on the same Givens rotation matrix $Q(\vartheta)$ analogous to that used for sign restrictions defined in (31), and search on the interval $[0, \pi]$ for the set of rotation angles $\hat{\vartheta}$ that minimizes the dependence of the orthogonalized shocks

³³Bacchiocchi and Fanelli (2015) suggest a heterogeneous SVAR model that is capable of capturing the changes in the structural parameters.

$\{\varepsilon_t(\hat{\vartheta}) = Q(\hat{\vartheta})^{-1}C^{-1}u_t\}_{t=1}^T$. This dependence is measured by either the so-called distance covariance $\mathcal{U}_T(\vartheta)$ or the Cramér-von-Mises distance $\mathcal{C}_T(\vartheta)$ in this study, i.e.³⁴

$$\hat{\vartheta}^{dCov} = \arg \min_{\vartheta} \mathcal{U}_T(\vartheta) \text{ or } \hat{\vartheta}^{CvM} = \arg \min_{\vartheta} \mathcal{C}_T(\vartheta).$$

Conditional on the estimated optimal set of rotation angels $\hat{\vartheta}^\bullet$, the estimator for matrix B is given by $\hat{B}^\bullet = B(\hat{\vartheta}^\bullet)$, $\bullet \in \{dCov, CvM\}$. In this study, independence-based identification using distance covariance relies on function `steadyICA` from the R package `steadyICA` (Risk et al. 2015). For independence diagnostics based on CvM distance, a two-stage optimization algorithm implemented in R package `svars` (Lange et al. 2020) is employed.

B DSGE models

In the Monte Carlo simulation, we employ a stylized three-equation dynamic DSGE model that has been widely used as a baseline framework for MP analysis (Carlstrom et al. 2009, Castelnovo 2012, Herwartz and Plödt 2016). A small-scale model would typically allow us to have a better control of the mapping between structural innovations and observables. Moreover, it is reasonable to assume statistical independence of the structural shocks and hence the feasibility of identification through non-Gaussianity in a small-scale system, whereas this would be more difficult to justify for larger system, which might involve more related economic processes (Herwartz 2018). For a detailed description of the model, we refer to Herwartz and Plödt (2016). The log-linearized version of the model reads as:

$$x_t = \gamma E_t x_{t+1} + (1 - \gamma)x_{t-1} - \delta_x(r_t - E_t \pi_{t+1}) + \omega_{x,t} \quad (33)$$

$$\pi_t = (1 + \alpha\beta)^{-1}\beta E_t \pi_{t+1} + (1 + \alpha\beta)^{-1}\alpha\pi_{t-1} + \kappa x_t + \omega_{\pi,t} \quad (34)$$

$$r_t = \tau_r r_{t-1} + (1 - \tau_r)(\tau_\pi \pi_t + \tau_x x_t) + \omega_{r,t} \quad (35)$$

$$\omega_{\bullet,t} = \rho_\bullet \omega_{\bullet,t-1} + \varepsilon_{\bullet,t}, \quad \bullet \in \{x, \pi, r\}, \quad t = 1, \dots, T, \quad (36)$$

where x_t , π_t and r_t denote the output gap, inflation and nominal interest rate, respectively. Equation (33) is the IS curve in which γ stands for the weight allocated to a forward-looking component and δ_x governs the impact of the ex-ante real interest rate. Equation (34) is a New Keynesian Phillips curve with indexation to past inflation α , discount factor β and slope parameter κ . The policy parameters in the Taylor rule (35) are denoted by τ_r , τ_π and τ_x . All structural innovations in (36) are assumed to follow mutually-independent autoregressive processes of order one with coefficient ρ_x , ρ_π and ρ_r for demand shock, supply shock and MP shock, respectively. In line with Herwartz and Plödt (2016), the DSGE model is calibrated in the common setting. The parameter values together with intervals for purpose of AoA-construction are summarized in Table A1.

As shown by Carlstrom et al. (2009), under the condition of equilibrium determinacy ($\tau_\pi > 1$),

³⁴For exact definitions of \mathcal{U}_T and \mathcal{C}_T , we refer to Appendix A in Herwartz et al. (2019).

Parameter	Calibration	Interval	Parameter	Calibration	Interval
β	0.99	0.99	α	0.5	[0.4, 0.6]
κ	0.05	[0.03, 0.07]	γ	0.5	[0.4, 0.6]
δ_x	0.1	[0.05, 0.15]	τ_x	0.5	[0.3, 0.7]
τ_π	1.8	[1.6, 2.0]	τ_r	0.6	[0.4, 0.8]
ρ_x	0.5	[0.4, 0.6]	ρ_π	0.5	[0.4, 0.6]
ρ_r	0.5	[0.4, 0.6]			

Table A1: Calibration of DGP parameters and intervals for AoA.

the model can be expressed as:

$$\begin{bmatrix} x_t \\ \pi_t \\ r_t \end{bmatrix} = \Phi \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix} + B \begin{bmatrix} \omega_{x,t} \\ \omega_{\pi,t} \\ \omega_{r,t} \end{bmatrix} \quad (37)$$

Denoting $y_t = (x_t, \pi_t, r_t)'$ and $\omega_t = (\omega_{x,t}, \omega_{\pi,t}, \omega_{r,t})'$ we obtain

$$y_t = \Phi y_{t-1} + B \omega_t,$$

which implies the explicit parameterization

$$\Phi = \begin{bmatrix} 0.74 & -0.09 & -0.16 \\ 0.13 & 0.44 & -0.06 \\ 0.24 & 0.30 & 0.53 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2.32 & -0.48 & -0.41 \\ 0.72 & 2.32 & -0.22 \\ 0.98 & 1.57 & 0.76 \end{bmatrix}$$

under the parameter choices encountered in Table A1.

As shown in Ravenna (2007) and Carlstrom et al. (2009), the system has a VAR(2) representation:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + B \varepsilon_t, \quad (38)$$

where the autoregressive coefficient matrices are $A_1 = \Phi + BFB^{-1}$ and $A_2 = -BFB^{-1}\Phi$.

C Detailed simulation results

T		120				240				480			
		Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$	Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$	Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$
\mathcal{L}_I	IV	.639	.930	.618	.370	.734	.979	.735	.488	.821	.996	.847	.619
	IV-W	.376	.656	.292	.181	.387	.682	.299	.180	.394	.717	.291	.173
	IV-E	.376	.804	.240	.083	.391	.894	.207	.072	.399	.942	.185	.068
	SR	.456	.806	.371	.189	.469	.861	.369	.176	.462	.888	.331	.168
	CV	.649	.923	.605	.419	.753	.975	.747	.536	.843	.996	.858	.674
	dCov	.632	.915	.581	.399	.720	.970	.697	.495	.805	.995	.811	.609
	CvM	.545	.866	.470	.298	.625	.947	.576	.352	.705	.991	.686	.438
$\mathcal{L}_{II}^{(a)}$	IV	.638	.930	.608	.376	.735	.982	.731	.493	.822	.999	.851	.616
	IV-W	.373	.652	.276	.191	.391	.692	.292	.189	.392	.722	.291	.163
	IV-E	.380	.809	.246	.084	.393	.891	.215	.073	.400	.947	.186	.068
	SR	.454	.812	.368	.183	.467	.856	.366	.178	.469	.894	.344	.169
	CV	.644	.918	.600	.415	.749	.978	.745	.524	.847	.998	.872	.670
	dCov	.634	.908	.589	.405	.731	.974	.704	.515	.823	.997	.838	.634
	CvM	.536	.853	.466	.289	.633	.950	.584	.366	.705	.986	.691	.438
$\mathcal{L}_{II}^{(b)}$	IV	.634	.928	.614	.359	.730	.985	.739	.467	.820	.997	.854	.608
	IV-W	.372	.662	.274	.180	.382	.686	.283	.178	.396	.711	.292	.185
	IV-E	.372	.791	.241	.084	.390	.892	.207	.071	.397	.938	.186	.068
	SR	.451	.798	.370	.183	.468	.863	.368	.173	.471	.891	.351	.172
	CV	.646	.918	.594	.426	.750	.981	.738	.531	.834	.997	.856	.649
	dCov	.646	.918	.610	.410	.753	.982	.744	.533	.841	.997	.868	.657
	CvM	.552	.858	.493	.304	.645	.949	.605	.382	.730	.989	.716	.487
$\mathcal{L}_{III}^{(a)}$	IV	.637	.929	.602	.379	.734	.983	.738	.482	.823	.998	.855	.617
	IV-W	.369	.639	.285	.183	.389	.701	.291	.175	.397	.704	.304	.183
	IV-E	.377	.800	.247	.085	.390	.886	.212	.073	.398	.940	.185	.068
	SR	.450	.799	.367	.184	.467	.858	.361	.182	.464	.888	.338	.167
	CV	.644	.916	.588	.429	.741	.979	.729	.515	.842	.997	.861	.669
	dCov	.653	.919	.603	.438	.754	.979	.743	.539	.845	.996	.867	.672
	CvM	.554	.860	.494	.309	.672	.948	.636	.433	.762	.985	.762	.538
$\mathcal{L}_{III}^{(b)}$	IV	.633	.932	.594	.372	.738	.978	.735	.500	.829	.997	.860	.629
	IV-W	.382	.675	.287	.183	.385	.688	.288	.181	.394	.694	.300	.186
	IV-E	.379	.807	.243	.087	.391	.888	.212	.072	.398	.941	.185	.068
	SR	.448	.803	.358	.183	.464	.855	.359	.178	.465	.886	.338	.172
	CV	.646	.925	.585	.428	.749	.977	.728	.541	.845	.996	.870	.668
	dCov	.631	.914	.578	.400	.736	.973	.715	.521	.836	.995	.847	.667
	CvM	.522	.830	.445	.292	.616	.922	.549	.376	.712	.970	.689	.477

Table A2: Average acceptance frequencies of for $i = 0, \dots, 4$.

T		120				240				480			
		Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$	Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$	Avg.	$\varepsilon_1 \rightarrow$	$\varepsilon_2 \rightarrow$	$\varepsilon_3 \rightarrow$
\mathcal{L}_I	IV	.448	.552	.456	.334	.574	.654	.593	.476	.699	.753	.721	.624
	IV-W	.362	.494	.309	.284	.441	.578	.382	.362	.518	.673	.455	.425
	IV-E	.400	.487	.337	.375	.495	.596	.437	.452	.569	.688	.517	.501
	SR	.349	.491	.336	.221	.418	.587	.419	.248	.481	.674	.498	.272
	CV	.466	.556	.460	.382	.592	.658	.603	.516	.712	.753	.734	.650
	dCov	.468	.557	.455	.390	.599	.657	.604	.535	.714	.755	.727	.659
	CvM	.471	.572	.439	.400	.604	.670	.609	.532	.715	.755	.735	.656
$\mathcal{L}_{II}^{(a)}$	IV	.446	.548	.451	.337	.570	.650	.580	.479	.716	.766	.743	.640
	IV-W	.357	.486	.292	.294	.436	.574	.371	.363	.522	.684	.458	.424
	IV-E	.406	.488	.346	.384	.494	.587	.438	.457	.576	.699	.534	.495
	SR	.350	.494	.339	.215	.416	.574	.420	.254	.488	.687	.516	.261
	CV	.471	.552	.465	.396	.585	.651	.592	.512	.728	.766	.752	.666
	dCov	.467	.554	.455	.390	.589	.654	.591	.523	.731	.766	.752	.675
	CvM	.473	.575	.443	.401	.599	.671	.595	.530	.718	.764	.731	.659
$\mathcal{L}_{II}^{(b)}$	IV	.452	.552	.463	.342	.565	.654	.584	.458	.708	.770	.732	.623
	IV-W	.362	.496	.291	.299	.438	.578	.375	.360	.524	.684	.464	.423
	IV-E	.409	.494	.349	.385	.487	.591	.433	.438	.574	.699	.524	.500
	SR	.355	.496	.351	.218	.412	.582	.415	.240	.494	.685	.519	.277
	CV	.474	.554	.466	.403	.588	.654	.597	.514	.720	.769	.739	.653
	dCov	.474	.556	.476	.392	.589	.654	.595	.520	.719	.769	.741	.648
	CvM	.472	.571	.444	.400	.602	.660	.609	.535	.721	.767	.738	.659
$\mathcal{L}_{III}^{(a)}$	IV	.445	.539	.460	.335	.580	.661	.596	.482	.704	.763	.719	.630
	IV-W	.349	.471	.297	.280	.443	.587	.381	.360	.528	.677	.466	.440
	IV-E	.401	.481	.344	.379	.499	.598	.440	.459	.579	.694	.528	.514
	SR	.342	.478	.340	.209	.421	.589	.422	.253	.489	.675	.513	.281
	CV	.470	.543	.473	.393	.598	.663	.605	.526	.716	.763	.734	.652
	dCov	.467	.545	.470	.384	.598	.663	.606	.524	.718	.764	.727	.663
	CvM	.469	.564	.451	.392	.596	.661	.595	.533	.714	.765	.718	.659
$\mathcal{L}_{III}^{(b)}$	IV	.434	.538	.437	.327	.585	.664	.595	.496	.712	.764	.735	.635
	IV-W	.348	.470	.293	.281	.436	.582	.372	.354	.525	.679	.457	.437
	IV-E	.396	.478	.336	.375	.498	.603	.437	.455	.576	.700	.524	.504
	SR	.342	.485	.335	.206	.425	.596	.427	.253	.496	.687	.531	.271
	CV	.461	.540	.452	.389	.601	.667	.603	.532	.723	.764	.746	.658
	dCov	.459	.546	.450	.381	.594	.666	.595	.523	.722	.765	.736	.665
	CvM	.463	.576	.427	.386	.566	.651	.560	.487	.715	.774	.723	.649

Table A3: Average acceptance frequencies of for $i = 5, \dots, 15$.

	\mathcal{L}_I			$\mathcal{L}_{II}^{(a)}$			$\mathcal{L}_{II}^{(b)}$			$\mathcal{L}_{III}^{(a)}$			$\mathcal{L}_{III}^{(b)}$		
T	120	240	480	120	240	480	120	240	480	120	240	480	120	240	480
IV	.682	.761	.875	.667	.779	.853	.658	.802	.869	.688	.771	.865	.679	.773	.862
IV-W	.166	.184	.188	.184	.207	.187	.200	.195	.200	.211	.211	.183	.197	.213	.193
IV-E	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CV	.608	.736	.823	.592	.739	.835	.584	.716	.819	.614	.717	.820	.639	.694	.837
dCov	.562	.569	.670	.588	.643	.699	.637	.728	.838	.688	.775	.867	.664	.797	.871
CvM	.300	.269	.329	.330	.316	.402	.401	.427	.532	.497	.581	.671	.454	.590	.631

Table A4: Frequency of a unique monetary policy (UMPS) shock for proxy SVARs and data-based identification schemes. Agnostic sign restrictions obtain a UMPS by construction.

T		120				240				480			
		Avg.	$\rightarrow x$	$\rightarrow \pi$	$\rightarrow r$	Avg.	$\rightarrow x$	$\rightarrow \pi$	$\rightarrow r$	Avg.	$\rightarrow x$	$\rightarrow \pi$	$\rightarrow r$
\mathcal{L}_I	IV	0.726	0.890	0.930	0.358	0.518	0.639	0.660	0.255	0.400	0.461	0.540	0.200
	IV-W	2.192	4.559	1.516	0.501	2.254	4.538	1.744	0.480	2.245	4.625	1.640	0.470
	SR	2.460	6.011	1.066	0.303	2.483	6.058	1.103	0.289	2.487	6.058	1.121	0.284
	CV	0.786	0.737	1.310	0.311	0.543	0.457	0.955	0.217	0.394	0.319	0.706	0.158
	dCov	0.730	0.825	1.027	0.338	0.590	0.530	0.944	0.298	0.507	0.373	0.886	0.261
	CvM	1.519	2.406	1.538	0.614	1.152	1.331	1.527	0.598	1.021	0.978	1.482	0.602
$\mathcal{L}_{II}^{(a)}$	IV	0.690	0.929	0.813	0.329	0.519	0.640	0.664	0.253	0.397	0.476	0.516	0.198
	IV-W	2.123	4.507	1.399	0.465	2.127	4.508	1.443	0.430	2.281	4.745	1.602	0.496
	SR	2.450	5.980	1.067	0.304	2.489	6.077	1.099	0.292	2.497	6.088	1.120	0.285
	CV	0.788	0.762	1.293	0.309	0.553	0.467	0.966	0.227	0.387	0.322	0.681	0.158
	dCov	0.675	0.751	0.961	0.314	0.552	0.479	0.890	0.287	0.439	0.346	0.734	0.236
	CvM	1.448	2.224	1.520	0.599	1.159	1.326	1.529	0.621	1.021	0.930	1.509	0.625
$\mathcal{L}_{II}^{(b)}$	IV	0.715	0.933	0.858	0.355	0.549	0.658	0.709	0.278	0.395	0.449	0.529	0.207
	IV-W	2.146	4.496	1.461	0.480	2.102	4.326	1.520	0.459	2.202	4.603	1.549	0.454
	SR	2.446	5.987	1.043	0.308	2.462	6.000	1.091	0.294	2.482	6.038	1.118	0.288
	CV	0.784	0.747	1.285	0.321	0.565	0.477	0.982	0.236	0.420	0.350	0.734	0.177
	dCov	0.623	0.657	0.894	0.319	0.505	0.453	0.793	0.270	0.364	0.324	0.563	0.205
	CvM	1.287	1.795	1.465	0.601	1.013	1.052	1.411	0.576	0.933	0.871	1.372	0.555
$\mathcal{L}_{III}^{(a)}$	IV	0.723	0.948	0.872	0.349	0.541	0.667	0.693	0.264	0.394	0.467	0.512	0.204
	IV-W	2.249	4.691	1.573	0.483	2.271	4.597	1.707	0.508	2.047	4.133	1.562	0.445
	SR	2.467	6.039	1.058	0.305	2.476	6.036	1.099	0.292	2.493	6.070	1.124	0.284
	CV	0.790	0.728	1.319	0.323	0.580	0.503	1.002	0.234	0.404	0.313	0.734	0.165
	dCov	0.632	0.686	0.897	0.312	0.445	0.454	0.649	0.232	0.336	0.308	0.517	0.185
	CvM	1.158	1.548	1.351	0.574	0.934	1.063	1.224	0.514	0.824	0.808	1.179	0.484
$\mathcal{L}_{III}^{(b)}$	IV	0.720	0.952	0.867	0.343	0.533	0.654	0.678	0.266	0.404	0.482	0.530	0.198
	IV-W	2.106	4.335	1.518	0.463	2.159	4.506	1.509	0.462	2.097	4.346	1.498	0.447
	SR	2.442	5.961	1.064	0.301	2.480	6.051	1.100	0.289	2.490	6.063	1.123	0.283
	CV	0.744	0.742	1.198	0.294	0.582	0.485	1.025	0.235	0.400	0.337	0.704	0.159
	dCov	0.654	0.778	0.888	0.298	0.500	0.521	0.722	0.256	0.349	0.349	0.508	0.190
	CvM	1.306	1.999	1.355	0.565	1.033	1.284	1.276	0.539	0.844	0.895	1.150	0.486

Table A5: Relative root mean squared error of estimated effects of UMPS. As documented in Table A4, proxy SVARs that condition on invalid instrument fail throughout to obtain a UMPS. Therefore, performance statistics for IV-E are not provided.

D Instruments for the US monetary policy shock

Table A6: Summary of employed external MP shock series.

Shock series	Type	Sample period	F -statistics
RR	narrative	1969M1–2007M12	32.435
RR 83	narrative	1983M1–2007M12	7.489
FF1	high frequency	1988M1–2012M12	5.266
FF4	high frequency	1988M1–2012M12	5.039
FF1 gb	high frequency	1988M1–2012M12	0.443
FF4 gb	high frequency	1988M1–2012M12	0.650
MA	high frequency	1990M2–2009M12	2.536
GSS	high frequency	1990Q1–2004Q4	1.592
SZ	SVAR	1960M2–2003M3	39.336
SW	DSGE	1959Q1–2004Q4	19.810

Notes: **RR** denotes the narrative shock series of Romer and Romer (2004), constructed as the residuals from a regression of the federal funds *target* rate on Greenbook forecasts for FOMC meeting dates, which has been estimated over an extended sample until 2007M12 by Ramey (2016) (quarterly sums, demeaned). **RR 83** is the re-estimated original RR series starting in 1983M1 using long-horizon Greenbook forecasts (quarterly sums, demeaned). **FF1** (**FF4**) denotes the movements in the current federal funds futures rate (three month ahead futures rate) within a 30 minute window surrounding FOMC announcements from Gertler and Karadi (2015) but have been converted to a monthly basis by Ramey (2016) (quarterly average). **FF1 gb** and **FF4 gb** follow the same methodology but have been orthogonalized to the Romer Greenbook variables (Ramey 2016) (quarterly average). **MA** stands for the monetary surprise series that has been purged of anticipatory effects from monetary surprises, i.e. the reaction of agents to information revealed by the central bank’s monetary decision (Miranda-Agrippino 2016) (quarterly average). **GSS** is the estimated factor that measures the unexpected changes in the current target for FFR in Gürkaynak et al. (2005). **SZ** is the Sims and Zha (2006) MP shock from a SVAR specification allowing for regime-dependent volatility (quarterly average). **SW** is the MP shock resulting from the Smets and Wouters (2007) DSGE model. All series except for the **GSS**, **SW**, **SZ** and **MA** have been obtained from Valerie Ramey’s webpage (<https://econweb.ucsd.edu/~vramey/research.html#data>), while **GSS**, **SW** and **SZ** are available at Mark Watson’s webpage (<http://www.princeton.edu/~mwatson/publi.html>, downloaded at 9/3/20 for this study). The **MA** monetary surprise is available on Miranda-Agrippino’s webpage (<http://silviamirandaagrippino.com/research/>). The shock series are adjusted to be free of serial correlation. For critical values for F -statistics see Table 5 and Lunsford (2015).

E Monetary policy, information effects and endogeneity

Constructing a direct measure of MP shocks requires disentangling exogenous fluctuations from endogenous (i.e. systematic) responses of policy indicators. In practice, this could be complicated by the signaling channel (or forward guidance) adopted by monetary authorities, or through informational asymmetries between the monetary authority and economic agents. The implications for MP identification in an imperfect information context have recently attracted attention in theoretical and empirical studies (with further references, see Noh 2018, Miranda-Agrippino and Ricco 2017, Miranda-Agrippino 2016). This literature builds upon the intuition that if a central bank has superior information about the future (and possibly the current) state of the economy compared with the private sector (more generally, if their forecasts differ), policy measures comprise monetary surprises and information about fundamentals, and agents will adjust their expectations according to both components. Hence, if an instrument under scrutiny does not account for such an ‘informational asymmetry’, estimated causal relations are at risk of being biased, for two distinct reasons: on the one hand, differences in forecasts lead to predictability of instruments conditional on past structural shocks or equivalently lagged endogenous variables (i.e. the so-called lead/lag exogeneity assumption in Stock and Watson 2018); and on the other hand, if nowcasts differ, informational frictions result in contemporaneous contamination of instruments with non-MP shocks.³⁵

Regarding the violation of the lag exogeneity, this is equivalent to saying that the VAR is (partially) non-invertible and can potentially be cured by including enough information, such that the reduced-form residuals span the space of the structural shocks. From arguments in Noh (2018) one can conclude that non-invertibility is implied if instrumental information is Granger causal for the variables in the VAR. Conducting respective tests for our small-scale VAR and its factor-augmented counterparts obtains the results documented in Table 7. As it turns out, SW is the only instrument that does not Granger cause the VAR variables and the extracted common factors. By contrast, macroeconomic variables can be predicted by instruments RR and SZ, thus further calling into question the structural model conditioning on these instruments.

r	SW		RR		SZ	
	<i>F</i>	<i>p</i> -val.	<i>F</i>	<i>p</i> -val.	<i>F</i>	<i>p</i> -val.
0	1.685	0.066	1.494	0.122	2.400	0.005
1	1.280	0.203	1.496	0.095	2.081	0.008
2	1.082	0.363	1.507	0.071	2.058	0.004
3	1.104	0.332	1.714	0.018	1.671	0.023

Table 7: Test for Granger (non-)causality in (FA)VARs with r factors. $H_0 : z \not\stackrel{GC}{\rightarrow} (y, F)$. Instruments are adjusted to be serially uncorrelated. The lag order in all (FA)VARs is four.

³⁵Arguably, models of forward guidance or signaling also result in a violation of the lag exogeneity.