Monetary Policy, External Instruments and Heteroskedasticity*

Malte Rieth†

Thore Schlaak‡

Maximilian Podstawski§

This version: February 13, 2019

Abstract

We develop a vector autoregressive framework that combines an external instrument and heteroskedasticity for the identification of monetary policy shocks. We show that exploiting both types of information sharpens structural inference, allows testing both the relevance and exogeneity condition for instruments, and largely dispenses the proxy-VAR approach from weak instrument problems. Building on the proposed framework, we document that surprise monetary contractions identified with a valid instrument lead to a significant and medium-sized decline in economic activity. Models with external instrument that neglect the identifying information in heteroskedasticity are less efficient and tend to underestimate the effects of monetary policy.

Keywords: Monetary policy, structural vector autoregressions, identification with external instrument, heteroskedasticity, Markov switching.

JEL classification: E52, C32, E58, E32.

---

*We are thankful to Dario Caldara, Lutz Kilian, Alexander Kriwoluzky, Helmut Lütkepohl, Geert Peersman, Dieter Nautz, Konstantinos Theodoridis, Mathias Trabandt, Rafael Wouters, as well as the conference and seminar participants at the Workshop on Structural VAR Topics Queen Mary University London, Freie Universität Berlin, and DIW Berlin.

†Corresponding author, DIW Berlin, Department of Macroeconomics, Mohrenstr. 58, 10117 Berlin, Germany, email: mrieth@diw.de

‡Freie Universität Berlin, email: schlaak@zedat.fu-berlin.de

§DIW Berlin, email: maximilian.podstawski@gmail.com

1 Introduction

Estimating the effects of monetary policy is a central element of macroeconomic analysis. While the economy reacts to policy decisions, monetary policy is also endogenous to the state of the economy, posing the issue of isolating exogenous variation in monetary policy. In the empirical literature, structural vector autoregressions (SVARs) are a main tool for studying the causal effects of monetary interventions. Departing from the classical identification via zero restrictions (Sims, 1980; Christiano, Eichenbaum and Evans, 1999), two identification approaches are receiving increasing attention in the literature. On the one hand, authors use extraneous data on monetary surprises to identify latent monetary shocks in SVARs.[1] On the other hand, many papers draw on volatility changes in macroeconomic and financial data to identify monetary shocks.[2] Both identification strategies are popular because they are parsimonious in terms of identifying assumptions and because they incorporate further information into the model.

Identification via external instrument allows for a contemporaneous response of monetary policy to asset prices. Moreover, it adds a potentially large information set to the model through a narrative or financial data-based instrument. Finally, it accounts for measurement error in the instrument, which reduces the attenuation bias in models treating the proxy as the true shock (Mertens and Ravn, 2013; Carriero, Mumtaz, Theodoridis and Theophilopoulou, 2015). However, these advantages rely on the presumption that the instrument is valid, that is, sufficiently strong and exogenous.

Identification through heteroskedasticity adds information from time-varying second moments to the model and relies on even weaker identifying assumptions. While an instrument for monetary policy shocks needs to move interest rates without correlating with other structural shocks, a significant relative increase in the variance of monetary shocks is sufficient to trace out the response of the other variables in the system to these shocks. The relative variance shift can be viewed a 'probabilistic instrument' that increases the likelihood that monetary policy shocks occur (Rigobon, 2003). Again, these minimal assumptions are not costless. The statistically identified shocks are often economically difficult to interpret.

This paper proposes a framework that combines both sources of information in order to improve identification within SVARs and to address some of the limitations that each of the two identification approaches has in isolation. The framework

[1]See Gertler and Karadi (2015), Cesa-Bianchi, Thwaites and Vicondoa (2016), Miranda-Agrippino and Ricco (2017), Stock and Watson (2018), Rogers, Scotti and Wright (forthcoming), Hachula, Piffer and Rieth (forthcoming), Caldara and Herbst (2019).

[2]See Rigobon and Sack (2004), Normandin and Phaneuf (2004), Lanne and Lütkepohl (2008), Wright (2012), Herwartz and Lütkepohl (2014), Bacchiocchi and Fanelli (2015), Nakamura and Steinsson (2018).

makes use of an external instrument, drawing on instruments for monetary policy shocks proposed in the literature. In addition, it exploits time-variation in the second moments of the data. The combination of both types of identifying information into a 'heteroskedastic proxy-VAR' has three main advantages relative to models using either type of information separately for identification.

First, the encompassing framework improves the identification of the structural model and, hence, the suitability of the model for policy analysis. We conduct an extensive simulation study. It suggests that the incorporation of identifying information from an external instrument and from time-varying second moments yields more accurate estimates of the true model according to the cumulated mean squared errors of impulse response functions. The heteroskedastic proxy-VAR is more efficient than either of the two existing identifcation approaches in isolation. This facet of our framework is similar in spirit to the analysis of Antolín-Díaz and Rubio-Ramírez (forthcoming) who combine narrative information and sign restrictions to sharpen inference in SVARs.

We use our framework with tightened grip on the structural model to provide new estimates of the macroeconomic effects of monetary policy shocks in the United States. A common and well documented feature of U.S. real and financial data is time-varying volatility (Stock and Watson, 2002; Justiniano and Primiceri, 2008; Carriero, Clark and Marcellino, 2016). Standard statistics provide strong evidence that changes in volatility are also present in our sample. We model them within a Markov switching in heteroskedasticity framework and use them for identification. As second central piece of identifying information we include the measure of unanticipated changes in the intended federal funds rate of Romer and Romer (2004) into the model. We find that an unexpected increase in the federal funds rate by 25 basis points leads to a cumulative fall in economic activity of about 0.7 percent. These effects are twice as large as estimates obtained from a standard proxy-VAR that does not exploit the heteroskedasticity. Modeling changes in volatility also allows us to evaluate whether the importance of monetary policy shocks changes across volatility regimes. Our results indicate that monetary shocks were more volatile during the 1970s and 1980s than in the 1990s and 2000s, and that this is associated with a substantially larger role for them in driving real and financial variables in these decades. During the Great Moderation monetary shocks are essentially irrelevant for business cycle fluctuations.

A second contribution of our framework is that it allows testing the validity, that is, exogeneity and relevance, of an instrument. Our framework includes the instrument as an endogenous variable in an augmented SVAR, as in Caldara and Herbst (2019). When using the heteroskedasticity in the residuals of the augmented model, both the exogeneity and the relevance condition become testable. This conveniently reduces to testing zero restrictions on the structural impact matrix of the augmented SVAR. We propose a testing sequence using likelihood ratio (LR) tests for that purpose. Monte Carlo evidence suggests that our exogeneity test has desirable properties in terms of size and power. Testing the exogeneity assumption has so far been unresolved in the literature but is of particular interest as the violation of instrument exogeneity may lead to erroneous conclusions regarding the validity of the instrument and the effects of latent structural shocks.

In the empirical analysis, we first test the narrative measure of Romer and Romer (2004). It is contemporaneously exogenous to technology and financial shocks. Then, we compare alternative instruments for monetary policy shocks proposed in the literature. We find that model-based measures (Bernanke and Mihov, 1998) and high-frequency instruments (Gertler and Karadi, 2015) are also exogenous proxies, and that they produce similar effects as the narrative measure. All three types of instruments imply a significant decline in economic activity during the period of the Great Moderation (see the discussion in Barakchian and Crowe, 2013; Ramey, 2016; Caldara and Herbst, 2019). The exogeneity test complements the invertibility test for SVAR models identified with external instruments of Stock and Watson (2018). Our augmented SVAR provides a natural way for implementing that test as well. Since the instrument enters all equations of the model we can simply test whether it Granger-causes the endogenous variables of the system. We find that it does not, implying that our SVAR is invertible.

The third contribution of the paper is related to the relevance condition for instruments and the literature on identification through heteroskedasticity. Our framework largely dispenses the proxy-VAR approach from weak instrument problems (Lunsford, 2015; Olea, Stock and Watson, 2018). If there is sufficient time-variation in the second moments for identification – a condition that can be checked after estimation – the model is statistically identified. Then, the relevance condition is no longer necessary for statistically valid inference. Whether the instrument is relevant reduces to an economic question about the informational content of the instrument and the interpretation of the associated structural shock. The Monte Carlo evidence suggests that our LR-test reliably discriminates between relevant and irrelevant instruments. The test complements existing versions of F-tests for instrument strength (Stock, Wright and Yogo, 2002; Stock and Watson, 2012; Mertens and Ravn, 2013). It has more power than the F-test because it uses all information both under the null and the alternative hypothesis. Thereby, our framework also simplifies the economic interpretation of the shock of interest, addressing a main challenge in the literature on identification through heteroskedasticity (Rigobon and Sack, 2003; Herwartz and Lütkepohl, 2014). In this class of models, structural shocks are identified statistically. They need to be labeled by the researcher after estimation. While the literature has developed several devices for that purpose, this task is often difficult and can leave doubts about the economic meaning of the structural shocks. Through the inclusion of a relevant (and exogenous) proxy into the model, the shock of interest is pinned down by prior economic reasoning.

The remainder of the paper is structured as follows. The next section introduces the heteroskedastic proxy-VAR framework and discusses the identification, testing, and estimation of the model. Section 3 presents simulation results in support of the framework. In Section 4, we use the heteroskedastic proxy-VAR to shed new light on the efficacy of monetary policy and to test a range of instruments discussed in the literature. Finally, Section 5 concludes.

2 The SVAR framework

The vector autoregressive (VAR) model is

$$y_t = \gamma + A(L)y_{t-1} + u_t, \tag{1}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a $(K \times 1)$-vector of observable variables, $A(L)$ is a lag matrix polynomial capturing the autoregressive component of the model, γ collects constant terms, and the u_t are K-dimensional serially uncorrelated observable residuals. The reduced form residuals u_t are linearly related to white noise structural shocks ε_t, according to

$$u_t = B\varepsilon_t. \tag{2}$$

We assume that the VAR is invertible and has a Wold moving average representation $y_t = \gamma + \sum_{i=0}^{\infty} \Phi_i u_{t-i}$. We test this assumption in the empirical analysis and find no significant evidence against it.

2.1 Identification via external instrument

We assume that there is an instrumental variable s_t which is correlated with the structural shock of interest, but uncorrelated with other structural shocks and hence fulfills

$$\mathbb{E}[s_t\varepsilon_{1t}] = \phi \neq 0 \tag{3}$$
$$\mathbb{E}[s_t\varepsilon_{jt}] = 0 \quad \forall\ j = 2, \ldots, K, \tag{4}$$

where ϕ is an unknown correlation between the instrument s_t and the structural shock of interest ε_{1t}. The latter is ordered first without loss of generality. In the literature, (3) is usually called the relevance condition and assumption (4) the exogeneity condition. A valid instrument satisfies both (3) and (4). It allows to recover ε_{1t} and, hence, the corresponding response vector from the reduced form residuals. Rewriting (2) with $B = [b_1, B^*]$, where b_1 is the response vector corresponding to ε_{1t} and B^* contains the responses of the remaining shocks, yields

$$u_t = b_1\varepsilon_{1t} + B^*\varepsilon_t^*. \tag{5}$$

Substituting (5) into $\mathbb{E}(s_t u_t)$ and using (3) and (4) allows uncovering the (relative) impact of the structural shock of interest on every variable in the system, that is, the jth element of b_1 (Stock and Watson, 2012; Mertens and Ravn, 2013; Piffer and Podstawski, 2018). By using the sample moments $\hat{\mathbb{E}}(u_t s_t)$, the instrument s_t implies the following $k-1$ identifying restrictions

$$b_1 = b_{11}\left(1, \frac{\hat{\mathbb{E}}(u_{2t}s_t)}{\hat{\mathbb{E}}(u_{1t}s_t)}, \dots, \frac{\hat{\mathbb{E}}(u_{Kt}s_t)}{\hat{\mathbb{E}}(u_{1t}s_t)}\right)', \tag{6}$$

posing identification of shock ε_{1t} up to the scaling factor b_{11}.

2.2 A heteroskedastic proxy-VAR

A common feature of macroeconomic and financial data are changes in volatility over time (see, among others, Stock and Watson, 2002; Justiniano and Primiceri, 2008; Carriero et al., 2016). Rigobon and Sack (2004), Normandin and Phaneuf (2004), and Lanne and Lütkepohl (2008) show that this holds in particular for the analysis of monetary policy where changes in volatility of the data feed into heteroskedastic residuals in monetary SVARs. Against this backdrop, we allow for heteroskedastic residuals in (1).[3] We assume that the volatility changes are driven by a first order Markov switching (MS) process $S_t \in \{1, \dots, M\}$ with M states and transition probabilities $p_{kl} = P(S_t = l|S_{t-1} = k), k, l = 1, \dots, M$. Furthermore, the reduced form residuals are normally and independently distributed conditional on a given state $u_t|S_t \sim \text{NID}(0, \Sigma(S_t))$, where all Σ_m, $m = 1, \dots, M$ are distinct.

Modeling heteroskedasticity in the structural shocks holds direct implications for the external variable instrumenting one of the shocks as the instrument is likely to be heteroskedastic itself. We follow Caldara and Herbst (2019) in assuming that the process generating the potentially heteroskedastic instrument s_t has the following linear form:

$$s_t = \beta\varepsilon_t + \eta\nu_t, \tag{7}$$

[3]We refrain from introducing additional nonlinearity into the model by allowing state-dependency in the constant or autoregressive parameters as we are interested in the heteroskedasticity features of the data for identification purposes.

where ε_t is the $K \times 1$ vector of structural shocks, $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$ is a $1 \times K$-coefficient vector, $\nu_t \sim N(0, \sigma_m^2)$ is a measurement error uncorrelated with the structural shocks ε_t, and η scales the effect of the noise. β_1 and η may be interpreted as weighting parameters of signal to noise, defining the quality of the instrument s_t. The instrument's quality is flawed by the noise ν_t and potentially by the influence of other structural shocks on the instrument through β_j with $j = 2, \ldots, K$, determining the degree of its endogeneity.[4]

We compile the system by appending model (1) with the process generating the external instrument in (7). The augmented VAR system is

$$z_t = \delta + \Gamma(L) z_{t-1} + e_t, \tag{8}$$

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

$$\begin{aligned} e_t &= D\mu_t \\ &= \begin{bmatrix} B_{(K\times K)} & 0_{(K\times 1)} \\ \beta_{(1\times K)} & \eta \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix}. \end{aligned} \tag{9}$$

Using (9), we rewrite the augmented VAR in (8) in structural form as

$$z_t = \delta + \Gamma(L) z_{t-1} + D\mu_t. \tag{10}$$

Since the state dependency in the variances of the reduced form residuals in (8), $var(e_t|m) = \tilde{\Sigma}_m$ with $m = 1, \ldots, M$, translates into the structural form, we have $E[\mu_t] = 0$ and $E[\mu_t \mu_t'] = \Lambda_m$, where Λ_m is a diagonal matrix satisfying the orthogonality condition of the structural innovations.

Beyond identifying information from the external instrument, the heteroskedasticity pattern provides a valuable source of identifying information (Rigobon and Sack, 2004; Normandin and Phaneuf, 2004; Lanne and Lütkepohl, 2008). Under the assumption of a constant instantaneous impact matrix D, for each volatility regime a decomposition

$$\tilde{\Sigma}_m = D\Lambda_m D' \tag{11}$$

[4]Time-variation in the second moments of the data may imply a time-varying correlation ϕ_m in (3). However, under the assumption of a time-invariant impact matrix B, which is standard in the literature on external instruments (Stock and Watson, 2012; Mertens and Ravn, 2013; Gertler and Karadi, 2015), this state-dependency does not imply any changes in the use of the instrument for identification, as ϕ (or ϕ_m) does not enter the relative impulse vector (6).

exists, where $\Lambda_m = \text{diag}(\lambda_{1,m}, \ldots, \lambda_{K+1,m})$. We normalize $\Lambda_1 = I_{K+1}$. For $m \geq 2$, the Λ_m are diagonal matrices with strictly positive elements that can be interpreted as the changes of the variances of the structural innovations in the respective regime relative to the first regime. Lanne, Lütkepohl and Maciejowska (2010) state conditions for local uniqueness of matrix D. Local uniqueness implies that D is identified up to the signs of the parameters in each column as well as to column permutations. The conditions for local uniqueness of D are: (i) the structural impact matrix D is time-invariant; (ii) the structural innovations μ_t are orthogonal; and (iii) there are sufficiently many and distinct changes in the variances of the structural innovations. The first assumption is standard in structural VARs identified with external instruments.[5] The second assumption is common in structural VAR analysis more generally. The third assumption can be checked after estimation by comparing the estimated variances λ_{lm}, with $l = 1, \ldots, K + 1$.

To see how the model setup relates to the conventional identification of a proxy-VAR note that the $(K + 1, i)$-element of $\tilde{\Sigma}_m$ in (11) is $cov(u_{it}s_t)$. This term is equated to the corresponding element of $D\Lambda_m D'$ which is $\beta_1\lambda_{1m}b_{i1}$ for $i = 1 \ldots K$ and $\beta_2 = \cdots = \beta_K = 0$. It follows that $\frac{cov(u_{it}s_t)}{cov(u_{1t}s_t)} = \frac{b_{i1}}{b_{11}}$. This ratio is usually termed the relative impulse vector. It summarizes the restrictions on the structural parameters of the model implied by the instrument.

2.3 Testing the validity of an external instrument

If the conditions for local uniqueness are met, the heteroskedasticity in the residuals allows for estimating all structural parameters of D of the augmented SVAR model (10). Any additional restrictions on D are then over-identifying and, hence, testable. This is particularly interesting in our context as it allows for testing both the relevance and the exogeneity of the instrument and, thus, its validity. Such tests conveniently reduce to testing zero restrictions on β, that is, the last row of the structural impact matrix D. Given full identification of the model via heteroskedasticity, this may be done with likelihood ratio tests (LR-tests), as in Lanne and Lütkepohl (2008) because the elements of β are fixed parameters under the null hypothesis. Testing zero restrictions on these parameters then implies evaluating nested models against each other. This furthermore implies that the distribution of the LR-tests is a standard χ^2-distribution and the degrees of freedom are equal to the number of restrictions.

[5] Using alternative identification schemes, Owyang and Ramey (2004), Primiceri (2005) and Sims and Zha (2006) examine the role of changes in the monetary policy rule over time. While they find some evidence for regime-switches, they conclude that these changes explain only a small part of U.S. business cycle fluctuations (Ramey, 2016). Other authors find little or no evidence of changes in the policy coefficients (Bernanke and Mihov, 1998; Leeper and Zha, 2003; Hanson, 2006).

We propose a two-stage testing sequence. First, we assess the *exogeneity condition* of the instrument by comparing the likelihood of an appropriately restricted version of model (10), that is, restricting $\beta = (\beta_1, 0, \ldots, 0)$, with an unrestricted model (10) where $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$. Formally, we test

$$H_0 : \beta_2 = \cdots = \beta_K = 0$$
$$H_1 : \exists \beta_j \text{ s.t. } \beta_\mathrm{j} \neq 0 \quad \text{for} \quad \mathrm{j} = 2, \ldots, \mathrm{K}.$$

Rejecting the null indicates endogeneity of the instrument. Otherwise we proceed to the second stage.

Here, we assess the *relevance condition* by comparing an appropriately restricted version of model (10), that is, setting $\beta_1 = 0$, with model (10) where β_1 is unrestricted. Under both the null and the alternative hypothesis $\beta_2 = \ldots, \beta_K = 0$. Formally, we test

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0.$$

Rejecting the null indicates the relevance of the instrument and, together with not rejecting the null of instrument exogeneity at the first stage, that it is valid. If the instrument is valid, we set $\beta = (\beta_1, 0, \ldots, 0)$ and refer to model (10) as a 'heteroskedastic proxy-VAR'.

Local uniqueness in our setup implies that *a priori* we cannot identify the column of the impact matrix D that belongs to a certain structural shock. Practically and in the simulations this is of little concern because a valid instrument for a monetary policy shock imposes additional restrictions on the covariance matrix. Consequently the shock that is most consistent with these restrictions is ordered to the column with the unrestricted β-coefficient. This pins down the monetary policy shock. Furthermore, assessing the endogeneity of instrument does not require a particular ordering of the remaining structural shocks as the test for exogeneity will reject the null of all but one β-element equal to zero in case of endogeneity. A weak external instrument could potentially prevent the shock of interest to be ordered in the column with the unrestricted β-element. However, this will be detected at the second stage of the testing sequence through not rejecting the null, which indicates an uninformative instrument.

2.4 Estimation

The parameters of (10) are estimated by means of the expectation maximization (EM) algorithm proposed by Herwartz and Lütkepohl (2014). Crucial for the analysis is to incorporate the regime-switching nature of the covariance matrix described in (11), given the restrictions on D and Λ_m. All other parameters are assumed to be regime-independent and do not vary across states. We use the following concentrated out log likelihood function in the maximization step of the EM algorithm and refer to Appendix A.1 for further computational details[6]:

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

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

Once the EM algorithm has converged, we obtain standard errors of the point estimates of the parameters through the inverse of the negative Hessian matrix evaluated at the optimum. We use these standard errors as a statistic to determine whether the estimated structural variances change significantly and by differing amounts across states. We check that the standard errors do not overlap. This is a requirement for statistical identification and, hence, for the testing sequence. For the dynamic analysis, we compute bootstrapped impulse responses. Given the heteroskedasticity, classical residual bootstrapping may be problematic in generating reliable confidence intervals. Any re-sampling scheme needs to preserve the second order characteristics of the data so that the structural parameters are maintained and can be estimated from bootstrapped samples. Therefore, we use a recursive design wild bootstrap with $e_t^\star = \varphi_t \widehat{e}_t$, where φ_t is a random variable independent of z_t following a Rademacher distribution. φ_t is either 1 or –1 with probability 0.5 and the bootstrap procedure is repeated 5,000 times. This is a commonly used technique for these types of models (Herwartz and Lütkepohl, 2014; Podstawski and Velinov, 2018). Alternatively, Lütkepohl and Schlaak (forthcoming) show that a related bootstrap method based on a normal distribution performs well for a model with volatility changes driven by GARCH processes. In Figure A.8 in Appendix A.3 we show that our results are robust to using this bootstrap.

3 Simulation study

To explore the properties of the proposed framework and LR-tests, we conduct an extensive Monte Carlo simulation. First, we evaluate how the tests behave for different degrees of instrument endogeneity and relevance. Second, we assess whether the accuracy of the estimation of the structural model increases systematically due to the explicit modeling of heteroskedasticity. We also discuss how the proposed framework largely resolves problems of weak instruments in proxy-VARs.

[6]The online appendix is available at `https://www.dropbox.com/s/qe8otvjp1nb9h58/online_appendix_aejm.pdf?dl=0`.

3.1 Setup of Monte Carlo study

We assume that the data generating process is of the form (10). The process implies that y_t and s_t are jointly normally distributed conditional on state $m = 1, \ldots, M$. In the simulation, we generate data for y_t and then for the instrument s_t contingent on the realizations of y_t. We use the following parameters for a first-order autoregressive model with instantaneous effect matrix B, which are taken from the New Keynesian DSGE-model of An and Schorfheide (2007):

$$\begin{bmatrix} r_t \\ x_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} 0.79 & 0.00 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0.00 & 0.62 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ x_{t-1} \\ \pi_{t-1} \end{bmatrix} + \begin{bmatrix} 0.69 & 0.61 & 0 \\ -1.10 & 1.49 & 1 \\ -0.75 & 1.49 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t^r \\ \varepsilon_t^z \\ \varepsilon_t^g \end{bmatrix},$$

where r_t is the interest rate, x_t is output and π_t is the inflation rate. The structural shocks are characterized as a monetary policy shock (ε_t^r), a productivity shock (ε_t^z), and a government spending shock (ε_t^g).

The variances of the structural innovations are driven by a discrete Markov switching process with $M = 2$ states and transition probabilities

$$P = \begin{bmatrix} 0.975 & 0.025 \\ 0.050 & 0.950 \end{bmatrix},$$

which are used to generate the Markov states S_t for $t = 1, \ldots, T$. Following standard conventions, we normalize the variances of the structural innovations in the first state to unity. We set the relative variances in the second state by choosing rather distinct variances in the range of parameters used in comparable studies (Lütkepohl and Schlaak, 2018):

$$\Lambda_2 = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}.$$

Given that the B matrix may be identified up to column signs and permutations only, we assure that the models are uniquely determined by sorting the estimated coefficients of Λ_2 in descending order and by adjusting the columns of the estimated impact matrix correspondingly. With appropriate starting values $y_0 = (0, 0, 0)'$, we generate data recursively by drawing from

$$\varepsilon_t \sim \begin{cases} N(0, I), & \text{for} \quad m = 1 \\ N(0, \Lambda_2), & \text{for} \quad m = 2, \end{cases}$$

using $B\varepsilon_t = u_t$ to calculate the reduced form residuals.

With the structural innovations at hand, we generate the instrument s_t, using (7). We set $\eta = 1$ and the variances of the noise parameter ν_t such that

$$\nu_t \sim \begin{cases} N(0,1), & \text{for} \quad m = 1 \\ N(0,12), & \text{for} \quad m = 2. \end{cases}$$

This setup implies a time-varying volatility of the instrument which can be observed in many time series of instruments that are used in the literature (see, for example, Romer and Romer, 2004; Gertler and Karadi, 2015).

The relationship between the structural shocks and the artificial instruments is modeled through β. We set $\beta = (\beta_1, \beta_2, 0)$, where β_1 captures the relevance of the instrument for the monetary shock ε_t^r, while β_2 measures the endogeneity to the second structural shock ε_t^z. We equate β_3 to zero to focus the simulation study, concentrating on cases where endogeneity stems only from one source. We construct a set of different instruments for the target monetary policy shock, using the following values: $\beta_1 \in [0, 0.2, 0.4, 0.6]$. Similarly, we consider $\beta_2 \in [0, 0.05, 0.20, 0.30, 0.40]$ and thereby introduce different degrees of endogeneity. These parameter combinations for β_1 and β_2 imply different correlations of the instrument with the monetary shock (ρ_1) and the non-monetary shock (ρ_2), which we list in the simulation results to facilitate a comparison to empirical applications. Finally, we use two sample sizes, $T = 200$ and $T = 500$, which are within the typical range of macroeconomic datasets. The number of replications for each simulation design is $R = 500$.

3.2 Fitted models

As a reference model for the test evaluation, we fit a MS(2)-VAR(1) with unrestricted β to the data. Then, we estimate and compare the following three models:

Model A Heteroskedastic proxy-VAR with $\beta = (\beta_1, 0, 0)$, that is, the instrument s_t is assumed to be exogenous.

Model B Heteroskedastic SVAR with $\beta = (0, 0, 0)$, that is, the model is identified by the time-varying volatility only.

Model C Standard proxy-VAR using the identifying information from the external instrument only.

The reference model with unrestricted β and models A and B allow for a time-varying covariance, which we model within a Markov switching in heteroskedasticity framework with two latent states. We place alternative restrictions on β as discussed in Section 2.2. Model C fits a standard proxy-VAR with the two stage least squares procedure suggested by Mertens and Ravn (2013) to evaluate a situation where the volatility in the data is ignored. Here, the response of the first variable to the identified structural shock is normalized to a have a positive sign. This model has *a priori* a disadvantage compared to the other models, given that the generated data feature volatility changes. Given that the reference model and models A and B are nested, we can compute χ^2-distributed LR-statistics to test for the exogeneity and the relevance of the generated instrument in each replication.[7] To test the exogeneity condition, we test the heteroskedastic reference VAR with β unrestricted against model A. To test the relevance of the instrument, we test model A against the more restricted model B.

To assess the benefits of combining identification via external instrument and via heteroskedasticity, we calculate the cumulated mean squared errors (MSE) of the estimated structural impulse response functions for models A-C relative to the true parameters of the data generating process. This metric summarizes the accuracy of the structural estimates relative to the true model. The cumulated MSE up to horizon h for variable k induced by shock l is calculated as

$$MSE_h(\theta_{kl,\bullet}) = \sum_{i=0}^{h-1} \left(\frac{1}{R} \sum_{r=1}^{R} (\theta_{kl,i} - \hat{\theta}_{kl,i}(r))^2 \right),$$

where $\hat{\theta}_{kl,i}(r)$ denotes the estimate of the structural impulse response $\theta_{kl,i}$ obtained in the r^{th} replication of our simulation experiment.[8] As the data generating VAR(1) parameters imply substantial persistence, we calculate cumulated MSE for a propagation horizon of up to $h = 25$ such that we capture the impact of differing estimates of both the impact matrix B and the autoregressive part of the model. Finally, we assess the accuracy of the estimates for the monetary policy shock only in order to accommodate the fact that the identification via a single external instrument, as in model C, facilitates the identification of one shock per instrument at most.

3.3 Simulation results

Table 1 shows the relative rejection frequencies of the LR-test for exogeneity at a nominal significance level of 10% for the two different sample sizes. The complete

[7] If $\beta = 0$, the heteroskedastic proxy-VAR reduces to a standard heteroskedastic SVAR where the distributions of s_t and u_t are independent. In this case, the structural parameters are identified using the heteroskedasticity of the data only.

[8] The structural impulse responses of the models are obtained as the elements of the matrices $\Theta_i = \Phi_i B$, $i = 0, 1, \ldots$, where Φ_i is the coefficient matrix of the i^{th} propagation horizon of the Wold moving average representation of the VAR. More precisely, the kl^{th} element of Θ_i, denoted by $\theta_{kl,i}$, is interpreted as the response of variable k to the l^{th} structural shock after a propagation horizon of i periods.

set of simulation results may be found in Appendix A.2. Exogenous instruments with $\rho_2 = 0$ are rejected with frequencies close to their expected nominal levels (see first column). This suggests that the test is neither under nor over-sized. When moving to the right across columns, the LR-test has power against the null hypothesis of an exogenous instrument. The rejection frequencies steadily increase with higher instrument endogeneity for both sample sizes. For $T = 200$, endogeneity is detected in roughly 50% of the cases, when the sample correlation of the non-monetary shock with the instrument is elevated. For $T = 500$ and $\rho_2 \geq 0.16$, the rejection frequencies are usually higher than 90%, depending on the relevance of the instrument. For a nominal significance level of 5% the rejection frequencies tend to be lower, but for $T = 500$ and average sample correlations $\rho_2 \geq 0.12$, endogeneity is reliably detected in more than 80% of the cases.

Table 1: Relative rejection frequencies at nominal significance level of 10% of LR-test for exogeneity of instrument.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0.0,0.0)	(0.05,0.03)	(0.20,0.12)	(0.30,0.16)	(0.40,0.22)
200	(0.00,0.00)	0.12	.	.	.	.
	(0.20,0.16)	0.10	0.13	0.39	.	.
	(0.40,0.30)	0.09	0.12	0.31	0.52	0.63
	(0.60,0.43)	0.10	0.10	0.25	0.42	0.54
500	(0.00,0.00)	0.12	.	.	.	.
	(0.20,0.16)	0.10	0.17	0.85	.	.
	(0.40,0.30)	0.09	0.14	0.70	0.90	0.96
	(0.60,0.43)	0.09	0.12	0.50	0.78	0.91

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

Table 2 displays the relative rejection frequencies of the LR-test for instrument relevance. We focus on a 5% nominal significance level as the test displays relative rejection frequencies very close to one against false null hypotheses for a significance level of 10% (see Table A.2 in Appendix A.2). For a white noise instrument without any identifying information ($\rho_1 = \rho_2 = 0$), the test shows an expected nominal rejection rate around 5%. The rejection frequency rapidly increases for higher correlations of the instrument with the monetary shock. The null of an uninformative instrument is rejected reliably in all cases and for both samples if $\rho_1 \geq 0.30$, irrespective of the endogeneity of the instrument.

To obtain an impression of the power of the LR-test and the relevance of

Table 2: Relative rejection frequencies at nominal significance level of 5% of LR-test for relevance of instrument.

Sample Size	Relevance (β_1,ρ_1)	Endogeneity (β_2,ρ_2)				
		(0.0,0.0)	(0.05,0.03)	(0.20,0.12)	(0.30,0.16)	(0.40,0.22)
200	(0.00,0.00)	0.06	.	.	.	.
	(0.20,0.16)	0.74	0.75	0.82	.	.
	(0.40,0.30)	1.00	1.00	1.00	1.00	1.00
	(0.60,0.43)	1.00	1.00	1.00	1.00	1.00
500	(0.00,0.00)	0.05	.	.	.	.
	(0.20,0.16)	0.98	0.98	0.98	.	.
	(0.40,0.30)	1.00	1.00	1.00	1.00	1.00
	(0.60,0.43)	1.00	1.00	1.00	1.00	1.00

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

the artificial instruments, we compare our test to the well-established F-test for instrument strength (Stock et al., 2002; Stock and Watson, 2012; Mertens and Ravn, 2013). Table 3 contains the relative rejection frequencies at a nominal significance level of 5% and the corresponding F-statistics for exogenous instruments of different strength. The first column shows that the size of the F-test is close to its nominal level although the test tends to reject the null slightly too often when it is true and the sample is small. The rejection frequencies increase in instrument relevance, that is, when moving right across columns, for both sample sizes. However, the increase is substantially slower than for the LR-test (see first column of Table 2). The latter detects a relevant instrument with 100% probability if $T \geq 200$ and $\rho_1 \geq 0.30$, whereas the F-test does so only if $\rho_1 > 0.43$ for $T = 200$ or if $\rho_1 \geq 0.43$ for $T = 500$. In other words, the LR-test has more power against the null hypothesis of irrelevant instrument when the alternative is true.

The decrease in type-II error is useful for practical purposes. It implies that fewer relevant instrument are discarded. For $T = 200$ and $\rho_1 = 0.30$ the LR-test suggests keeping all candidate relevant instruments, whereas the F-test retains only 75%. For $T = 500$ and $\rho_1 = 0.16$ the LR-test finds 98% of relevant instruments, whereas the F-test detects only 55%. The advantage of having a test with more power becomes even more visible when departing from the 5% significance level for the F-test and using the stricter criterion of an F-statistic of 10, which is commonly used to shield against weak instrument problems. Table 3 suggests that between 50% and 60% of relevant instruments are then erroneously discarded in samples

Table 3: Relative rejection frequencies at nominal significance level of 5% of robust F-test for instrument relevance.

	Sample $T = 200$			Sample $T = 500$		
Relevance (β_1,ρ_1)	Rejection frequency	Frequency $F > 10$	Robust F-statistic	Rejection frequency	Frequency $F > 10$	Robust F-statistic
(0.00,0.00)	0.07	0.01	1.19	0.05	0.00	1.02
(0.20,0.16)	0.31	0.06	3.30	0.55	0.17	5.55
(0.30,0.23)	0.66	0.17	5.86	0.86	0.51	11.3
(0.40,0.30)	0.75	0.38	9.33	0.97	0.81	19.19
(0.60,0.43)	0.96	0.78	18.61	1.00	1.00	40.66

Notes: The table shows the relative rejection frequencies of robust F-tests for instrument strength at a nominal significance level of 5%, the relative frequencies that $F > 10$, and the average F-statistics, based on 500 replications for each instrument. Endogeneity is assumed to be absent, that is $\beta_2 = \rho_2 = 0$.

of 500 and 200 observations, respectively.

Another feature of the proposed framework is that it largely resolves the problem of weak instruments, provided that there is sufficient heteroskedasticity in the data. Table 4 displays the evaluation of models A-C using the MSE of the structural impulse responses as accuracy criterion. We normalize the MSE by those of model A and focus on the results based on a sample size of $T = 500$.[9] For a white-noise instrument ($\rho_1 = \rho_2 = 0$), models A and B yield essentially the same MSE. This implies that the use of a weak instrument is unproblematic for inference if the data contain changes in volatility and if they are used for identification. For relevant and exogenous instruments, that is moving south across rows, the heteroskedastic proxy-VAR systematically yields the smallest MSE for all variables and parameterizations compared to both competing models. These gains increase with instrument relevance and are substantial. For instruments with 43% correlation with the monetary shock, the improvement relative to model B is 27% on average across parameters. Model C performs extremely poorly in all cases. Given that the variances of the instrument and of the other endogenous variables are time-varying in our setup, fitting a standard proxy-VAR that does not account for heteroskedasticity leads to serious distortions in the estimates of the structural parameters, irrespective of the relevance of the exogenous instrument. Overall these results suggest that the explicit modeling of volatility changes when they are a feature of the data and using the information of a valid proxy improves structural inference in SVARs.

[9]Appendix A.2 shows that the results are robust to changes of the propagation horizon and sample size.

Table 4: Comparison of MSE of impulse responses to monetary policy shock.

Relevance (β_1,ρ_1,F)	Endogeneity (β_2,ρ_2)														
	(0.0,0.0)			(0.05,0.03)			(0.20,0.12)			(0.30,0.16)			(0.40,0.22)		
Model	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}	θ_{11}	θ_{21}	θ_{31}
(0.0,0.0,1.3)															
Model A	1.00	1.00	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model B	1.00	0.99	1.00	.	.	.	.	.	.	.	.	.	.	.	.
Model C	82.30	50.48	140.88	.	.	.	.	.	.	.	.	.	.	.	.
(0.20,0.16,6.9)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.	.	.	.	.	.
Model B	1.02	1.10	1.01	1.01	0.92	0.98	0.86	0.24	0.79	.	.	.	.	.	.
Model C	42.45	21.96	54.48	40.79	18.65	50.87	25.03	9.79	26.27	.	.	.	.	.	.
(0.40,0.30,23.6)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.05	1.37	1.05	1.04	1.09	1.02	0.85	0.27	0.80	0.50	0.11	0.45	0.36	0.07	0.30
Model C	34.04	15.35	48.45	33.35	10.41	46.90	22.40	3.64	28.69	10.47	2.41	11.88	6.23	1.98	5.86
(0.60,0.43,49.0)															
Model A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model B	1.06	1.68	1.08	1.07	1.35	1.08	0.96	0.37	0.95	0.77	0.18	0.75	0.51	0.10	0.47
Model C	32.98	16.26	49.49	32.80	11.16	48.94	26.42	3.15	37.99	18.60	2.14	25.75	10.33	1.68	12.97

Notes: The table shows the cumulated MSE of fitted models (1)-(3) relative to model (1) for a propagation horizon up to $h = 25$ and sample size $T = 500$. Each entry is based on 500 replications of each simulation design. Dots (.) denote combinations of values for β_1 and β_2 that produce lower correlations between the instrument s_t and the target structural shock ε_t^r than between the instrument and the endogenous structural shock ε_t^z. These cases are not taken into account in the analysis.

This conclusion also holds for slightly endogenous instruments ($\rho_2 = 0.03$) if the proxy is sufficiently relevant. When the correlation between the instrument and the monetary shock is larger than $\rho_1 = 0.16$, model A still and consistently yields the smallest MSE. When the endogeneity increases further, the estimation precision of model A deteriorates considerably. Then, model B, which ignores the misspecified instruments for identification, yields more precise estimates. This finding underscores the importance of being able to test for instrument exogeneity. As before, model C performs worst in all cases.

Summarizing the simulation results, both LR-tests are helpful tools to assess the validity of instruments. Relevant instruments are detected reliably already in small samples at the 5% significance level. Moreover, the LR-test has more power than the widely used F-test. A detection of endogeneity requires somewhat larger samples and higher correlations between the instrument and the non-monetary shocks. The rejection frequencies suggest using the 10% significance level. Regarding the estimation precision, the heteroskedastic proxy-VAR recovers the true model best, even in cases of slightly endogenous instruments. As endogeneity increases, a standard heteroskedastic SVAR not using the instrument performs better and the heteroskedastic proxy-VAR yields seriously distorted estimates. This stresses the importance of having a means for evaluating instrument exogeneity. Finally, both models using time-varying volatility for identification yield sharper inference than a standard proxy-VAR, and the use of heteroskedasticity largely eliminates the problem of weak instruments.

4 Monetary policy analysis in a heteroskedastic proxy-VAR

We use our framework to provide new – and in light of the Monte Carlo evidence potentially sharper – estimates of the impact of monetary policy on the macro-economy. Our baseline model consists of three endogenous variables and an instrument for monetary policy shocks in the vector $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$. We use the first difference of the log of industrial production as a measure of real economic activity Δx_t, the federal funds rate as the monetary policy indicator ff_t, and a measure of corporate bond spreads ebp_t.[10] For the latter, we employ the excess bond premium of Gilchrist and Zakrajšek (2012) which approximates the tightness of credit markets. Caldara and Herbst (2019) show that it is important to control for the endogenous response of monetary policy to financial conditions to identify monetary policy shocks. Following the same authors, we use log differences of industrial production and include six lags of the endogenous variables to account for the persistence in the data. As an instrument for the latent monetary policy shocks, we take the narrative-based measure of unexpected changes in the intended fed funds rate of Romer and Romer (2004), rr_t.[11] This proxy starts in 1969M1 and, thereby, is the longest available potential instrument provided by the literature.

We estimate the VAR on monthly frequency data within the sample 1973M1 to 2007M6. The start is dictated by the availability of data on the excess bond premium, while the end is chosen such as to ensure that our sample is not affected by the zero lower bound or by unconventional monetary policy. We perform an extensive sensitivity analysis of the main results. We change the number of lags as well as states, the sample period, the monetary policy indicator, and the transformation of the variables. We also estimate a model including producer and commodity prices. The robustness results are summarized in Appendix A.3 and show that our main results hold.

[10] Industrial production is series $INDPRO$ downloaded from the FRED database of the Federal Reserve Bank of St. Louis, the federal funds rate is series $FEDFUNDS$ from the same database. The excess bond premium of non-financial firms is identical to the data used in Gilchrist and Zakrajšek (2012) and was downloaded from Gilchrist's website.

[11] We use the updated version of the original Romer and Romer (2004) constructed by Wieland and Yang (2016) downloaded from Wieland's webpage.

4.1 Model specification

An important choice in our framework is the way changes in volatility are modeled. The functional form of the model affects the likelihood and thereby the estimators and tests. Therefore, we perform an extensive model comparison before turning to the model based inference. As candidates we model heteroskedasticity through a smooth transition in variances using either a 12-month trailing moving average of industrial production or time as the transition variable, an exogenous break point iterating over all potential break points in the sample, a multivariate GARCH process, and a Markov switching framework with two states.[12] Appendix A.3 describes all models formally.

Table 5: Model selection.

Reduced form models	$log(L_T)$	SC	AIC	HQ
White noise residuals	-692.91	2046.78	1605.81	1780.32
Smooth transition in variances (IP)	-452.29	1637.66	1148.59	1342.13
Exogenous breakpoint	-400.33	1527.73	1042.67	1234.63
Smooth transition in variances (time)	-293.41	1319.90	830.82	1024.36
GARCH residuals	-286.41	1317.91	820.81	1017.53
Markov switching	-212.60	1158.27	669.20	862.74

Note: L_T denotes the likelihood function evaluated at the optimum, $AIC = -2log(L_T) + 2f$, $SC = -2log(L_T) + log(T)f$ and $HQ = -2log(L_T) + 2f \times log(log(T))$, where f is the number of free parameters and T the number of observations.

Table 5 shows specification statistics for a linear model and the five alternative volatility models. Information criteria are shown to work well for judging the performance of MS models (Psaradakis and Spagnolo, 2006), whereas standard tests are problematic for this purpose as some parameters might not be identified under the null hypothesis of a smaller number of states than under the alternative (Hansen, 1992). Lütkepohl and Schlaak (2018) show that the AIC succesfully chooses between alternative volatility models. The models are ordered descending according to their log-likelihood values. This ranking coincides with that implied by all three information criteria, leaving little amigbuity about the relative performance of the alternative models. The table conveys two important results. First, the linear model is dominated by all models that allow for changes in volatility. This is strong evidence in support of the assumption of heteroskedasticity. In this case, using any time-varying volatility model estimates structural impulse

[12] We also estimate smooth transition in variances models using either 6 or 24-month moving averages of industrial production. Both models perform worse than the 12-month version so we do not report them in the table.

Table 6: Test for VAR invertibility.

Equation	$\Delta\ln(\text{IP})$	Fed funds	EBP	Instrument
p-value	0.21	0.56	0.83	0.65

Notes: The table shows p-values for a robust F-statistic testing the null hypothesis that the coefficients on six lags of s_t are jointly equal to zero in each of the VAR equations of the Markov switching proxy-VAR.

responses more precisely than a linear model (Lütkepohl and Schlaak, 2018). Second, the Markov switching model is clearly preferred over the other heteroskedastic models according to both the log-likelihood and all three information criteria. Lütkepohl and Schlaak (2018) show that the Markov switching model is usually the best choice even in cases where the volatility specification does not coincide with the data generating process. Modeling changes in volatility through a latent variable gives full voice to the data, reducing the risk of misspecification of the transition variables, functions or break points.

Based on Table 5 we choose the Markov switching model, but we show that our result are robust to the choice of the volatility model. In Section 4.5, we use smooth transition models. They are popular in the VAR literature as they provide economic insights into the driving forces of the states. In Appendix A.3, we show that our results also hold in a three-state Markov switching model. We favor a two-state model as the baseline since two states are economically more intuitive and easier to interpret. Moreover, a two-state model leads to more stable and precise estimates given that the third state contains only few observations.

Finally, to see whether the VAR model is invertible, we exploit that the augmented model (8) naturally lends itself for Granger-causality tests. Following Stock and Watson (2018), we test the null hypothesis that the six lags of the instrument are jointly equal to zero in each of the VAR equations. Table 6 indicates that there is no statistically significant evidence against the hypothesis of invertibility. This result is in line with the findings of Stock and Watson (2018), who do not reject the assumption of invertibility of the VAR model of Gertler and Karadi (2015), which is similar to our specification.

To facilitate the comparison with the literature using the narrative measure of Romer and Romer (2004), we set the autoregressive coefficients for the instrument, that is the respective elements of $\Gamma(L)$ in (10), to zero in the subsequent analysis. In Table A.7 in Appendix A.3 we show that our main results do not change when we leave the autoregressive and constant part of the model fully unrestricted.

4.2 Volatility regimes and identification

Table 7 reports the estimated state-dependent reduced form covariance matrices indicating whether the model detects switches in volatility, which are one important element in our identification and testing strategy. This information also helps us interpret our endogenously and agnostically identified regimes. Clearly, there are increases in volatility in state 2 for all residuals. The error variances in the equations for industrial production, the federal funds rate, the excess bond premium and the instrument increase by factors of approximately 3, 54, 10 and 20, respectively. In particular, the volatility of the error in the interest rate equation changes strongly across regimes. We read this as further evidence that the sample is characterized by changes in volatility. Moreover, the model seems able to detect and separate them.

The table also shows that the covariances increase (in absolute value) in state 2 as well, and by larger factors than the variances. These changes in the covariances illustrate the idea behind identification through heteroskedasticity. In a period where interest rates are highly volatile, we learn more about the relation between the federal funds rate and economic activity as the covariance between both temporarily increases. Monetary policy shocks are then more likely to occur and can be used as a 'probabilistic instrument' (see Rigobon, 2003) to trace out the response of production.

Table 7: Estimated state covariance matrices ($\times 10^3$) of reduced form model (8) with $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$.

State 1: $\tilde{\Sigma}_1$	State 2: $\tilde{\Sigma}_2$
$\begin{bmatrix} 281.54 & & & \\ 10.60 & 32.96 & & \\ -1.03 & 0.40 & 22.00 & \\ 0.18 & 7.73 & -0.92 & 21.08 \end{bmatrix}$	$\begin{bmatrix} 864.71 & & & \\ 497.08 & 1835.58 & & \\ -46.06 & 27.88 & 194.85 & \\ 206.86 & 415.70 & 57.08 & 413.46 \end{bmatrix}$

To achieve identification from a statistical point of view and to be able to test the validity of the external instrument, we need significant and differential changes in the volatility of the structural innovations μ_t. Table 8 shows the estimated variances of the structural model (10) with unrestricted β in state 2, which are contained in Λ_2. Given the restrictions on D and that the instrument is ordered last in z_t, λ_{42} captures the change in the variance of the noise in the measurement of the instrument. As the ordering of the remaining $\lambda_{\cdot 2}$s is arbitrary, we simply order them from largest to smallest. All of these three estimates are significantly larger than one, implying that the volatility of all structural shocks increases when switching from state 1 to state 2. Thus, together with the evidence in Table 7, we label state 2 the high volatility state. Identification requires that the variance shifts are all distinct from each other. This seems to be the case according to their estimated standard errors, which suggests that the decomposition in (11) is locally unique and can be used to test the validity of the instrument.

Table 8: Estimates and standard errors of relative variances.

Parameter	Estimate	Standard error
λ_{12}	55.57	10.16
λ_{22}	8.90	3.03
λ_{32}	2.62	0.53
λ_{42}	16.79	4.80

Notes: The standard errors are obtained from the inverse of the negative Hessian evaluated at the optimum of the structural model (10) with $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$.

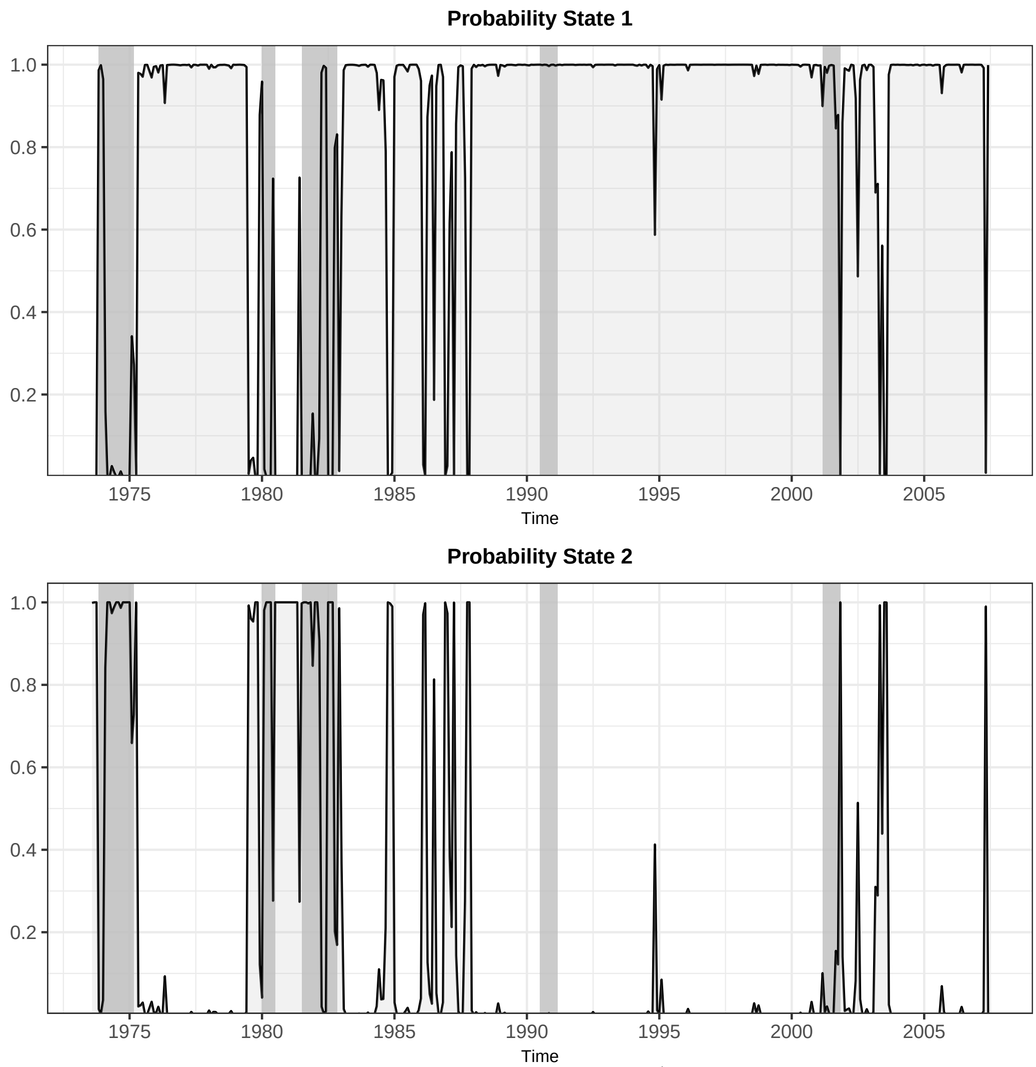
To develop an economic notion about the statistically identified regimes we plot the smoothed state probabilities in Figure 1. The upper part corresponds to state 1 and the lower part to state 2. State 1 dominates the sample, in particular its second half. The model detects a long spell of low volatility during a period that is often referred to as the 'Great Moderation' in the 1990s and 2000s with stable growth and inflation under the chairmanship of Alan Greenspan. The high volatility regime appears more often during the first part of the sample. Many of the spikes in the probability of being in state 2 are associated with specific events in the economic history of the U.S. For instance, there are peaks around the energy crisis and the subsequent recession in the middle of the 1970s and at the beginning of the 1980s. There is also a longer-lasting switch to state 2 which coincides with the chairmanship of Paul Volcker at the end of the 1970s and the first half of the 1980s. In the second part of the sample, there are peaks around the burst of the dot-com bubble in 2001, the 9/11 attacks, and the subsequent recession. Overall, this short narrative, while only suggestive, indicates that the endogenously determined volatility regimes capture relevant developments in the U.S. real economy and in the conduct of monetary policy in our sample.

4.3 Instrument validity

We now use these significant and distinct changes in the variances of the structural innovations to test the validity of the instrument. Given validity, we combine the information in the instrument with that in the second moments of the data and estimate the dynamic effects of monetary policy shocks.

First, we test for exogeneity. We leave β_1 unrestricted, thereby ordering the

Figure 1: Smoothed state probabilities.



Notes: The figure shows the smoothed state probabilities $\hat{\xi}_{mt|T}$ for $m = 1$ in the upper panel and for $m = 2$ in the lower panel of model (10) with $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$. The shaded vertical bars mark recession periods defined by the NBER.

monetary shock first. The null hypothesis is that the instrument is exogenous to the non-monetary policy shocks, $H_0 : \beta_2 = \beta_3 = 0$, against the alternative that the instrument is endogenous. Table 9 shows that the data do not reject the assumption of exogeneity. The LR-statistic is quite small and the p-value not close to conventional significance levels. Thus, the instrument passes the first stage.

To assess its relevance, we test the null hypothesis that the instrument is unrelated to all structural shocks, $H_0 : \beta = 0$, against the alternative that it is

Table 9: Instrument validity.

	Exogeneity	Relevance
LR statistic	0.04	43.32
p-value	0.97	0.00
Restrictions	2	1

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

significantly related to at least one structural shock. If the null is rejected, this will be the monetary policy shock given the first stage result and that the instrument is constructed to have a high correlation with the monetary shock and a low (zero) correlation with the other shocks. The table shows that the instrument is highly relevant. The null is rejected at the 1% significance level. We conclude that the narrative-based measure of Romer and Romer (2004) is also relevant and, thus, a valid instrument for monetary policy shocks.

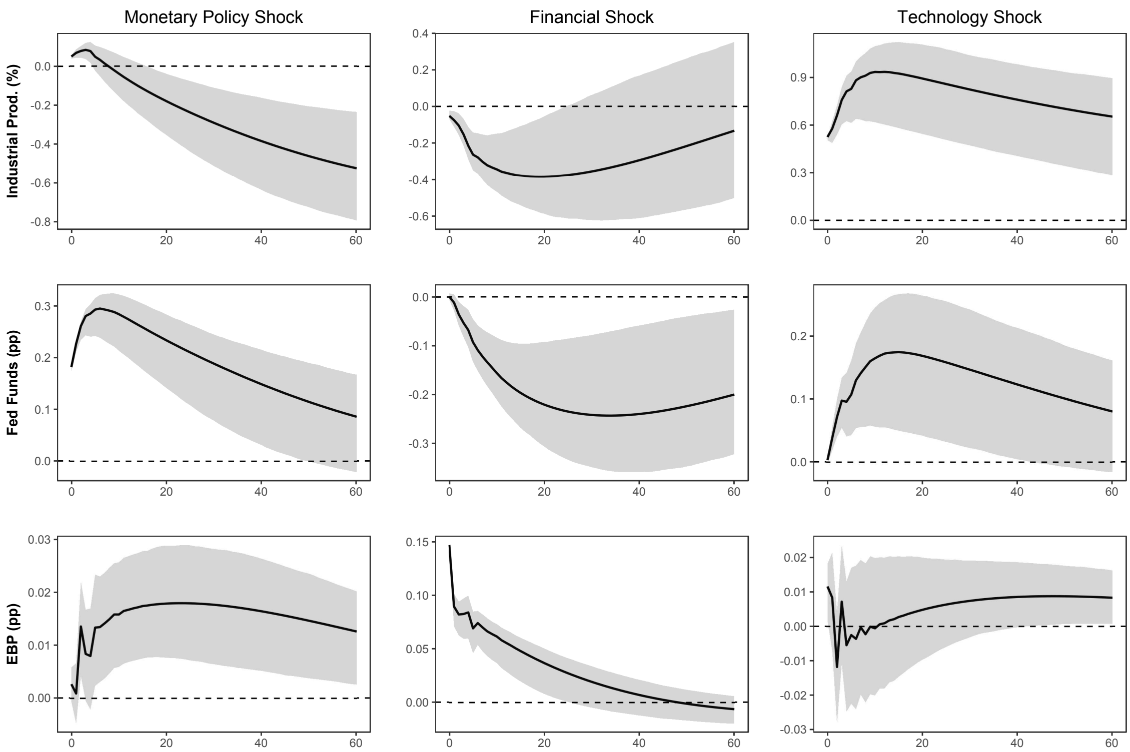
4.4 Dynamic effects and importance of monetary shocks

We now estimate the dynamic effects of monetary shocks and quantify their economic importance for output and credit spread fluctuations. Based on the testing sequence, we leave β_1 unrestricted and set $\beta_2 = \beta_3 = 0$. This implies that the estimation combines the information contained in the instrument and in the second moments of the data for identification of a heteroskedastic proxy-VAR.

Figure 2 shows the impulse responses to all three shocks in columns on the endogenous variables in rows. The inclusion of the proxy into the model is a key advantage over traditional identification through heteroskedasticity where a main challenge is the economic labeling of the statistically identified shocks (Herwartz and Lütkepohl, 2014). Due to our restrictions on β, we can clearly label the monetary shock, which is pinned down in the first column of D. A one standard deviation monetary surprise corresponds to an increase in the federal funds rate by 18 basis points. According to the 95 percent confidence bands, credit spreads increase significantly a few months after the shock and remain elevated for several years, before gradually returning back to trend. Economic activity declines significantly a year and a half after the occurrence of the shock. The response bottoms toward the end of the propagation horizon with a cumulative effect of –0.5 percent.

Quantitatively, the dynamics of real activity are similar to those implied by the hybrid VAR models of Romer and Romer (2004) and Coibion (2012), who include

Figure 2: Impulse responses for heteroskedastic proxy-VAR model.

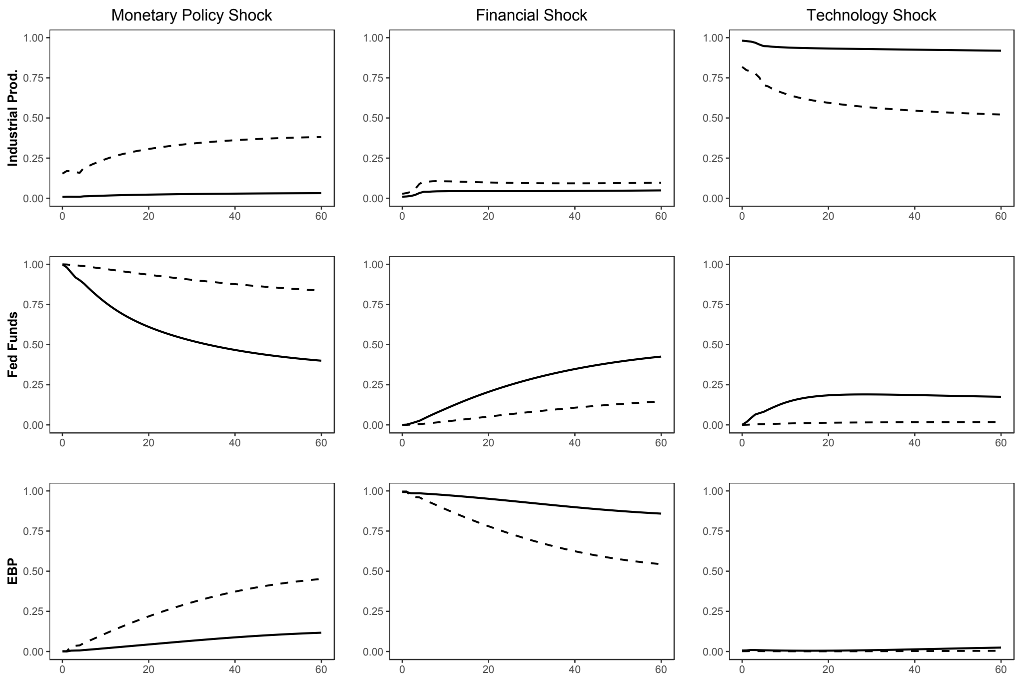


Notes: The figure shows the impulse responses to one standard deviation shocks in state $m = 1$ of the heteroskedastic proxy-VAR(6) model with $M = 2$ states for $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$. The sample is 1973M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95 percent pointwise confidence intervals based on 5,000 bootstrap replications.

the cumulated monetary policy surprise measure directly into a VAR together with industrial production and producer prices. They document a 2.9 percent trough effect for a 100 basis points shock. Qualitatively, our results differ from their estimates which suggest a short-lived increase in real activity but then a quick and significant decline after about six months. Romer and Romer (2004) trace the initial hump back to a single observation (a large negative surprise in the intended fed funds rate on April 1980 coupled with a strong decline in production) and the sampling error due to this extreme event. While we also obtain the hump, the following decline is more sluggish and more persistent. Output falls below trend only after a year and a half. These dynamics are similar to Ramey (2016).

We next assess the economic importance of monetary policy shocks for output and credit spread fluctuations. The regime-specific forecast error variance decompositions in Figure 3 show that the contribution of monetary shocks to the variability of the endogenous variables is highly state-dependent. In the high volatility regime, monetary shocks account for up to 40 percent of the variance of production and spreads at longer horizons. In the low volatility regime, they each explain less than 10 percent. They also account for a much larger share of the variance in the federal funds rate in the high volatility regime than in the low volatility regime.

Figure 3: Variance decompositions for heteroskedastic proxy-VAR.



Notes: The figure shows the regime-specific forecast error variance decompositions (solid line - state 1; dashed line - state 2) for the structural shocks in columns on the endogenous variables in rows for the heteroskedastic proxy-VAR(6) model with $M = 2$ states and $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$. The sample is 1973M1-2007M6 and the instrument for monetary policy shocks is the narrative-based measure of Romer and Romer (2004).

We briefly discuss the effects of the other two structural shocks, which are mainly identified using the changes in volatility and yet need to be labeled (if they are of interest to the researcher). Our framework also simplifies this task compared to traditional identification through heteroskedasticity. The inclusion of a valid proxy for monetary policy into the model gives an economic interpretation to the main shock of interest and separates it from the remaining shocks. Therefore, the latter are easier to label. This is reflected in relatively clear sign patterns of the impulse responses for the other two shocks and in the forecast error variance decomposition, which both suggest two simple labels.

The second shock accounts for virtually all of the variability in the excess bond premium upon impact and for more than 60 percent in both states in the long-run. Thus, we label it a financial shock. An exogenous 15 basis points increase in credit spreads leads to an immediate contraction in real activity (see Figure 2). This is followed by a hump-shaped negative response, with a trough of –0.4 percent, and a gradual return back to trend after about three years. The monetary authority responds by lowering the policy rate by up to 20 basis points after two years to offset the adverse effect of tighter financial conditions on production.

The remaining shock accounts for a minimum of 80 percent of the variance of industrial production on impact. At longer horizons, it explains more than half of the variance in the high volatility state and more than 90 percent in the low volatility state. Thus, we label it a technology shock. In response, industrial production jumps up immediately and reaches a peak after one year. Monetary policy aims to counter the expansion by raising the interest rate by 15 basis points after one year, which thereafter reverts to trend together with real activity. The excess bond premium hardly responds to the shock. Toward the end of the horizon there is a mildly positive response, consistent with the monetary tightening.

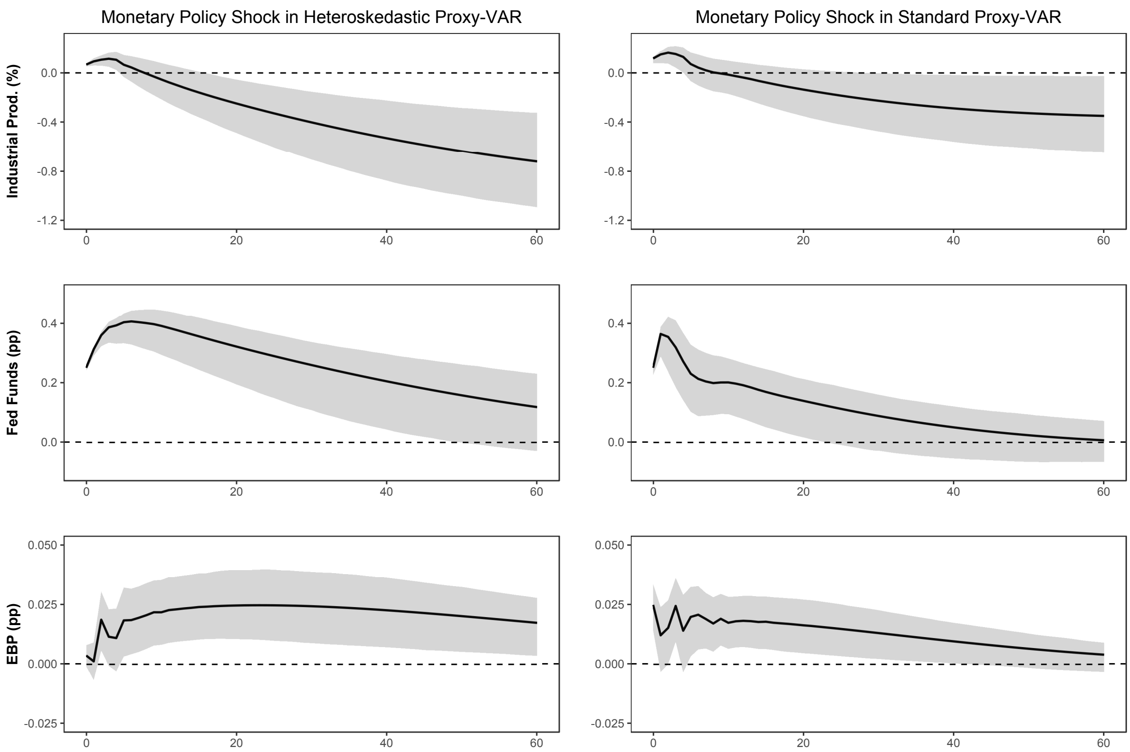
In light of the impulse response analysis and the results from the testing sequence, it is not surprising that we find similar effects of monetary shocks as Romer and Romer (2004). The tests show that the narrative-based measure is a valid instrument for latent monetary policy shocks. The impulse response analysis shows that the instrument is neither endogenous to economic activity shocks nor to financial shocks. Thus, a main contribution of our approach is to increase the confidence in estimates based on this instrument. The results suggest that it can be used in models like ours or in the Bayesian model of Caldara and Herbst (2019) to reliably estimate the dynamic effects of monetary policy shocks.[13]

The simulation study suggests that another advantage of modeling heteroskedasticity – which is present in our sample as shown in Section 4.2 – is that the information contained in the second moments helps identifying the structural model and, hence, the effects of monetary policy more accurately. This is corroborated by Figure 4, which compares the impulse responses to a monetary policy shock from the heteroskedastic (left column) to those from a standard proxy-VAR (right column). The shock is scaled to 25 basis points for comparison. Qualitatively, both models yield the same conclusions. Corporate bond spreads increase and production declines.

However, quantitatively and in terms of economic significance, there are notable differences. In the heteroskedastic proxy-VAR, the monetary shock is more

[13]The advantages of using the measure as an instrument and accounting for measurement error instead of using it directly as a variable in a regression or VAR are discussed in Stock and Watson (2012) and Mertens and Ravn (2013).

Figure 4: Comparison of heteroskedastic and standard proxy-VAR.



Notes: The figure shows the impulse responses to a monetary policy shock of 25 basis points in a heteroskedastic (left column) and standard proxy-VAR (right column) on the endogenous variables in rows. The model contains $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$, the sample is 1973M1-2007M6 and the instrument for latent monetary shocks is the narrative-based measure of Romer and Romer (2004). The shaded bands denote 95 percent pointwise confidence intervals based on 5,000 bootstrap replications.

hump-shaped, reaching a peak only after ten months, and estimated to be more persistent. The federal funds rate remains significantly above trend for about 48 months, whereas in the standard proxy-VAR it is indistinguishable from zero after roughly 24 months. This stronger and longer-lasting monetary contraction leads to a quicker, larger and more persistent drop in industrial production, which falls significantly below trend after a year and a half, declining cumulatively 0.7 percent. In contrast, in the standard model, the decline in economic activity is more sluggish, the effect is only borderline significant after three years, and the trough is only –0.4 percent. Similarly, the effect of the monetary shock on credit spreads is stronger, longer-lasting, and more statistically significant in the model using the time-varying volatility.

These differences are due to alternative uses of the existing information in the data. While the heteroskedastic proxy-VAR draws on both the instrument and changes in volatility for identification, the standard proxy-VAR discards the second piece of information. To see whether the assumption of heteroskedastic structural innovations is consistent with the data we now employ information criteria for the comparison of the *structural* models. Since the standard proxy-VAR point-identifies only one column, it is under-identified and its likelihood and information scores are the same as for the reduced form model. They are shown in Table 5. In contrast, the heteroskedastic proxy-VAR is over-identified, such that its likelihood deteriorates slightly (to –216.29) relative to the reduced form Markov switching model. Nevertheless, the information criteria favor the more parsimonious structural over the reduced form model. All three criteria improve; to 1135.62, 666.69, and 852.20 for the SC, AIC, and HQ, respectively. This reflects the earlier result from the LR-test for instrument exogeneity of not rejecting the over-identifying restrictions. More importantly, all information criteria clearly prefer the heteroskedastic over the standard proxy-VAR. We conclude that the former provides sharper inference because it exploits time-variation in second moments and that the latter underestimates the effects of monetary policy.

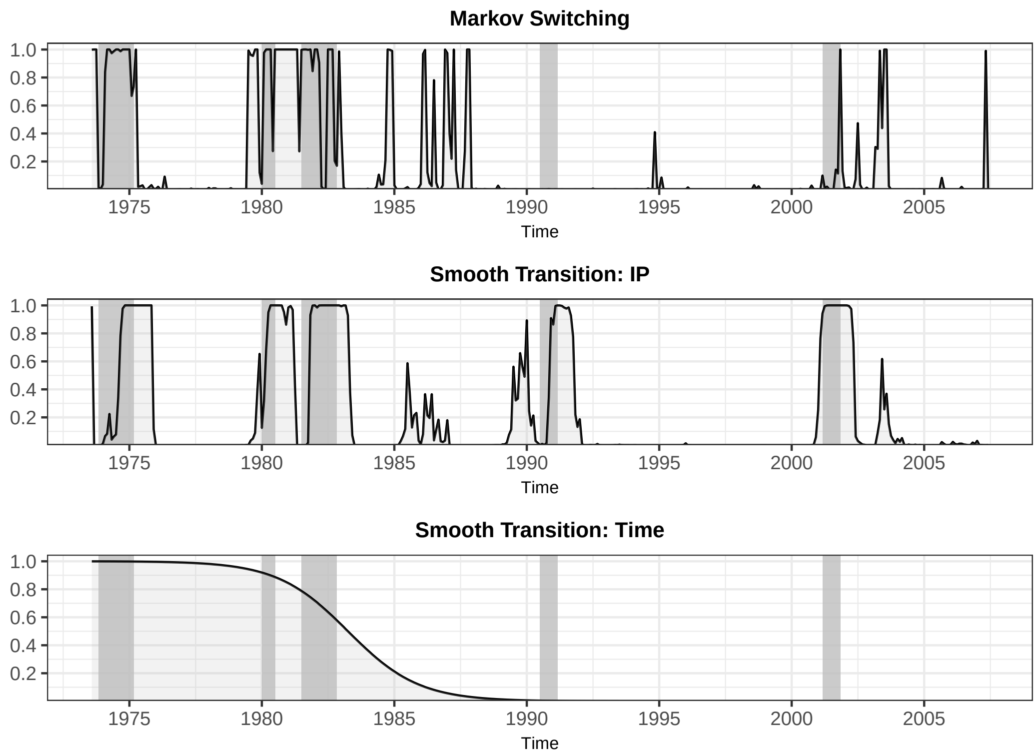
4.5 Smooth transition in variances

In this subsection, we assess whether our results are sensitive to using an alternative volatility model as both the test results and the impulse responses depend on the functional form of the heteroskedasticity. We focus on the smooth transition in variances specifications as they add economic insight to the drivers of the states. We consider the version with a 12-month trailing moving average of industrial production and with time as transition variable to see whether there is a systematic relation between high volatility and recessions and a transition in shock variances from the Volcker to the Greenspan area, respectively.

Figure 5 compares the volatility states across models. The smooth transition based on industrial production estimates roughly similar states as the Markov switching model. The correlation between the two states is 0.36. Moreover, it captures most of the NBER recessions, although with some delay which reflects the use of a trailing moving average as transition variable. The model based on time also does an acceptable job. It matches the transition in the Fed Chairmanship occurring in 1987. The correlation with the Markov switching state is 0.43.

Table 10 shows the estimated structural variances and their standard errors to assess whether the smooth transition in variances adds identifying information to the model. The variance increases in state 2 have the same ranking in both models and as in the Markov switching model. The monetary policy shock and the measurement error shift most in variances, followed by the financial and technology shock, although the variances tend to be less precisely separated between regimes

Figure 5: Volatility states 2 of Markov switching and smooth transition models.



Notes: The figure shows the state 2 probability of heteroskedastic proxy-VARs, using Markov switching (upper panel) and smooth transition in variances based on a 12-month moving average of industrial production (middle panel) or on time (lower panel) as transition variable. The model is $z_t = [\Delta x_t, ff_t, ebp_t, rr_t]'$. The shaded vertical bars mark recession periods defined by the NBER.

than in the Markov switching model. Importantly, the structural shocks are all identified precisely according to the standard errors.

Therefore, we proceed to test whether the instrument is valid. Table 11 shows that this is the case. Both models come to the same conclusion as the Markov switching model. The narrative-measure of Romer and Romer (2004) is exogenous and relevant. Finally, we study the impulse responses implied by the smooth transition models. Figure 6 shows that they are not statistically distinguishable from those of the Markov switching model, although both imply smaller effects more similar to the proxy-VAR that does not exploit the heteroskedasticity. The latter finding reflects that both models are less successful in separating the volatility regimes, using less of the identifying information in the data and yielding blunter inference. Overall, however, the subsection shows that our framework and results do not depend on a specific volatility model.

Table 10: Relative structural variances for smooth transition models.

Transition variable	IP		Time	
Parameter	Estimate	S.E.	Estimate	S.E.
λ_{12}	22.41	1.99	69.61	13.32
λ_{22}	3.01	0.34	2.93	0.51
λ_{32}	2.25	0.26	0.94	0.19
λ_{42}	6.18	0.59	11.14	2.64

Table 11: Test for instrument validity based on smooth transition in variances.

Transition variable	Exogeneity p-value	Relevance p-value
Industrial production	0.313	0.000
Time	0.715	0.000

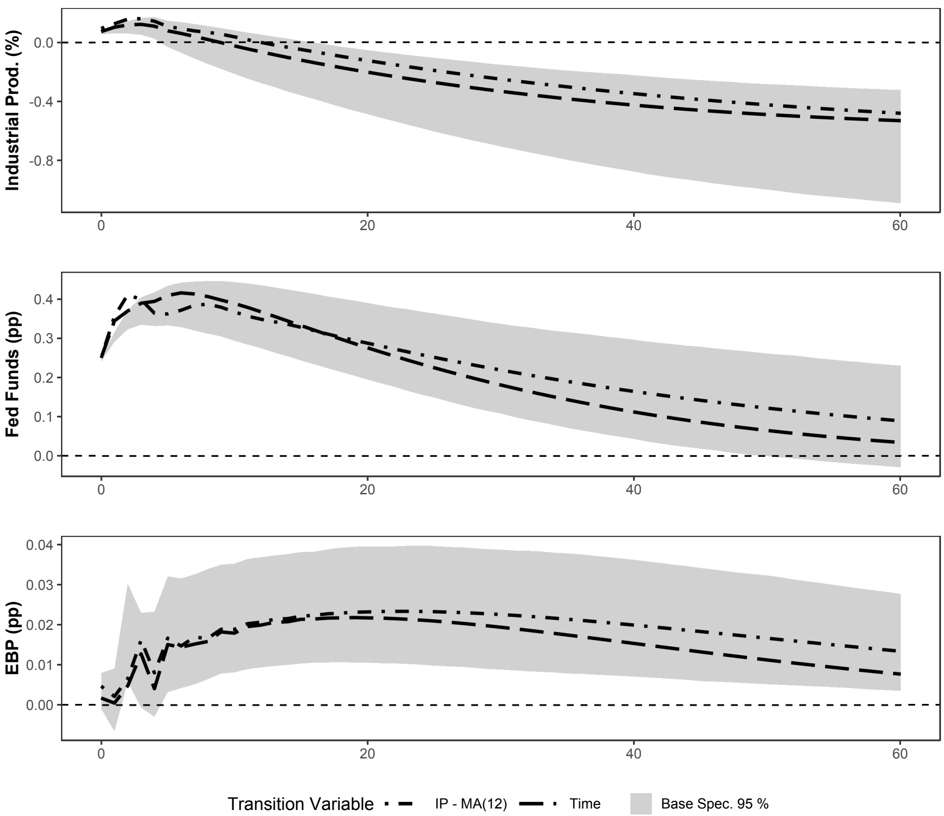
Notes: The table shows the p-values of LR-tests for the exogeneity and relevance of the measure of Romer and Romer (2004) as instrument s_t for monetary policy shocks in smooth transition in variances models using different transition variables. The model is $z_t = [\Delta x_t, ff_t, ebp_t, s_t]'$ and sample period is 1973M1-2007M6.

4.6 Testing alternative proxies for monetary shocks

In this section, we test and compare alternative measures of monetary surprises proposed in the literature to study the effects of monetary policy. In addition to the narrative measure of Romer and Romer (2004), we consider the identified monetary shocks from the SVAR of Bernanke and Mihov (1998) and monetary surprises identified using high(er) frequency data. For the latter, we employ measures derived from changes in federal funds futures data around policy announcements using a daily window (see Barakchian and Crowe, 2013), a 30-minutes window (see Gertler and Karadi, 2015), and a 30-minutes window including further cleaning of the surprises by regression on a range of control variables (see Miranda-Agrippino and Ricco, 2017). We consider the potential instruments one at a time.

To establish a level playing field and to facilitate a clean comparison, we use a common sample period for the evaluation of the proxies although they are available for slightly different periods. As most high(er) frequency proxies start only in the 1990s, we use the same sample (and model) as Caldara and Herbst (2019), which is 1994M1-2007M6. The first row in Table 12 shows the test results when re-estimating the model using the narrative-based measure on the shorter sample period. The conclusion from above based on the longer sample hold. The

Figure 6: Comparison of smooth transition with baseline Markov switching model.



Notes: The figure shows impulse responses to a 25 basis points monetary policy shock of the smooth transition in variances SVAR using industrial production as transition variable (dashed line) or time (dash-dotted line). The shaded area denotes 95 percent pointwise confidence intervals based on 5,000 bootstrap replications of the baseline Markov switching proxy-VAR(6).

instrument is valid. The next row shows that the same assertion applies to the model-based measure.

The picture is more mixed for the instruments based on high-frequency data. While none appear to suffer from endogeneity, only the plain changes in the fourth federal funds futures in a 30-minute window around policy announcements are a relevant proxy.[14] Both the instrument using daily data and the cleaned instrument do not meet the relevance condition according to the LR-test and conventional significance levels. One possible explanation for this finding is that the former

[14]Bertsche and Braun (2017) aim at assessing the estimated moment conditions implied by the instruments of Romer and Romer (2004) and Gertler and Karadi (2015) in non-nested models using a stochastic volatility framework. However, the distribution and properties of their test statistic are unclear as the parameters under the null hypothesis are estimated and thus random variables.

Table 12: Testing alternative proxies.

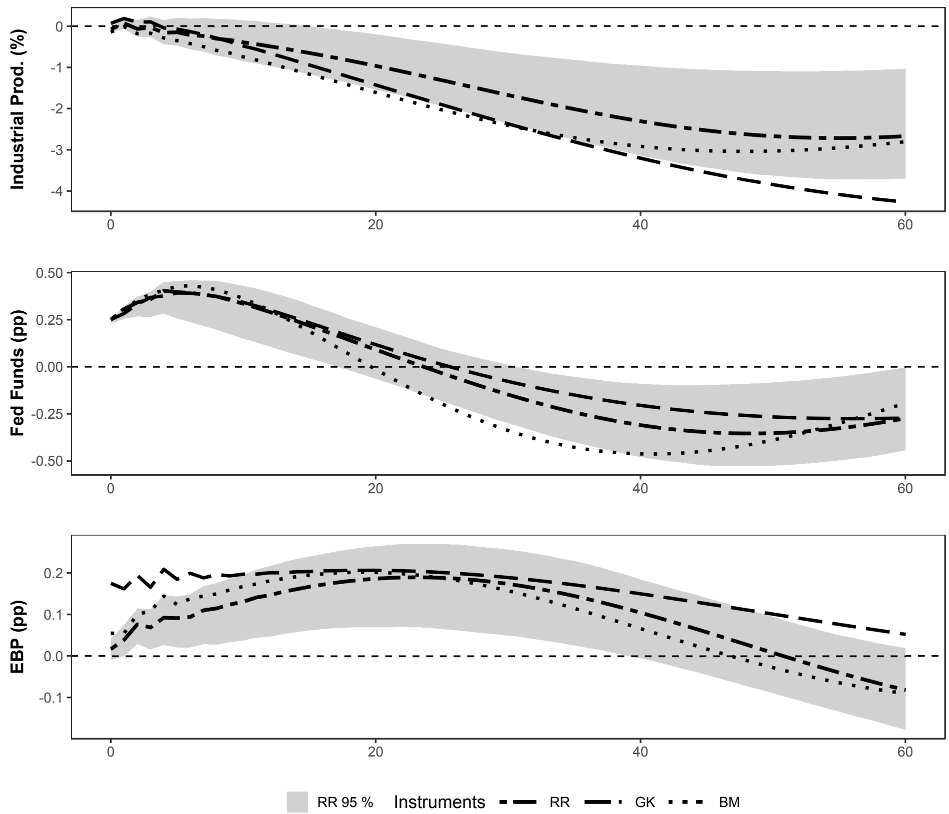
Instrument	Exogeneity p-value	Relevance p-value
Narrative-based	0.427	0.003
Model-based	0.830	0.000
High(er) frequency data		
Daily data	0.485	0.302
30-minute window	0.599	0.002
30-minute window and cleaned	0.897	0.238

Notes: The table shows the p-values of LR-tests for the exogeneity and relevance of different instruments s_t in the model with $z_t = [\Delta x_t, ff_t, ebp_t, s_t]'$, testing them one at a time. The sample period is 1994M1-2007M6. The narrative-based measure is of Romer and Romer (2004), the model-based measure is of Bernanke and Mihov (1998), and measures based on high(er) frequency data are taken from Barakchian and Crowe (2013), Gertler and Karadi (2015) and Miranda-Agrippino and Ricco (2017), respectively.

instrument may be too noisy and the latter stripped of too much relevant information through the regression on further control variables. These results, of course, are conditional on the model specification and sample used.

Focusing on the valid instruments for the given model and sample, Figure 7 compares the implied effects of a monetary policy shock on the endogenous variables. The solid lines show the estimates using the narrative-based measure together with 95 percent confidence bands as a baseline for the comparison. Relative to the estimates using the full sample, the initial short increase in industrial production vanishes in the shorter sample starting in January 1994, consistent with the argument of Romer and Romer (2004) that this hump is related to the April 1980 observation. The other two proxies produce similar effects. Over most of the propagation horizon, the responses are not distinguishable from those implied by the narrative-based measure. The monetary shock identified using the high-frequency proxy has a larger effect on the excess bond premium upon impact and on industrial production toward the end of the propagation horizon. Overall, however, the three valid proxies lead to similar conclusions. This is reassuring and suggests that the main findings of the previous sections based on a longer sample provide a reasonable description of the effects of monetary policy. A natural next step is to include several valid proxies for monetary shocks simultaneously into the model. However, this is beyond the scope of the paper and left for future research.

Figure 7: Responses for heteroskedastic proxy-VAR using different instruments.



Notes: The figure shows the impulse responses to a 25 basis points monetary policy shock. The sample is 1994M1-2007M6 and the different instruments, using one at a time, are the narrative measure of Romer and Romer (2004) (solid line with shaded 95 percent pointwise confidence intervals based on 5,000 bootstrap replications), the high-frequency proxy of Gertler and Karadi (2015) (dashed line), and the model-based measure of Bernanke and Mihov (1998) (dotted line).

5 Conclusions

We propose an econometric framework in the form of a structural vector autoregression that combines the information contained in an external instrument and in time-varying second moments of the data for identification of latent monetary policy shocks in the U.S. We show that the framework improves the identification of the structural model and leads to sharper inference. Moreover, it allows testing the validity of the chosen instrument, thereby increasing the credibility and reliability of the estimation results for policy analysis. Given sufficient heteroskedasticity in the data the framework also largely dispenses the proxy-VAR approach from problems arising from weak instruments. Finally, it facilitates an economic interpretation of the structural shock of interest, which is not only identified statistically through heteroskedasticity but also through prior economic reasoning contained in the instrument.

We apply the framework to test the validity of using the narrative measure of monetary surprises of Romer and Romer (2004) as an instrument for monetary policy shocks. We find that it is a valid instrument in our model and sample. We use it and combine it with the heteroskedasticity in the data to provide new and potentially sharper estimates of the dynamic effects of monetary policy on the macro-economy. We find that a surprise monetary contraction of 25 basis points in the federal funds rate leads to a significant increase in corporate bond spreads and to a significant decline in real economic activity of cumulatively 0.7 percent. In contrast, a standard proxy-VAR that does not use the time-variation in second moments implies substantially smaller effects. The results further suggest significant changes in the volatility of monetary shocks over time and that the shocks explain a large share of real and financial fluctuations in the 1970s and 1980s, but only a small share during the Great Moderation under the chairmanship of Alan Greenspan.

Finally, we evaluate different proxies for monetary policy proposed in the literature and find that instruments based on intra-daily data that are not further cleaned (Gertler and Karadi, 2015) and instruments from time-series models (Bernanke and Mihov, 1998) are also valid in our model and sample. They lead to qualitatively and quantitatively similar results as the narrative-based proxy.

References

An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models, *Econometric Reviews* **26**(2-4): 113–172.

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (forthcoming). Narrative sign restrictions for svars, *American Economic Review* .

Bacchiocchi, E. and Fanelli, L. (2015). Identification in structural vector autoregressive models with structural changes, with an application to us monetary policy, *Oxford Bulletin of Economics and Statistics* **77**(6): 761–779.

Barakchian, S. M. and Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data, *Journal of Monetary Economics* **60**(8): 950–966.

Bernanke, B. S. and Mihov, I. (1998). Measuring monetary policy, *The Quarterly Journal of Economics* **113**(3): 869–902.

Bertsche, D. and Braun, R. (2017). Identification of structural vector autoregressions by stochastic volatility, Discussion Paper, University of Konstanz.

Caldara, D. and Herbst, E. (2019). Monetary policy, real activity, and credit spreads: Evidence from bayesian proxy svars, *American Economic Journal: Macroeconomics* **11**(1): 157–92.

Carriero, A., Clark, T. E. and Marcellino, M. (2016). Common drifting volatility in large bayesian VARs, *Journal of Business & Economic Statistics* **34**(3): 375–390.

Carriero, A., Mumtaz, H., Theodoridis, K. and Theophilopoulou, A. (2015). The impact of uncertainty shocks under measurement error: A proxy svar approach, *Journal of Money, Credit and Banking* **47**(6): 1223–1238.

Cesa-Bianchi, A., Thwaites, G. and Vicondoa, A. (2016). Monetary policy transmission in an open economy: new data and evidence from the United Kingdom, Mimemo, London School of Economics and Political Science, LSE Library.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end?, *Handbook of macroeconomics* **1**: 65–148.

Coibion, O. (2012). Are the effects of monetary policy shocks big or small?, *American Economic Journal: Macroeconomics* **4**(2): 1–32.

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

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.

Hachula, M., Piffer, M. and Rieth, M. (forthcoming). Unconventional monetary policy, fiscal side effects and euro area (im)balances, *Journal of the European Economic Association* .

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP, *Journal of applied Econometrics* **7**(S1).

Hanson, M. S. (2006). Varying monetary policy regimes: A vector autoregressive investigation, *Journal of Economics and Business* **58**(5-6): 407–427.

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

Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604–41.

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

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**(6): 1131–1149.

Leeper, E. M. and Zha, T. (2003). Modest policy interventions, *Journal of Monetary Economics* **50**(8): 1673–1700.

Lunsford, K. (2015). Identifying structural vars with a proxy variable and a test for a weak proxy, *Technical report*, Federal Reserve Bank of Cleveland.

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

Lütkepohl, H. and Schlaak, T. (forthcoming). Bootstrapping impulse responses of structural vector autoregressive models identified through GARCH, *Journal of Economic Dynamics and Control* .

Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states, *American Economic Review* **103**(4): 1212–1247.

Miranda-Agrippino, S. and Ricco, G. (2017). The transmission of monetary policy shocks, Mimemo, London School of Economics and Political Science, LSE Library.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: The information effect, *The Quarterly Journal of Economics* **133**(3): 1283–1330.

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

Olea, M., Stock, J. and Watson, M. (2018). Inference in structural vector autoregressions identified with an external instrument, *Technical report*, mimemo, Columbia University.

Owyang, M. T. and Ramey, G. (2004). Regime switching and monetary policy measurement, *Journal of Monetary Economics* **51**(8): 1577–1597.

Piffer, M. and Podstawski, M. (2018). Identifying uncertainty shocks using the price of gold, *The Economic Journal* .

Podstawski, M. and Velinov, A. (2018). The state dependent impact of bank exposure on sovereign risk, *Journal of Banking & Finance* **88**: 63–75.

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

Psaradakis, Z. and Spagnolo, N. (2006). Joint determination of the state dimension and autoregressive order for models with markov regime switching, *Journal of Time Series Analysis* **27**(5): 753–766.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation, *Handbook of Macroeconomics*, Vol. 2, Elsevier, pp. 71–162.

Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.

Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market, *The Quarterly Journal of Economics* **118**(2): 639–669.

Rigobon, R. and Sack, B. (2004). The impact of monetary policy on asset prices, *Journal of Monetary Economics* **51**(8): 1553 – 1575.

Rogers, J. H., Scotti, C. and Wright, J. H. (forthcoming). Unconventional monetary policy and international risk premia, *Journal of Money, Credit and Banking* .

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

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica: Journal of the Econometric Society* **48**(1): 1–48.

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

Stock, J. H. and Watson, M. W. (2002). Has the business cycle changed and why?, *NBER macroeconomics annual* **17**: 159–218.

Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 120–157.

Stock, J. H. and Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments, *The Economic Journal* **128**(610): 917–948.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments, *Journal of Business & Economic Statistics* **20**(4): 518–529.

Wieland, J. F. and Yang, M.-J. (2016). Financial Dampening, *NBER Working Papers 22141*, National Bureau of Economic Research, Inc.

Wright, J. H. (2012). What does monetary policy do to long-term interest rates at the zero lower bound?, *The Economic Journal* **122**(564): 447–466.