

Same, but different? Testing monetary policy shock measures*

Stephanie Ettmeier[†] Alexander Kriwoluzky[‡]

20th August 2018

Abstract

In this study, we determine the reliability of three popular monetary policy measures introduced by Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015) respectively, and estimate the causal effects of monetary policy. To this end, we employ the Proxy-SVAR model proposed by Stock and Watson (2012) and Mertens and Ravn (2013). We carry out impulse response analysis employing different test statistics to determine the shock measures' information content. We find that Gertler and Karadi's (2015) series qualifies as the most relevant monetary policy shock measure.

JEL classification: C12, C32, E32, E52

Keywords: Monetary policy shock measures, Proxy-SVAR, Weak proxies, F-test

*We would like to thank Martin Altemeyer-Bartscher, Georgios Georgiadis, Oliver Holtemöller, Kurt Lunsford, Chi Hyun Kim, Martin Kliem, Gernot Müller, Moritz Schularick, and the Nordic Econometric Meeting 2017 and annual International Association for Applied Econometrics 2017 conference participants.

[†]DIW Berlin, Germany

[‡]Corresponding author: DIW Berlin, Germany, akriwoluzky@diw.de, +49 30 89789-430

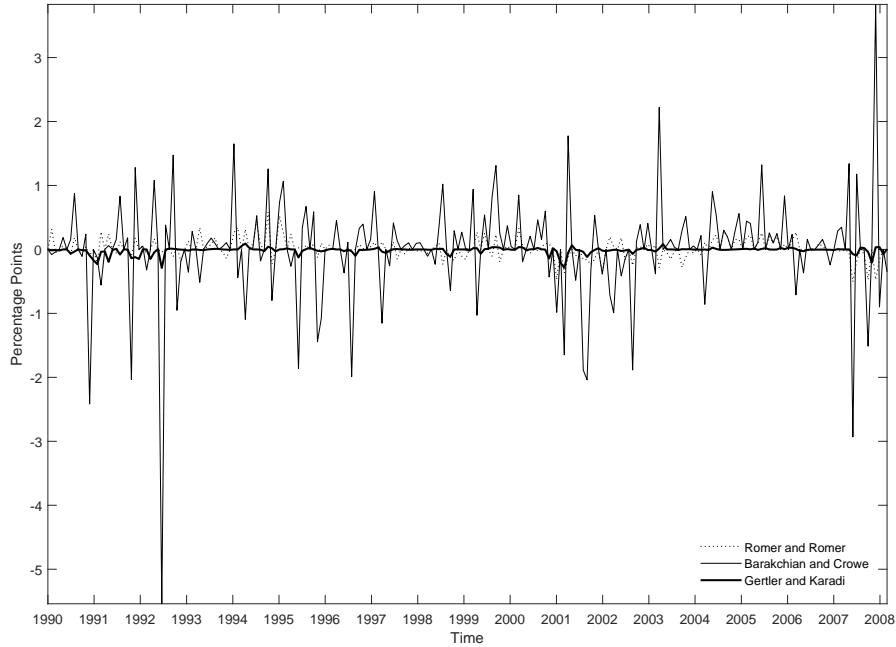
1 Introduction

What effect does a monetary policy shock have on the economy? This question has been addressed by Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015) in recent studies. The common underlying feature of these three studies is that they directly measure monetary policy shocks. All these measures have become popular, and many recent studies have applied them synonymously.¹ Interestingly, as Figure 1 shows, the three studies aim to measure the same structural shock, but the measures of these studies differ greatly from one another. One potential reason for this discrepancy brought forward in the literature is that a directly measured shock series is prone to measurement error and consequently does not represent the structural shock of interest in its entirety. Stock and Watson (2012) and Mertens and Ravn (2013) proposed with the Proxy-SVAR model an empirical framework for causal inference that accounts for this limitation by applying the shock measures as proxies for the structural shock.²

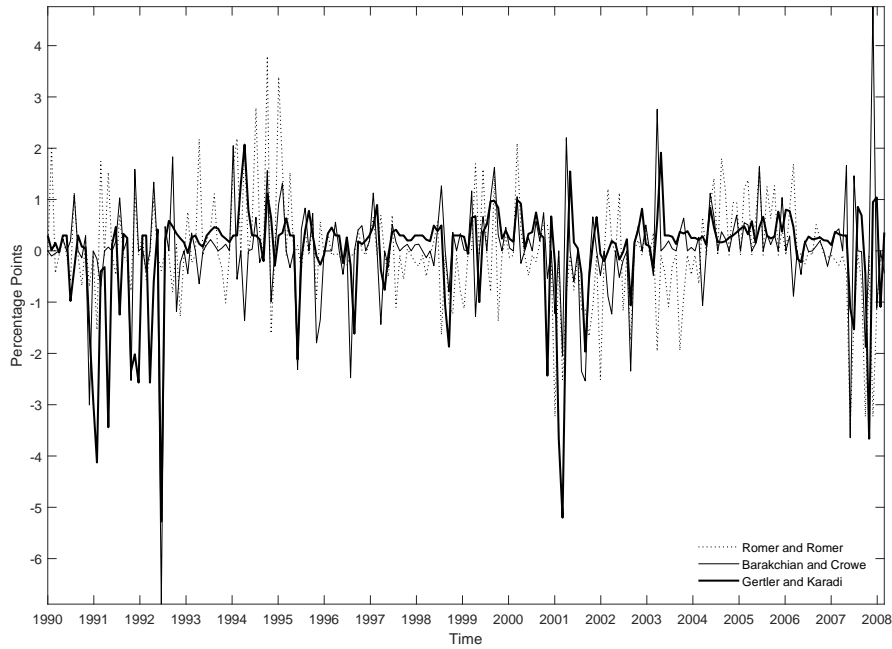
For capital- and labor-income tax shocks, Mertens and Ravn (2018) demonstrated that the degree of correlation between the shock measure and the structural shock is the key to reliable inference. In this study, we examine this aspect for the monetary policy shock measures of Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015). More precisely, we determine how reliable the shock measures estimate the causal effects of monetary policy on the economy. We first provide evidence that the choice of monetary policy shock measure does indeed influence the estimation results. To this end, we estimate a Proxy-SVAR model using the three different monetary policy shock measures as proxy variables and compute the corresponding impulse responses. Our model includes a variable for interest rates, output, prices, and credit costs. We derive the confidence bands

¹Examples include Midrigan (2011), Stock and Watson (2012), Georgiadis (2017), Miranda-Agrippino and Rey (2015), Passari and Rey (2015), Zeev et al. (2015), Balakrishnan et al. (2016), Piffer (2016), Sinclair et al. (2016), and Tenreyro and Thwaites (2016).

²The Proxy-SVAR approach enjoys great popularity in the empirical macroeconomic literature. Recent studies include Caldara and Herbst (2018), Mertens and Montiel Olea (2018), Braun and Brüggemann (2017), Caldara and Kamps (2017), Mian et al. (2017), Cesa-Bianchi et al. (2016), Zanetti and Li (2016), and Carriero et al. (2015).



(a) Original shock measures



(b) Standardized shock measures

Figure 1. Comparison of monetary policy shock measures for their overlapping period 1990:M1 to 2008:M6.

Note: The dotted line depicts Romer and Romer (2004) updated by Barakchian and Crowe (2013), the solid line depicts Barakchian and Crowe (2013), and the bold solid line depicts Gertler and Karadi (2015). To compute the standardized series in panel (b), we demeaned the proxies and normalized their variance to 1.

of the impulse response functions (IRFs) using the method of Montiel Olea et al. (2016). We choose this particular method for inference because it allows for computing asymptotically valid probability bands even when the proxy variable's information content is low. As the width of the confidence bands nevertheless depends on the proxy variable, this method allows for a first comparison of the three monetary policy shock measures. We find that the responses in our Proxy-SVAR model to the three proxy variables are comparable in sign and magnitude, but they differ in statistical significance. Only Gertler and Karadi (2015)'s measure leads to statistically significant estimates for all the IRFs.

To gain further insights, we next test the proxy variables' information content, that is their relevance, to the underlying structural shock. For this, we employ two different test statistics used in the literature to determine the relevance of proxy variables in a VAR model (Gertler and Karadi, 2015; Lunsford, 2015). Both test statistics are based on the F-statistic that determines whether the correlation between a proxy variable and the structural shock of interest is close to zero, but they start from different setups. On the one hand, Gertler and Karadi (2015) follow the microeconomic instrumental variable approach and employ the proxy as an independent variable. In particular, they regress the reduced-form error related to the monetary policy variable on the proxy to derive the F-statistic. On the other hand, Lunsford (2015), following Stock and Watson (2012), applies the proxy as a dependent variable. He calculates the F-statistic by regressing the proxy variable on a constant and the entire vector of reduced-form errors. We find that the measure of Gertler and Karadi (2015) is the only shock series allowing for rejection of the null hypothesis of weak relevance for both test statistics. From these test results and the preceding impulse response analysis, we conclude that the shock series of Gertler and Karadi (2015) qualifies as the most relevant monetary policy shock measure.

Our choice of the three monetary policy shock measures, is driven by the consideration to cover a representative cross section of the most recent and openly accessible series. In general, monetary policy shock measures can be broadly split into narrative shock and high-

frequency shock measures. As representative of narrative shock measures, we consider the seminal work of Romer and Romer (2004), who study the archives of the Federal Reserve System. The high-frequency approach, on the other hand, measures the changes in interest-rate futures on days when the Federal Open Market Committee (FOMC) meets. Here, we follow two approaches in the literature to construct high-frequency series. On the one hand, Gürkaynak et al. (2005), Campbell et al. (2012), Barakchian and Crowe (2013), and Nakamura and Steinsson (2017) employ a factor model to condense the shock measure from different futures contracts. From this strand, we choose the shock measure of Barakchian and Crowe (2013), since it qualifies as one of the most recent and openly accessible high-frequency series. On the other hand, Kuttner (2001), Cochrane and Piazzesi (2002), Faust et al. (2004), Gertler and Karadi (2015), and Hanson and Stein (2015) calculate the first differences of a single contract. From this group, we choose the series of Gertler and Karadi (2015) because it constitutes the most recent and openly accessible high-frequency monetary policy shock measure.

The remainder of this paper is structured as follows. Section 2 describes the three monetary policy shock measures and shows how different they are from one another. Section 3 sets up our empirical model and applies several econometric tools to compare the monetary policy shock measures. To draw conclusions for causal inference, we compute IRFs and formally test the proxy variables' strength. The final section concludes the study.

2 Three different monetary policy shock measures

In this section, we briefly describe the three monetary policy shock measures considered in this study. We first illustrate the narrative approach employed by Romer and Romer (2004) and then outline the approaches used by Barakchian and Crowe (2013) and Gertler and Karadi (2015).

Romer and Romer (2004) construct their monetary policy shock measure after carefully studying the Federal Reserve archives. First, the authors use the *Record of Policy Actions of the FOMC*, the *Minutes of the FOMC*, and the *Monetary Policy Alternatives (Bluebook)* to determine the intended federal funds rate (\tilde{i}_t). Romer and Romer (2004) then specify the information set of the FOMC using the *Greenbook*, that is, a report on the *Current Economic and Financial Conditions*. From this source, they derive forecasts of real output growth (Δx_t^f), inflation (p_t^f), and the unemployment rate (u_t^f). Finally, the authors determine the monetary policy shock measure (s_t^{RR}) as the residual in the following regression:

$$\begin{aligned} \Delta \tilde{i}_t = & \beta_0 + \beta_1 \tilde{i}_t + \sum_{h=-1}^2 \beta_{2,h} \Delta x_{t,h}^f + \sum_{h=-1}^2 \beta_{3,h} (\Delta x_{t,h}^f - \Delta x_{t-1,h}^f) \\ & + \sum_{h=-1}^2 \beta_{4,h} p_{t,h}^f + \sum_{h=-1}^2 \beta_{5,h} (p_{t,h}^f - p_{t-1,h}^f) + \beta_6 u_t^f + s_t^{RR}. \end{aligned} \quad (1)$$

Barakchian and Crowe (2013) and Gertler and Karadi (2015) follow the approach of Söderström (2001), Kuttner (2001), Cochrane and Piazzesi (2002), Faust et al. (2004), and Gürkaynak et al. (2005). They employ high-frequency data on interest-rate futures to identify a monetary policy shock. We denote the interest-rate futures traded at date d in month m for the federal funds rate in month $m+h$ as f_d^h . Assuming that on average the FOMC succeeds in keeping the federal funds rate close to the target rate, and that the risk premium stays constant on the day of policy announcement, the change in the expected federal funds target rate ($E_d \bar{i}_{m+h}$) during the subsequent calendar months ($h \geq 1$) following a policy announcement on day d of month m is given by³

$$\Delta E_d \bar{i}_{m+h} = f_d^h - f_{d-1}^h \quad (2)$$

³The change following a policy announcement in the expected federal funds target rate ($E_d \bar{i}_{m+h}$) during the remainder of the current calendar month ($h = 0$) is given by $\Delta E_d \bar{i}_m = \frac{M}{M-d} (f_d^0 - f_{d-1}^0)$ respectively, where M denotes the length of the current month in days.

Barakchian and Crowe (2013) calculate the change in the expected federal funds target rate in the current month and up to five months ahead. The authors then follow Gürkaynak et al. (2005) and employ a factor model to construct the monetary policy shock measure. Their factor model is given by

$$\omega_t = \phi_t \Lambda + e_t \quad (3)$$

where the vector ω_t collects the changes in the expected federal funds target rate for different maturities, e_t comprises white noise disturbances for each entry in ω_t , $\phi_t = \begin{bmatrix} \phi_{1,t} & \phi_{2,t} \end{bmatrix}'$ includes two factors, and the vector Λ includes the corresponding loadings. Barakchian and Crowe (2013) argue that the first factor $\phi_{1,t}$ measures the monetary policy shock because it captures the portion of information revealed by the policy announcement shifting the yield curve of interest rates vertically at all horizons. It thus coincides with a transmission of monetary policy, where the changes in short-term interest rates influence the rates at longer horizons. Hence, they set $s_t^{BC} = \phi_{1,t}$.

In contrast to Barakchian and Crowe (2013), Gertler and Karadi (2015) do not estimate a factor model, but instead construct the measure directly from the difference between the expected federal funds target rates. Furthermore, they do not consider a 24-hour time window around the FOMC meetings, but rather measure the monetary policy shocks in a 30-minute window around the policy announcement. As their baseline monetary policy shock measure, Gertler and Karadi (2015) consider the change in three-month ahead futures contracts; that is, $s_t^{GK} = \Delta E_{30min} \bar{i}_{m+3}$. Here, ΔE_{30min} indicates the 30-minute window around the policy announcement.

Figure 1 plots the different monetary policy shock measures in their original and standardized forms.⁴ Although the three measures are supposed to measure the same structural shock, they appear strikingly different. Table 1 underlines the differences between them. The measures derived using the high-frequency data approach are weakly correlated with

⁴All measures were obtained from supplementary journal archives. Appendix A provides further information on the dataset. Note that we use the shock series of Romer and Romer (2004), specifically the updated version in Barakchian and Crowe (2013), throughout this study to extend the sample period to 2008:M6.

the narrative time series. In fact, even the correlation between the two measures constructed from the interest-rate futures data is only around 0.4.

Although these shock measures are different, all of them are subject to criticism. For example, Leeper (1997) and Coibion (2012) suggest that a narrative time series could still contain endogenous components, and Ramey (2016) points out that the measure constructed by Gertler and Karadi (2015) can be predicted through Greenbook forecasts. Despite these criticisms, we continue our analysis with the original monetary policy shock measures and determine their effect on the inference of monetary policy effects for two reasons. First, as the extensive list of citations in the introduction shows, these monetary policy shock measures are nevertheless employed in applied research. Second, we need to know which measure can provide valid inference on dynamic causal effects and should be further developed and improved.⁵

TABLE 1
Correlation between monetary policy shock measures

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| RR | 1.000 | | |
| BC | 0.193 | 1.000 | |
| GK | 0.190 | 0.405 | 1.000 |

3 Testing monetary policy shock measures

In this section, we first describe the data. We then set up the empirical model, compute IRFs, and statistically test the shock measures' relevance.

The empirical model

For our analysis, we employ the Proxy-SVAR model developed by Stock and Watson (2012) and Mertens and Ravn (2013). The VAR model contains four endogenous variables, one for

⁵For initial improvement approaches, see, for instance, Miranda-Agrippino (2016) and Paul (2017).

monetary policy, one for output, one for price level, and, following the suggestion of Caldara and Herbst (2018), one for credit costs. The variables are included in the vector y_t as follows:

$$y_t = \begin{bmatrix} i_t & x_t & p_t & ebp_t \end{bmatrix}' \quad (4)$$

where i_t denotes the federal funds rate, x_t is the logarithm of industrial production, p_t is the logarithm of consumer price index, and ebp_t is the excess bond premium of Gilchrist and Zakrajšek (2012). Throughout this study, we estimate the model over the sample time span from 1979:M7 to 2012:M6, but for identification, we consider the time span for which the three monetary policy shock series are jointly available; that is, from 1990:M1 to 2008:M6. Stock and Watson (2018) argue that estimating the VAR model over a longer time span improves efficiency.

In general, the VAR model with n endogenous variables is given by

$$y_t = B_0 + B_1 t + B(L)y_{t-1} + u_t, \quad u_t \sim \text{i.i.d.} (0, \Sigma_u) \quad (5)$$

where $B(L)$ denotes the reduced-form VAR model coefficients, B_1 is a linear time trend, and B_0 is the intercept term. We employ a lag order of 12 because all our data have monthly frequency.⁶ u_t denotes the $n \times 1$ vector of reduced-form errors with the corresponding variance-covariance matrix Σ_u . The reduced-form errors u_t are related to the structural errors ϵ_t as follows:

$$u_t = A\epsilon_t, \quad \epsilon_t \sim \text{i.i.d.} (0, I) \quad (6)$$

The identification issue in VAR models arises because we cannot determine A uniquely from $\Sigma_u = AA'$. The Proxy-SVAR model estimates the effects of a structural shock by using an additional measure for monetary policy as a proxy for the underlying structural shock. We denote the column in A related to the monetary policy shock as a_m . Furthermore, we

⁶In Appendix C, we provide various robustness checks. We show that our results are robust to the order of lags and do not hinge on the inclusion of a linear time trend.

partition the structural shocks and reduced-form errors as follows:

$$\epsilon_t = \begin{bmatrix} \epsilon_{m,t} & \epsilon_{2,t} \\ 1 \times 1 & n-1 \times 1 \end{bmatrix}' \quad (7)$$

$$u_t = \begin{bmatrix} u_{m,t} & u_{2,t} \\ 1 \times 1 & n-1 \times 1 \end{bmatrix}' \quad (8)$$

The first block ($\epsilon_{m,t}$ and $u_{m,t}$) is associated with the monetary policy shock, and the second block comprises the additional shocks. The three different monetary policy shock measures (s_t^{RR} , s_t^{BC} , or s_t^{GK} are represented by s_t^*) are subsequently assumed to be proxy variables correlated with the monetary policy shock $\epsilon_{m,t}$:

$$E[s_t^* \epsilon_{m,t}] = \Phi, \quad \Phi \neq 0 \quad (9)$$

and uncorrelated with the remaining structural shocks:⁷

$$E[s_t^* \epsilon_{2,t}] = 0 \quad (10)$$

Condition 10 postulates the exogeneity of the proxy variable, while Condition 9 states its relevance. In general, the more information of the structural shock a proxy variable contains, the more relevant it qualifies and the more reliable will be inference. In contrast, proxies having close to zero percent correlation with the structural shock (weak proxies) will lead to inconsistent estimates and thus complicate statements about causal relations (Stock and Watson, 2018, 2016; Lunsford, 2015). For our study, we exploit this relevance condition to draw conclusions on the monetary policy shock measures' impact on inference.⁸

⁷Conditions 9 and 10 mirror the relevance and exogeneity conditions from the microeconomic instrumental variable literature respectively. However, the proxy variable serves a different purpose in the framework of Stock and Watson (2012) and Mertens and Ravn (2013). Here, the problem is not that a_m is only inconsistently estimated because $\epsilon_{m,t}$ and $\epsilon_{2,t}$ are correlated, but rather that a_m is unidentifiable because $\epsilon_{m,t}$ is not observable. Because of this difference, statistical techniques robust to weak identification, as developed by, for example, Kleibergen (2005), Kleibergen and Mavroeidis (2009), or Montiel Olea and Pflueger (2013), cannot be applied in the Proxy-SVAR framework.

⁸In line with the literature (for example, Tenreiro and Thwaites (2016) or Stock and Watson (2018)), we take the exogeneity of the three monetary policy proxy variables as given.

Dynamic effects of monetary policy

Figure 2 compares the dynamic effects of a contractionary one-unit monetary policy shock as identified by the three shock measures.⁹

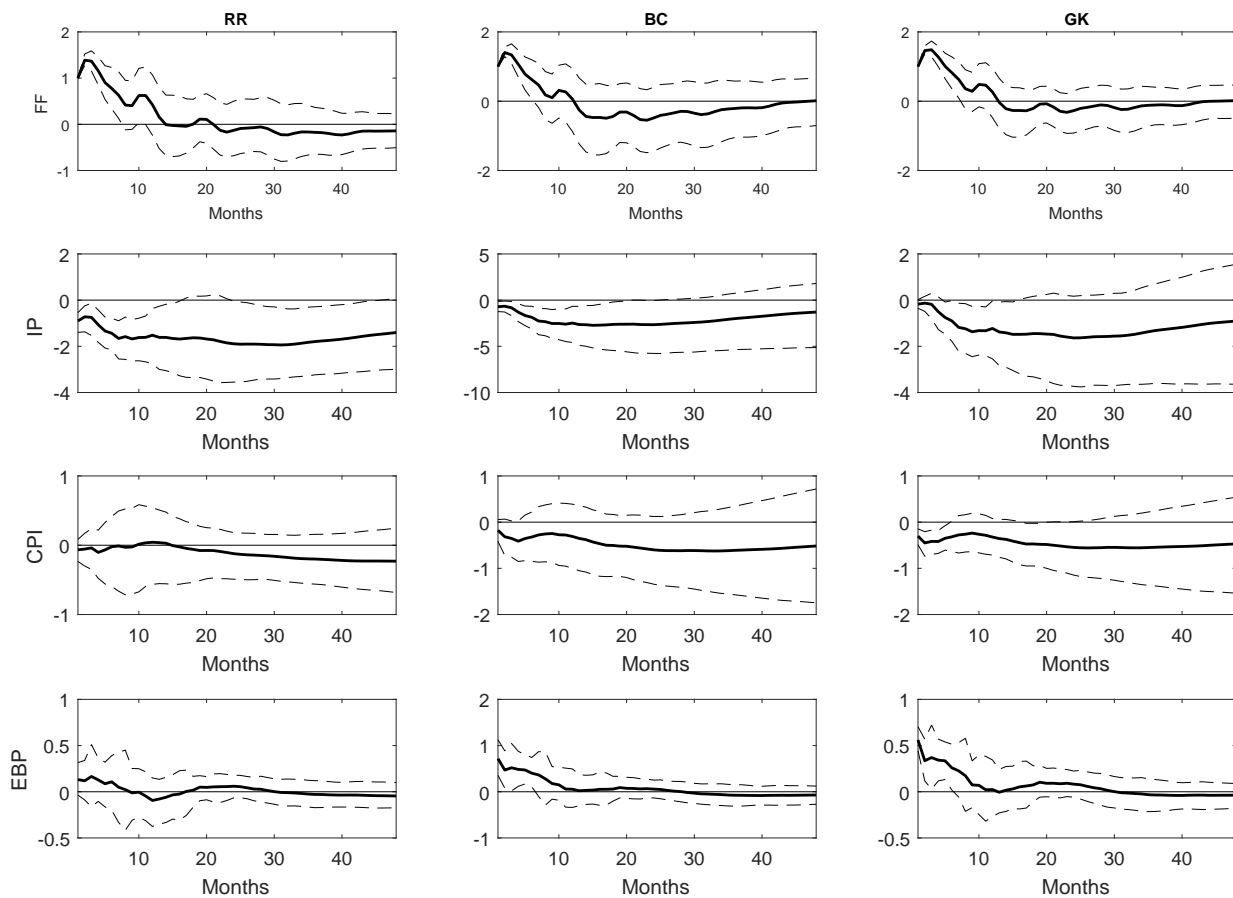


Figure 2. Dynamic effects of a monetary policy shock identified by the shock measures of Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015).

Note: FF stands for federal funds rate, IP for industrial production, CPI for consumer price index, and EBP for excess bond premium. Inference was made using the method of Montiel Olea et al. (2016). The bold line depicts the mean estimate, and the dashed lines show 68% confidence bands.

As stressed by Montiel Olea et al. (2016) and empirically demonstrated by Mertens and Ravn (2018), proxy strength is an important concern for inference in Proxy-SVAR models. In our application, we use the method of Montiel Olea et al. (2016) to compute confidence bands. The authors suggest confidence sets for IRFs that are asymptotically valid even in

⁹Appendix B provides an overview of the identified structural monetary policy shocks.

case of weak proxy variables. An appealing feature of this method is that the width of the confidence bands depends on the proxy. Applying the weak-proxy robust inference will allow for direct comparison of the performance of shock measures because the only factor we vary are the proxies themselves.¹⁰

Consistent with conventional theory and similarly for the three shock series, an increase in interest rates induces on impact a decline in industrial production and consumer prices, and a rise in credit costs. In general, the three shock measures produce very similar results in terms of size and magnitude. The impact effects differ only slightly in magnitude, and the IRFs exhibit comparable dynamics for longer horizons. The shock measure of Romer and Romer (2004) significantly influences the federal funds rate and industrial production on impact. However, it contains insufficient information to infer the monetary policy effects on the price level and credit costs. Except for consumer prices, the series of Barakchian and Crowe (2013) produces significant effects in the short and medium term. The measure of Gertler and Karadi (2015) is the only series enabling estimation of statistically significant effects on all four variables.

Two tests for weak proxies

Condition 9 of the Proxy-SVAR approach provides a hypothesis that allows for directly testing the shock measures' relevance. To carry out the test, we employ two different statistical tests that are usually applied in the Proxy-SVAR literature. Our first test is the F-test used by Gertler and Karadi (2015). Following the microeconomic instrumental variable approach, Gertler and Karadi (2015) use the proxy as an independent variable to derive the test statistic from the linear projection of the reduced-form error associated with monetary

¹⁰Alternative methods for inference in Proxy-SVAR models exist; for instance, the Delta method employed by Mertens and Montiel Olea (2018) or the moving block bootstrap method employed by Lunsford and Jentsch (2016). However, only the method of Montiel Olea et al. (2016) allows for asymptotic valid inference in case of weak proxy variables. Since we consider the three shock measures a priori equal, we choose the inference method robust to proxy strength.

policy on s_t^* :

$$u_{m,t} = \gamma (s_t^* - \bar{s}^*) + \eta_t \quad (11)$$

Here, \bar{s}^* denotes the mean of the proxy variable, and η_t is an *i.i.d.* error term with variance σ_η^2 ; it is assumed to be independent of ϵ_t and lags of y_t . To determine whether s_t^* is sufficiently correlated with $\epsilon_{m,t}$, Gertler and Karadi (2015) calculated the F-statistic of Equation (11) for the null hypothesis that $\gamma = 0$:

$$F_{IV} = (T - 1) \frac{\sum_{t=1}^T [\hat{u}_{m,t}^2 - [\hat{u}_{m,t} - \hat{\gamma}(s_t^* - \bar{s}^*)]^2]}{\sum_{t=1}^T [\hat{u}_{m,t} - \hat{\gamma}(s_t^* - \bar{s}^*)]^2} \quad (12)$$

Here, $\hat{\gamma}$ denotes the ordinary least squares (OLS) estimate of γ in Equation 11. For the test decision, they rely on the commonly accepted rule of thumb, which states that an F-statistic greater than 10 signals proxy strength (Stock and Yogo, 2005).

Table 2 presents the results for F_{IV} . According to this test, the measures of Romer and Romer (2004) and Gertler and Karadi (2015) constitute relevant proxy variables.

TABLE 2
Results F_{IV}

| | RR | BC | GK |
|----------|-----------|-----------|-----------|
| F_{IV} | 17.961* | 9.39 | 28.87* |

Note: * indicates significance according to the rule of thumb $F_{IV} > 10$.

Our second test is the F-test suggested by Lunsford (2015). In contrast to Gertler and Karadi (2015), Lunsford (2015) follows Stock and Watson (2012) and employs the proxy as a dependent variable. To compute the test statistic, first, s_t^* has to be regressed on the right-hand side variables of the VAR model, and then the fitted value of this regression, \hat{s}_t^* , has to be projected onto the entire vector of the reduced-form errors and a constant as follows:

$$\hat{s}_t^* = c_0 + u_t' \mu + \nu_t \quad (13)$$

where μ is a vector of dimension $n \times 1$, and ν_t is an *i.i.d.* error term with variance σ_ν^2 . Furthermore, ν_t is assumed to be independent of ϵ_t and lags of y_t . To test whether s_t^* constitutes a weak proxy variable for $\epsilon_{m,t}$, Lunsford (2015) suggested calculating the F-statistic of Equation (13) for the null hypothesis that $\mu = 0$:

$$F_{WP} = \left(\frac{T-n}{n} \right) \frac{\sum_{t=1}^T [(\hat{s}_t^* - \bar{s}^*)^2 - [(\hat{s}_t^* - \bar{s}^*) - \hat{u}_t' \hat{\mu}]^2]}{\sum_{t=1}^T [(\hat{s}_t^* - \bar{s}^*) - \hat{u}_t' \hat{\mu}]^2} \quad (14)$$

Here, $\hat{\mu}$ denotes the OLS estimate of μ in Equation 13.¹¹

Table 3 displays the results for F_{WP} . For the test decision, we use the critical values provided by Lunsford (2015).¹² According to this test, only the measure of Gertler and Karadi (2015) is a relevant proxy variable.

TABLE 3
Results F_{WP}

| | RR | BC | GK |
|----------|-----------|-----------|-----------|
| F_{WP} | 3.503 | 2.165 | 10.275* |

Note: * indicates significance at the 5% level according to the critical value of Lunsford (2015) $F_{WP} > 8.22$

In summary, only the proxy variable of Gertler and Karadi (2015) allows for rejection of the null hypothesis of weak relevance for the two test statistics. This finding supports the impulse response analysis where only Gertler and Karadi's measure had significant monetary policy effects on all variables. From our two exercises, we can conclude that the monetary policy shock measure of Gertler and Karadi (2015) allows for estimating the dynamic causal effects of monetary policy most reliably, because it contains the most information on the structural monetary policy shock.

¹¹We simplify the notation and denote the mean of \hat{s}_t^* also with \bar{s}^* .

¹²Consistent with the relevant studies in the literature, that is, Lunsford (2015), Stock and Yogo (2005), and Stock et al. (2002), the critical values we apply correspond to a tolerated asymptotic bias level of 10%. So far, Lunsford (2015) has not formally justified his test for VAR models having a deterministic trend. We present test results for the VAR model with no linear time trend and stationary variables in Appendix C as a robustness check.

4 Conclusion

In this study, we identify the monetary policy shock measure most relevant to study the effects of monetary policy. So far, the measures of Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015) have been used synonymously in the related literature. To this end, we employ the Proxy-SVAR model of Stock and Watson (2012) and Mertens and Ravn (2013). We carry out impulse response analysis and use two different test statistics to determine the shock measures' information content. We find that the series of Gertler and Karadi (2015) qualifies as the most relevant shock measure for two reasons. First, of the three series, only Gertler and Karadi's measure produced significant monetary policy effects on all variables. Second, it is the only shock measure that allows for rejecting the null hypothesis of weak relevance for the two test statistics.

References

- Balakrishnan, R., S. Laseen, and A. Pescatori (2016). ‘U.S. Dollar Dynamics; How Important Are Policy Divergence and FX Risk Premiums?’, IMF Working Papers 16/125, International Monetary Fund.
- Barakchian, S. M. and C. Crowe (2013). ‘Monetary policy matters: Evidence from new shocks data’, *Journal of Monetary Economics*, Vol. 60, pp. 950–966.
- Braun, R. and R. Brüggemann (2017): ‘Identification of SVAR Models by Combining Sign Restrictions With External Instruments,’ Working Paper Series of the Department of Economics, University of Konstanz 2017-07, Department of Economics, University of Konstanz.
- Caldara, D. and E. Herbst (2018). ‘Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs,’ *American Economic Journal: Macroeconomics*, forthcoming.
- Caldara, D. and C. Kamps (2017). ‘The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers,’ *The Review of Economic Studies*, Vol. 84, pp. 1015–1040.
- Campbell, J. R., C. L. Evans, J. D. Fisher, and A. Justiniano (2012). ‘Macroeconomic Effects of Federal Reserve Forward Guidance’, *Brookings Papers on Economic Activity*, Vol. 43, pp. 1–80.
- Carriero, A., H. Mumtaz, K. Theodoridis, and A. Theophilopoulou (2015). ‘The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach,’ *Journal of Money, Credit and Banking*, Vol. 47, 1223–1238.
- Cesa-Bianchi, A., G. Thwaites, and A. Vicendoa (2016): ‘Monetary policy transmission in an open economy: new data and evidence from the United Kingdom,’ Staff Working Paper 615, Bank of England.

- Cochrane, J. H. and M. Piazzesi (2002). ‘The Fed and Interest Rates - A High-Frequency Identification’, *American Economic Review*, Vol. 92, pp. 90–95.
- Coibion, O. (2012). ‘Are the effects of monetary policy shocks big or small?’, *American Economic Journal: Macroeconomics*, Vol. 4, pp. 1–32.
- Faust, J., E. T. Swanson, and J. H. Wright (2004). ‘Identifying VARs based on high frequency futures data’, *Journal of Monetary Economics*, Vol. 51, pp. 1107–1131.
- Georgiadis, G. (2017). ‘To bi, or not to bi? Differences between spillover estimates from bilateral and multilateral multi-country models’, *Journal of International Economics*, Vol. 107, pp. 1–18.
- Gertler, M. and P. Karadi (2015). ‘Monetary policy surprises, credit costs, and economic activity’, *American Economic Journal: Macroeconomics*, Vol. 7, pp. 44–76.
- Gilchrist, S. and E. Zakrajšek (2012). ‘Credit Spreads and Business Cycle Fluctuations’, *American Economic Review*, Vol. 102, pp. 1692–1720.
- Gürkaynak, R. S., B. P. Sack, and E. T. Swanson (2005). ‘Do actions speak louder than words? The response of asset prices to monetary policy actions and statements’, *International Journal of Central Banking*, Vol. 1, pp. 55–93.
- Hanson, S. G. and J. Stein (2015). ‘Monetary policy and long-term real rates’, *Journal of Financial Economics*, Vol. 115, pp. 429–448.
- Kleibergen, F. (2005). ‘Testing Parameters in GMM Without Assuming that They Are Identified’, *Econometrica*, Vol. 73, pp. 1103–1123.
- Kleibergen, F. and S. Mavroeidis (2009). ‘Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve’, *Journal of Business & Economic Statistics*, Vol. 27, pp. 293–311.

- Kuttner, K. N. (2001). ‘Monetary policy surprises and interest rates: Evidence from the Fed funds futures market’, *Journal of Monetary Economics*, Vol. 47, pp. 523–544.
- Leeper, E. M. (1997). ‘Narrative and VAR approaches to monetary policy: Common identification problems’, *Journal of Monetary Economics*, Vol. 40, pp. 641–657.
- Lunsford, K. G. (2016). ‘Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy’, Manuscript.
- Lunsford, K. G. and C. Jentsch (2016). ‘Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States,’ Working Paper 1619, Federal Reserve Bank of Cleveland.
- Mertens, K. and J. L. Montiel Olea (2018). ‘Marginal Tax Rates and Income: New Time Series Evidence,’ *The Quarterly Journal of Economics*, forthcoming.
- Mertens, K. and M. Ravn (2018). ‘The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply to Jentsch and Lunsford,’ Working Papers 1805, Federal Reserve Bank of Dallas.
- Mertens, K. and M. O. Ravn (2013). ‘The dynamic effects of personal and corporate income tax changes in the United States’, *American Economic Review*, Vol. 103, pp. 1212–1247.
- Mian, A., A. Sufi, and E. Verner (2017). ‘Household Debt and Business Cycles Worldwide*,’ *The Quarterly Journal of Economics*, Vol. 132, pp. 1755–1817.
- Midrigan, V. (2011). ‘Menu Costs, Multiproduct Firms, and Aggregate Fluctuations’, *Econometrica*, Vol. 79, pp. 1139–1180.
- Miranda-Agrippino, S. (2016). ‘Unsurprising shocks: information, premia, and the monetary transmission’, Bank of England working papers 626, Bank of England.
- Miranda-Agrippino, S. and H. Rey (2015). ‘World Asset Markets and the Global Financial Cycle’, Working Paper 21722, National Bureau of Economic Research.

- Montiel Olea, J. L. and C. Pflueger (2013). ‘A Robust Test for Weak Instruments’, *Journal of Business & Economic Statistics*, Vol. 31, pp. 358–369.
- Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2016), ‘Inference in SVARs identified with External Instruments,’ Manuscript.
- Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2016). ‘Uniform Inference in SVARs with External Instruments’, Manuscript.
- Nakamura, E. and J. Steinsson (2017). ‘High Frequency Identification of Monetary Non-Neutrality: The Information Effect’, *Quarterly Journal of Economics*, Vol. 133, pp. 1283–1330.
- Passari, E. and H. Rey (2015). ‘Financial Flows and the International Monetary System’, *The Economic Journal*, Vol. 125, pp. 675–698.
- Paul, P. (2017). ‘The Time-Varying Effect of Monetary Policy on Asset Prices’, Working Paper Series 2017-9, Federal Reserve Bank of San Francisco.
- Piffer, M. (2016). ‘Monetary Policy and Defaults in the US’, Discussion Papers of DIW Berlin 1559, DIW Berlin, German Institute for Economic Research.
- Ramey, V. A. (2016). ‘Macroeconomic Shocks and Their Propagation’, in J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, pp. 71–162.
- Romer, C. D. and D. H. Romer (2004). ‘A New Measure of Monetary Shocks: Derivation and Implications’, *American Economic Review*, Vol. 94, pp. 1055–1084.
- Sinclair, T., P.-L. Tien, and E. Gamber (2016). ‘Do Fed Forecast Errors Matter?’, Working Papers 2016-14, The George Washington University, Institute for International Economic Policy.
- Söderström, U. (2001). ‘Predicting monetary policy with federal funds futures prices’, *Journal of Futures Markets*, Vol. 21, pp. 377–391.

- Stock, J. H. and M. W. Watson (2012). ‘Disentangling the Channels of the 2007-09 Recession’, *Brookings Papers on Economic Activity*, Vol. 44, pp. 81–156.
- Stock, J. H. and M. W. Watson (2016). ‘Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics’, in J. B. Taylor and H. Uhlig (eds), *Handbook of Macroeconomics*, Vol. 2, pp. 415–525.
- Stock, J. H. and M. W. Watson (2018). ‘Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments,’ Working Paper 24216, National Bureau of Economic Research.
- Stock, J. H., J. H. Wright, and M. Yogo (2002). ‘A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments’, *Journal of Business & Economic Statistics*, Vol. 20, pp. 518–529.
- Stock, J. H. and M. Yogo (2005). ‘Testing for Weak Instruments in Linear IV Regression’, in D. W. K. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, pp. 80–108.
- Tenreyro, S. and G. Thwaites (2016). ‘Pushing on a String: US Monetary Policy Is Less Powerful in Recessions’, *American Economic Journal: Macroeconomics*, Vol. 8, pp. 43–74.
- Zanetti, F. and W. Li (2016): ‘The Effect of Monetary Policy Shocks in the United Kingdom: an External Instruments Approach,’ Department of Economics Discussion Paper Series 812, University of Oxford, Department of Economics.
- Zeev, N. B., C. M. Gunn, and H. Khan (2015). ‘Monetary News Shocks’, Carleton Economic Papers 15-02, Carleton University, Department of Economics.

Appendix A Data description

The frequency of all data is monthly.

Barakchian and Crowe (2013) Shock Measure: The high-frequency monetary policy shock measure is downloadable as *shock*.

<http://dx.doi.org/10.1016/j.jmoneco.2013.09.006> (01/30/2017).

Consumer Price Index: The consumer price index for all urban consumers and all items is seasonally adjusted, chained (reference year 1982–1984), and log transformed. This series is downloadable from the Federal Reserve Bank of St. Louis as *CPIAUCSL*.

<https://fred.stlouisfed.org/series/CPIAUCSL> (01/30/2017).

Excess Bond Premium: The excess bond premium is originally provided by Gilchrist and Zakrajsek (2012). Monthly updates are made available by the Board of Governors of the Federal Reserve System. The series is downloadable as *ebp*.

<https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/updating-the-recession-risk-and-the-excess-bond-premium-20161006.html> (01/10/2017).

Federal Funds Rate: The effective federal funds rate is retrieved as *FEDFUNDS* from the Federal Reserve Bank of St. Louis. The series is measured in percentage and not seasonally adjusted.

<https://fred.stlouisfed.org/series/FEDFUNDS> (01/30/2017).

Gertler and Karadi (2015) Shock Measure: The high-frequency monetary policy shock measure is provided by the authors as *ff4*.

https://www.aeaweb.org/aej/mac/data/0701/2013-0329_data.zip (01/30/2017).

Industrial Production Index: The industrial production index is taken from the Federal Reserve Bank of St. Louis. The series *INDPRO* is seasonally adjusted and chained (reference year 2012). It is used in its log transformation.

<https://fred.stlouisfed.org/series/INDPRO> (01/30/2017).

Romer and Romer (2004) Shock Measure: The original narrative monetary policy shock measure from 1969 to 1996 is provided by Romer and Romer (2004). Monthly updates till June 2008 are made available by Barakchian and Crowe (2013) and can be retrieved as *resid08*.

<http://dx.doi.org/10.1016/j.jmoneco.2013.09.006> (01/30/2017).

Appendix B Structural monetary policy shocks

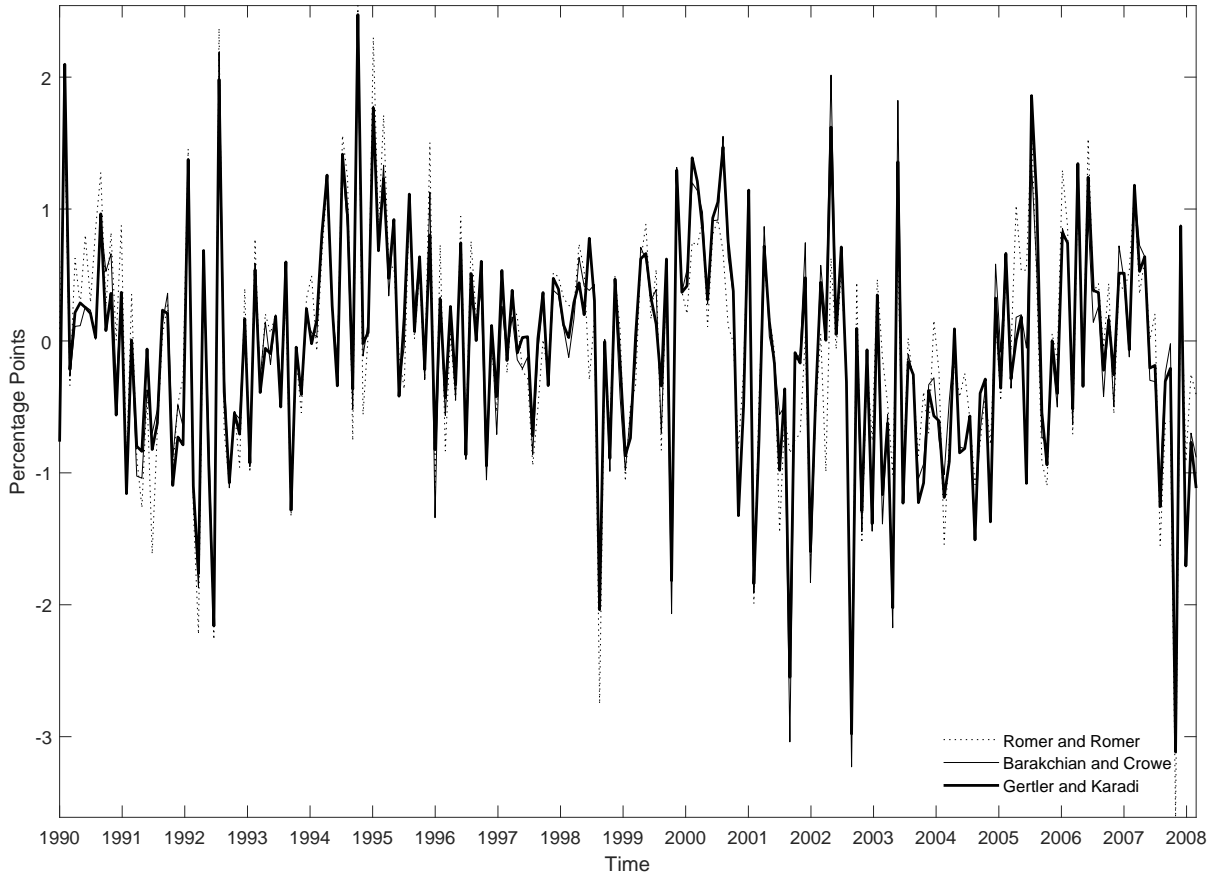


Figure 3. Comparison of structural monetary policy shocks for the period 1990:M1 to 2008:M6.

Note: Identification was achieved successively with one of the three monetary policy shock measures in a Proxy-SVAR model. The dotted line depicts the structural shock identified by the series of Romer and Romer (2004) updated by Barakchian and Crowe (2013), the solid line depicts that by Barakchian and Crowe (2013), and the bold solid line depicts that by Gertler and Karadi (2015).

The structural monetary policy shock $\epsilon_{m,t}$ can be recovered from the variance of the reduced-form residuals Σ_u and the covariance between the proxy and the reduced-form residuals $E[u_t s_t]$ respectively:¹³

$$\epsilon_{m,t} = u_t' \Sigma_u^{-1} E[u_t s_t] \{E[s_t u_t'] \Sigma_u^{-1} E[u_t s_t]\}^{-1/2}. \quad (15)$$

¹³As in Mertens and Ravn (2013) and Lunsford (2015), we normalize the variance of the structural shocks σ_ϵ^2 to 1.

Figure 3 plots the three structural monetary policy shocks identified by the shock measures of Romer and Romer (2004), Barakchian and Crowe (2013), and Gertler and Karadi (2015). Compared to the raw series in Figure 1, the three structural shocks move stronger in tandem. Table 4 lists their correlations. Interestingly, the structural shocks identified using the two high-frequency series of Barakchian and Crowe (2013) and Gertler and Karadi (2015) have the highest correlation, 0.978. However, the narratively identified shock series has strong correlation with the other two series at almost 90%.

TABLE 4
Correlation of structural monetary policy shocks

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| RR | 1.000 | | |
| BC | 0.897 | 1.000 | |
| GK | 0.899 | 0.978 | 1.000 |

Appendix C Robustness results

Through several exercises, we determine the robustness of our main test findings. First, as originally suggested by Lunsford (2015), we remove the linear time trend from our VAR model and include instead the federal funds rate, the growth rate of industrial production, and the price index as well as the level of the excess bond premium. Second, we vary the lag order of our baseline VAR model specification.

(i) Test results for the VAR model with stationary variables

Table 5 shows the F-statistics for the VAR model without linear trend. It includes the federal funds rate, the growth rate of industrial production, and consumer price index, as well as the excess bond premium.

TABLE 5
Results with no linear time trend and stationary variables

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| F_{IV} | 19.334* | 12.672* | 42.129* |
| F_{WPP} | 4.58 | 3.09 | 8.724° |

Note: * indicates significance according to the rule of thumb $F_{IV} > 10$; ° indicates significance at the 5% level according to the critical value of Lunsford (2015) $F_{WPP} > 8.22$

(ii) Alternative lag order of the VAR model

(a) Table 6 shows the F-statistics for the baseline VAR model with a lag order of 36. The model is estimated with a linear time trend and includes the federal funds rate, industrial production, consumer price index, as well as the excess bond premium. As in Romer and Romer's (2004) original study, this is to allow for longer-lasting monetary policy effects.

TABLE 6
Results with 36 lags

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| F_{IV} | 8.572 | 5.676 | 33.771* |
| F_{WPP} | 1.762 | 4.149 | 10.573° |

Note: * indicates significance according to the rule of thumb $F_{IV} > 10$; ° indicates significance at the 5% level according to the critical value of Lunsford (2015) $F_{WPP} > 8.22$

(b) Table 7 shows the F-statistics for the baseline VAR model with a lag order of 3 selected by the AIC. The model is estimated with a linear time trend and includes the federal funds rate, industrial production, consumer price index, and the excess bond premium. We considered a maximum lag length of 50.

TABLE 7
Results with lag order selected by the AIC

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| F_{IV} | 32.233* | 6.849 | 17.96* |
| F_{WPP} | 3.347 | 2.533 | 8.093 |

Note: * indicates significance according to the rule of thumb $F_{IV} > 10$; ° indicates significance at the 5% level according to the critical value of Lunsford (2015) $F_{WPP} > 8.22$

(c) Table 8 shows the F-statistics for the baseline VAR model with a lag order of 2 selected by the BIC. The model is estimated with a linear time trend and includes

the federal funds rate, industrial production, consumer price index, and the excess bond premium. We considered a maximum lag length of 50.

TABLE 8
Results with lag order selected by the BIC

| | RR | BC | GK |
|-----------|-----------|-----------|-----------|
| F_{IV} | 23.013* | 8.899 | 20.993* |
| F_{WIP} | 5.176 | 2.855 | 9.677° |

Note: * indicates significance according to the rule of thumb $F_{IV} > 10$; ° indicates significance at the 5% level according to the critical value of Lunsford (2015) $F_{WIP} > 8.22$