Bayesian Inference in Structural Vector Autoregression with Sign Restrictions and External Instruments *

Lam Nguyen, University of California at San Diego†

June 23, 2018
Revised: October 10, 2019

Abstract

Instrument validity cannot be tested in a just-identified model, and it is not clear what conclusion to draw when instrument validity is rejected in an over-identified model. In practice researchers tend to regard instruments as valid when they lead to sensible inferences. This paper uses Bayesian methods to formalize this idea. I develop a proxy structural vector autoregression in which prior information from both theory and the empirical literature is incorporated about signs and magnitudes of certain parameters and equilibrium impacts. I use my method to investigate the relevance and validity of three popular instruments for monetary policy shocks, developed by Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007). I find that all of them are strongly relevant but only that of Smets and Wouters is valid. Furthermore, the empirical analysis demonstrates that my framework can combine information from a relevant and valid instrument with prior information about sign restrictions to improve inference about structural impulse-response functions.

Keywords: Structural vector autoregressions, Proxy SVAR, dynamic causal effects, sign restrictions, instrumental variables, subjective Bayesian inference, model selection, set-identification, monetary policy.

JEL Classifications: C1, C32, E47

---

*I am deeply grateful for the invaluable support and guidance from my adviser, Professor James D. Hamilton. I am also thankful to Professor Brendan Beare, Professor Roger Gordon, Professor Michelle White, and Professor Prashant Bharadwaj for their helpful advice during the third-year original paper sequences. Lastly, I am highly indebted to Amaze Lusompa, participants at the UCSD Graduate Students Research Seminar, UCSD Macroeconomics Workshop, UCSD Econometrics Workshop, and the 2018 Midwest Econometrics Group Conference for many illuminating discussions. I bear all responsibility for any remaining errors and omissions.

†Department of Economics, University of California at San Diego, CA, 92093, USA. Email: lhn022@ucsd.edu
1 Introduction

Starting with Sims (1980), empirical macroeconomists have been developing new robust identification strategies to perform dynamic causal inference in structural vector autoregressions (SVARs). As conventional identifying assumptions have come under intense scrutiny (Ramey (2016), Nakamura and Steinsson (2017)), identification by instrumental variables has emerged as a more credible method for causal analysis. The method, also known as Proxy SVAR, was popularized by Stock and Watson (2012a, 2018), Mertens and Ravn (2013), Gertler and Karadi (2015), and has been applied to a wide variety of settings.

Is Proxy SVAR really more credible? Its central premise is that imperfect measures can be used as instruments for the true shocks.¹ The key identifying assumptions are that the instruments are correlated with the true shocks (relevance) and uncorrelated with other structural shocks (validity).² Although the relevance assumption can be formally tested (Stock and Watson (2018)), the validity assumption is often defended by ad hoc arguments. Stock and Watson (2012a) provide suggestive evidence showing that this assumption might fail to hold for the majority of instruments in macroeconomics. However, without further information, the data cannot tell us whether this assumption holds for any particular instrument.

How, then, can researchers select a good instrument among many potentially invalid instruments? I propose to evaluate both instrument relevance and validity by using a subjective Bayesian approach that incorporates information from both theory and the empirical literature in the form of sign restrictions. The literature on sign restrictions are pioneered by Uhlig (2005), Faust (1998), and Canova and De Nicolo (2002), and it has increasingly gained popularity with many recent contributions, including Rubio-Ramirez et al. (2010), Arias et al. (2018a), Arias et al. (2019), Antolín-Díaz and Rubio-Ramírez (2018), and Baumeister and Hamilton (2015, 2018, 2019). The surge in popularity is due to the fact that sign restrictions are more consistent with economic theory and thus more robust to model misspecification. My paper suggests using the identifying assumptions from sign restrictions to evaluate the underlying assumption of the instruments.

To implement the idea, I generalize the SVAR framework of Baumeister and Hamilton (2015) to incorporate both sign restrictions and external instruments. Investigation of the validity and relevance assumptions will then be framed as a model selection problem. In particular, instrument validity is investigated by a Bayesian model comparison between a model where the validity assumption holds and one where it does not. If the data favor the first model, the instruments are judged to be valid, otherwise, they are invalid. Similarly, the method can be used to evaluate the relevance assumption.

I apply the technique to analyze the relevance and validity of three monetary policy instruments, proposed by Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007).³ I first use various tests

²In this paper, I use the terminology in Hamilton (1994). In some other texts, such as Stock and Watson (2012a), the authors define a valid instrument as an instrument that is relevant and exogenous. Their definition of instrument relevance is the same as mine, and their definition of instrument exogeneity is equivalent to my definition of instrument validity.
³The instrument of Romer and Romer (2004) is constructed by the narrative method, that of Sims and Zha (2006) is estimated from a Markov-switching SVAR model, and that of Smets and Wouters (2007) comes from a medium-scale DSGE
of overidentifying restrictions in a frequentist Proxy SVAR framework to demonstrate that at least some of the instruments are invalid and to illustrate the inability of existing methods to differentiate valid instruments from invalid ones. Then, I use the sign-restricted SVAR proposed by Baumeister and Hamilton (2018) as a benchmark to evaluate the underlying assumptions of those instruments. My main finding is that all three instruments are strongly relevant, but only the Smets-Wouters instrument is valid. The reason is that the Romer-Romer and Sims-Zha instruments are highly correlated with both monetary and demand shocks identified from the sign-restricted SVAR. This finding sheds light on Stock and Watson (2012a)’s observation that monetary and fiscal policy shocks, identified by their respective instruments, are highly correlated with each other.

Furthermore, the empirical application establishes that a Bayesian framework can use instruments to identify structural shocks just as in conventional approaches. I first use Romer and Romer (2004)’s specification to study the dynamics of macro aggregates in response to a monetary shock. Then I show that the sign-restricted SVAR augmented with the instrument produces structural impulse-response functions (SIRFs) with similar patterns. In particular, I find that both the output gap and inflation have hump-shaped responses following a monetary policy shock. This finding is consistent with those from previous literature as surveyed in Ramey (2016).

My paper is the first to evaluate the instrument relevance and validity assumptions by sign restrictions, and in doing so, it addresses three strands of literature. First is the econometric literature on SVAR identified by sign restrictions and instrumental variables. The major distinction between my framework and others is that mine transparently and flexibly incorporates both sign restrictions and instrumental variables. In contrast to Arias et al. (2018b), Ludvigson et al. (2017), and Braun et al. (2017), I use explicit priors and sign restrictions to estimate the SVAR parameters directly without using draws of the rotation matrix to convert the reduced-form VAR, a practice known to impose implicit priors on the dynamic causal effects (Baumeister and Hamilton (2015), Wolf (2018), Watson (2019)). Moreover, my method can easily use multiple proxies to identify a single shock, a common objective which cannot be achieved using Arias et al. (2018b)’s algorithm. Compared to the robust Bayesian method of Giacomini et al. (2019), my parameterization makes better use of information from the previous literature because the SVAR parameters have economic interpretation. Relative to other approaches in Bayesian Proxy SVAR, my method not only enjoys all the advantages over the frequentist counterpart as discussed in Caldara and Herbst (2019), Bahaj (2014), and Drautzburg (2016) but it also use sign restrictions and informative priors on various model quantities to improve causal inference.

Second, my paper contributes to the econometric literature on instrumental variable regression with imperfect instruments. Early work, such as Conley et al. (2012), Nevo and Rosen (2012), and Chan and Tobias (2015), relaxed the validity assumption by allowing the instruments to have a small, direct effect on the dependent variable. Although those papers acknowledge that the instrument validity is likely violated,
none of them attempts to formally investigate that assumption. More recently, Ludvigson et al. (2017) and Braun et al. (2017) consider sign-restricted SVAR models with external instruments where they restrict the correlations between the structural shocks and the instruments. Because of independent evidence that many instruments are likely invalid, using the instruments to restrict the model could remove many good models from the identified set and negatively affect inference. In contrast to their papers, I recommend researchers first use sign restrictions to evaluate the underlying assumptions of the instruments, and only incorporate those that have passed the first stage in the final model.

Finally, my paper contributes to the literature that aims to shrink down the identified set by incorporating more credible information. Tamer (2010) points out that the identified set in a partially-identified model is often too large to be useful for policy analysis. Consequently, a large literature has tried to incorporate information that help sharpen the identified set, including Ludvigson et al. (2017), Braun et al. (2017), Arias et al. (2018b), Antolin-Diaz and Rubio Ramírez (2016), Amir-Ahmadi and Drautzburg (2017), and Ahmadi and Uhlig (2015). Relative to these papers, I directly model the structural shocks as a linear function of exogenous variables, and hence I can flexibly incorporate many different type of information and exogenous factors. Although my empirical application only uses information from the current value of the instrument, the algorithm can implement sign restrictions on the effect of the instruments on the structural shocks as suggested by Gafarov (2014) or incorporate lags of the instruments to deal with anticipation in rational expectation models as illustrated in Noh (2018).

The rest of this paper is organized as follows. Section 2 describes the framework and estimation strategy. Section 3 applies the method to investigate the relevance and validity of three monetary policy instruments. Section 4 briefly concludes. Additional technical details are shown in the appendices.

2 Sign-restricted SVAR with instrumental variables

2.1 Model Description

Suppose that the dynamics of the data are summarized by a SVAR(p)

$$Ay_t = k + B_1 y_{t-1} + B_2 y_{t-2} + \cdots + B_p y_{t-p} + u_t$$

$$u_t \sim \text{i.i.d } N(0, D^*)$$

where $y_t$ and $u_t$ are $(n \times 1)$ vector of observed variables and structural shocks at time t. $A$ is an $(n \times n)$ matrix that governs contemporaneous relationship between observed variables, $B_i$ ($i = 1, \ldots, p$) is an $(n \times n)$ matrix of lag coefficients, and $D^*$ is an $(n \times n)$ diagonal covariance matrix of the structural shocks. For

---

4Piffer (2017) and Lamne et al. (2016) propose Bayesian methods to assess the validity of sign restrictions, but those papers do not discuss instrumental variables.
simplicity, the model is rewritten as

$$Ay_t = Bx_{t-1} + u_t$$  \hfill (1)

where $B = \begin{bmatrix} B_1 & B_2 & \ldots & B_p \end{bmatrix}$ is $[n \times (np + 1)]$ matrix, and $x_{t-1} = \begin{bmatrix} y'_{t-1} & y'_{t-2} & \ldots & y'_{t-p} & 1 \end{bmatrix}'$ is $[(np + 1) \times 1]$ vector.

Suppose we want to incorporate a $(q \times 1)$ vector of instruments, $z_t = [z_{t}^{(1)}, z_{t}^{(2)}, \ldots, z_{t}^{(q)}]$, into the model. We can use the linear projection of the structural shocks on these instruments

$$u_t = Cz_t + w_t$$ \hfill (2)

$$w_t \sim \text{i.i.d } N(0, D)$$ \hfill (3)

where $C$ is a $(n \times q)$ matrix that governs the relationship between the instruments and the endogenous variables, and $D$ is again assumed to be diagonal. The general model with the instruments can be written as

$$Ay_t = Bx_{t-1} + Cz_t + w_t$$ \hfill (4)

Depending on different specifications of $C$ and the choice of the instruments $z_t$, model (4) can be used to (1) identify one structural shock by one instrument, (2) identify one structural shock by multiple instruments, (3) identify multiple structural shocks by multiple instruments, or (4) incorporate both the current and lagged values of the instruments and other exogenous factors.

The general model (4) nests three important special cases regarding the instrument properties. Without loss of generality, suppose we use all instruments to identify the last structural shock in the system. First, if the instruments are \textit{irrelevant but valid}, $C$ will be restricted to be identically zero. On the other extreme, if we have \textit{relevant but invalid} instruments, $C$ will be totally unrestricted. Lastly, \textit{relevant and valid} instruments will restrict the elements of the $C$ matrix to be

$$C = \begin{bmatrix} \mathbf{0}_{(n-1) \times q} \\ c_{n1} \\ \vdots \\ c_{nq} \end{bmatrix}$$ \hfill (5)

This structure of the matrix $C$ ensures that the instruments are correlated with only one structural shock (relevant instruments) but not with other shocks (valid instruments).

\subsection*{2.2 Discussion}

This section briefly discusses the identification problem in SVAR, how researchers use instrumental variables to identify the dynamic causal effects, and how they can use the identifying assumptions from sign restrictions to evaluate the instrument relevance and validity assumptions.
The SVAR(p) described in (1) admits a reduced-form VAR(p) representation

$$\mathbf{y}_t = \Pi_1 \mathbf{x}_{t-1} + \mathbf{e}_t$$

where $$\Pi_1 = \mathbf{A}^{-1} \mathbf{B}$$ and $$\mathbf{e}_t \sim \mathcal{N}(0, \Omega^*)$$ with $$\Omega^* = \mathbf{A}^{-1} \mathbf{D}^* \left( \mathbf{A}^{-1} \right)'$$. These parameters are identified from the data and can be consistently estimated by the Ordinary Least Squares (OLS) method. Nevertheless, the parameters $$\mathbf{A}$$, $$\mathbf{B}$$, and $$\mathbf{D}^*$$ are generally not identified because $$\mathbf{A}$$ and $$\mathbf{D}^*$$ may have more parameters than the reduced-form covariance matrix, $$\Omega^*$$. Conventional approaches, such as exclusion restrictions, use quantitative restrictions to achieve point-identification. For example, researchers can normalize the elements of the matrix $$\mathbf{D}^*$$ to one and set the matrix $$\mathbf{A}$$ to be lower-triangular. Those restrictions impose a particular causal order among the endogenous variables, and hence they are both conceptually and empirically controversial (Ramey (2016), Nakamura and Steinsson (2017)).

Identification by instrumental variables has become a promising alternative recently. To understand the method, consider the reduced-form representation of the SVAR(p) with instrumental variables in (4)

$$\mathbf{y}_t = \Pi_1 \mathbf{x}_{t-1} + \Pi_2 \mathbf{z}_t + \mathbf{e}_t$$

where $$\Pi_1 = \mathbf{A}^{-1} \mathbf{B}$$, $$\Pi_2 = \mathbf{A}^{-1} \mathbf{C}$$, and $$\mathbf{e}_t \sim \mathcal{N}(0, \Omega)$$ with $$\Omega = \mathbf{A}^{-1} \mathbf{D} \left( \mathbf{A}^{-1} \right)'$$. Let $$\mathbf{A}^{-1} = \mathbf{\Psi} = \left[ \psi^1 \psi^2 \ldots \psi^n \right]$$ and consider the specification of $$\mathbf{C}$$ described in (5), we have

$$\Pi_2 \mathbf{z}_t = \mathbf{\Psi} \mathbf{C} \mathbf{z}_t = \psi^n c_{n,1} z^{(1)}_t + \psi^n c_{n,2} z^{(2)}_t + \cdots + \psi^n c_{n,q} z^{(q)}_t$$

If the instruments are relevant and valid, they will identify the last column of $$\mathbf{A}^{-1}$$, $$\psi^n$$, up to some scale factors. The last column of $$\mathbf{A}^{-1}$$ will then be used to infer the dynamic causal effects of the last structural shocks on other macro aggregates. On the other hand, if $$\mathbf{C}$$ does not satisfy the restrictions in (5), the instruments will only identify some non-linear combinations of the structural parameters and hence be useless for dynamic causal inference.

Although instrument validity is crucial to the success of this approach, existing methods have little to say about its legitimacy. When researchers have only one instrument, they cannot determine whether the matrix $$\mathbf{C}$$, now a vector, satisfies the restrictions in (5). Consequently, they often must rely on ad hoc arguments to argue that the instrument is "plausibly exogenous". And although those restrictions can be tested when researchers use more than one instruments to identify a structural shock, it is not clear what conclusion to draw about the validity of any particular instrument without further information.

Nevertheless, in many cases, researchers do have other identifying assumptions from theory and the empirical literature about reasonable magnitudes and signs of different model quantities. These assumptions can be incorporated as prior belief and sign restrictions on the elements of $$\mathbf{A}$$ or $$\mathbf{A}^{-1}$$. Because sign restrictions
are grounded in economic theory and robust to model misspecification, this paper proposes to use them as
overidentifying assumptions in order to formally test the underlying assumptions of the instruments. To see
the intuition of my approach, consider the problem of a researcher who has some doubt about instrument
validity. From the Bayesian point of view, her belief after seeing the data can be expressed as an odds ratio
between two models: a model where the instrument is valid, and a model where it is not. Conditional on
the data, if the first model is more likely than the second, she will conclude that the instrument is valid.
Whereas, if the first model is less likely, she will conclude that it is not. Similarly, instrument relevance can
be formally assessed by comparing a model where the instrument is assumed to be valid and relevant with
a model where the instrument is assumed to be irrelevant.

Formally, let $M_0$ be a special case of (4) where the matrix $C$ is identically zero, $M_1$ be the model where
the matrix $C$ satisfies the restrictions in (5), and $M_2$ be the model where the matrix $C$ is left unrestricted.
Thus, $M_0$ imposes the restrictions that the instruments are not relevant, $M_1$ imposes the restrictions that
the instruments are valid, and $M_2$ does not impose any restriction. If a Bayesian model comparison favors
$M_1$ over $M_0$, the instruments are judged to be relevant. Similarly, if a Bayesian model comparison favors
$M_1$ over $M_2$, the instruments are deemed to be valid.

The robustness of sign restrictions and prior belief play important roles in the success of this method.
Wolf (2016) cautions that sign restrictions might identify both the true shocks and mixtures of other shocks
that have different dynamics, while Paustian (2007), Canova and Paustian (2011), and Gafarov (2014), and
Wolf (2018) assert that the sign-restricted SVAR is more robust and delivers the correct causal inference if
researchers are willing to impose an appropriate number of sign restrictions. The required number of sign
restrictions would depend on a particular application. In any case, I note that my algorithm can flexibly
accommodate sign restrictions in both structural parameters, $A$ and $C$, and impact effects of structural
shocks, $A^{-1}$, along with other conventional identifying strategies like short-run and long-run restrictions.
Furthermore, in set-identified models such as sign-restricted SVARs, inference is sensitive to prior belief even
in large sample. In particular, there exists quantities of which there is no Bayesian updating (Lindley (1957),
Poirier (1998)), and Bayesian credible set no longer satisfy frequentist coverage even with infinite amount of
data (Moon and Schorfheide (2012)). Following Baumeister and Hamilton (2015), I acknowledge and deal
with this problem by using explicit, informative priors carefully constructed from both theory and practice.

2.3 MCMC Algorithm for Estimation

This section describes the priors, the likelihood, and the MCMC algorithm to simulate from the posterior
distributions in details. For estimation purposes, the joint prior of $(A, B, C, D)$ can be decomposed as

I allow arbitrary priors on parameters of $A, C$ and employ natural conjugates for the two conditional priors $P(D|A, C)$ and $P(B|A, C, D)$ to ease computation of the posteriors $P(A, B, C, D|Y_T)$. Specifically, I use independent inverse-Gamma priors for the variances of the structural shocks

$$p\left(d_{ii}^{-1}|A, C\right) \sim \Gamma(\kappa_i, \tau_i)$$

(9)

$$P(D|A, C) = \prod_{i=1}^{n} p\left(d_{ii}^{-1}|A, C\right)$$

(10)

where $\Gamma(\kappa_i, \tau_i)$ denotes the Gamma distribution with parameters $\kappa_i$ and $\tau_i$. For the lag parameters, I use multivariate normal priors that are independent across equations

$$P(b_i|A, C, D) \sim N(m_i, d_{ii}M_i)$$

(11)

$$P(B|A, C, D) = \prod_{i=1}^{n} P(b_i|A, C, D)$$

(12)

where $N(m_i, d_{ii}M_i)$ denotes the multivariate normal distribution with mean $m_i$ and covariance matrix $d_{ii}M_i$. The overall prior is

$$P(A, B, C, D) = P(A, C) \prod_{i=1}^{n} p\left(d_{ii}^{-1}|A, C\right) \prod_{i=1}^{n} P(b_i|A, C, D)$$

(13)

I condition on the instrumental variables in the estimation and make use of the conditional likelihood, which is

$$P(Y_T|A, B, C, D, Z_T) = (2\pi)^{-Tn/2} \det |A|^T \det |D|^{-T/2} \times \exp \left[-(1/2) \sum_{t=1}^{T} (Ay_t - Cz_t - Bx_{t-1})' D^{-1} (Ay_t - Cz_t - Bx_{t-1}) \right]$$

(14)

Given the priors and the likelihood, the posteriors will be characterized by Bayes’ rule. The joint posterior of $A$ and $C$ is calculated by the random-walk Metropolis-Hasting algorithm, while the posteriors of $B$ and $D$ are their respective natural conjugates. The estimation procedure is stated formally below.

**Proposition 1.** Let the priors be given as in (8)-(13) and the likelihood function be given as in (14). Moreover, let $a_i'$ denote the $i$-th row of $A$, $c_i'$ denote the $i$-th row of $C$, $P_i$ denote the Cholesky factor of $M_i^{-1} = P_i P_i'$. Then, the posteriors are

$$P(A, B, C, D|Y_T, Z_T) = P(A, C|Y_T, Z_T) \prod_{i=1}^{n} p\left(d_{ii}^{-1}|A, C, Y_T, Z_T\right) \prod_{i=1}^{n} P(b_i|A, C, D, Y_T, Z_T)$$

(15)

---

6The motivation for using conditional likelihood is provided in Appendix A.
\[
P(A, C|Y_T, Z_T) = \frac{k_T P(A, C) \left[ \det \left( A\hat{\Omega}A^t \right) \right]^{T/2}}{\prod_{i=1}^{n} [2\tau_i^*/T]^{\kappa_i^*}} \tag{16}
\]

\[
p \left( d_{ii}^{-1}|A, C, Y_T, Z_T \right) \sim \Gamma (\kappa_i^*, \tau_i^*) \tag{17}
\]

\[
P(b_i|A, C, D, Y_T, Z_T) \sim N (m_i^*, d_{ii}^*) \tag{18}
\]

\[
\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t' \tag{19}
\]

\[
\hat{\epsilon}_t = y_t - \hat{\Phi} x_{t-1} \tag{20}
\]

\[
\hat{\Phi} = \left( \sum_{i=1}^{T} y_i x_{i-1}' \right) \left( \sum_{i=1}^{T} x_i x_{i-1}' \right)^{-1} \tag{21}
\]

\[
\kappa_i^* = \kappa_i + (T/2) \tag{22}
\]

\[
\tau_i^* = \tau_i + (\zeta_i^*/2) \tag{23}
\]

\[
\zeta_i^* = (\tilde{Y}_i' \tilde{Y}_i) - (\tilde{Y}_i' \tilde{X}_i) \left( \tilde{X}_i' \tilde{X}_i \right)^{-1} (\tilde{X}_i' \tilde{Y}_i) \tag{24}
\]

\[
m_i^* = (\tilde{X}_i' \tilde{X}_i)^{-1} (\tilde{X}_i' \tilde{Y}_i) \tag{25}
\]

\[
M_i^* = (\tilde{X}_i' \tilde{X}_i)^{-1} \tag{26}
\]

\[
\tilde{Y}_i = \left[ y_1 \ldots y_T y_T a_i - c_i z_{i1} \ldots y_T a_i - c_i z_{iT} \ldots \right]^t \tag{27}
\]

\[
\tilde{X}_i = \left[ x_0' \ldots x_{t-1}' P_i \right]^t \tag{28}
\]

The differences between this proposition and that in Baumeister and Hamilton (2015) are equations (16) and (27). In particular, the target for the Metropolis-Hasting algorithm in (16) takes into account the priors for the matrix \( C \), and (27) redefines \( \tilde{Y}_i \) to account for the instrumental variables \( z_i \). The estimation procedure and its proof are identical to those in their paper.

### 2.4 Bayesian Model Comparison

#### 2.4.1 Bayes factor

Bayesian model comparison is used to evaluate instrument relevance and validity. Generally, a Bayesian model comparison between model \( k \) and \( k' \) is done by calculating the posterior odds ratio, showing which model is more likely now that we have seen the data.\(^7\) This can be written as the product of the Bayes factor and the prior odds ratio.

\[
\frac{P(M_k|Y_T, Z_T)}{P(M_{k'}|Y_T, Z_T)} = \frac{P(Y_T|Z_T, M_k)}{P(Y_T|Z_T, M_{k'})} \frac{P(M_k)}{P(M_{k'})} \tag{29}
\]

Suppose before seeing the data, we assume model \( k \) is as likely as model \( k' \), then the prior odds ratio (i.e.

\(^7\)Since I condition on the instruments, they don’t affect the calculation of the odds ratio. Appendix A provides more details.
\( P(M_k) / P(M_{k'}) \) will be one and equation (29) simplifies to

\[
\frac{P(M_k | Y_T, Z_T)}{P(M_{k'} | Y_T, Z_T)} = \frac{P(Y_T | Z_T, M_k)}{P(Y_T | Z_T, M_{k'})}
\]  

(30)

The quantity on the right-hand side of equation (30) is the so-called Bayes factor, which is the ratio of two marginal likelihoods. The Bayes factor between model \( k \) and model \( k' \) is denoted as

\[
B_{k,k'} = \frac{P(M_k | Y_T, Z_T)}{P(M_{k'} | Y_T, Z_T)} = \frac{P(Y_T | Z_T, M_k)}{P(Y_T | Z_T, M_{k'})}
\]

The Bayes factor shows how likely model \( k \) is relative to model \( k' \). For example, the value of two means that model \( k \) is twice as likely as model \( k' \) after the data are observed.

2.4.2 Marginal likelihood estimation

The Bayes factor is the ratio of two high dimensional integrals. Let \( \theta \) denote all unknown parameters. The marginal likelihood of model \( k \) is defined as

\[
P(Y_T | Z_T, M_k) = \int P(Y_T | \theta, Z_T, M_k) P(\theta | M_k) \, d\theta
\]

(31)

where \( P(Y_T | \theta, Z_T, M_k) \) is the likelihood of model \( M_k \), and \( P(\theta | M_k) \) is the prior, which is assumed to be independent of \( Z_T \). Let \( f(\theta) \) be a known multivariate density function of \( \theta \). Bayes’ theorem implies that

\[
\frac{1}{P(Y_T | Z_T, M_k)} = \int \frac{f(\theta)}{P(Y_T | \theta, Z_T, M_k) P(\theta | M_k)} P(\theta | Y_T, Z_T, M_k) \, d\theta
\]

Thus, a natural estimator for the marginal likelihood from the posterior draws would be

\[
\hat{P}(Y_T | Z_T, M_k) = \left[ \frac{1}{N_0} \sum_{n=1}^{N_0} \frac{f(\theta^n)}{P(Y_T | \theta^n, Z_T, M_k) P(\theta^n | M_k)} \right]^{-1}
\]

(32)

where \( N_0 \) is the number of posterior draws after discarding the burn-in sample. The choice of \( f(\theta) \) is important in the calculation of the marginal likelihood. For instance, if we choose \( f(\theta) \) to be the prior distribution (i.e. \( f(\theta) = P(\theta | M_k) \)), we will have a harmonic mean estimator

\[
\hat{P}_{HMM}(Y_T | Z_T, M_k) = \left[ \frac{1}{N_0} \sum_{n=1}^{N_0} \frac{1}{P(Y_T | \theta^n, Z_T, M_k)} \right]^{-1}
\]

However, this estimator has infinite variance and is numerically inefficient. For the method to work well, \( f(\theta) \) needs to be a good approximation of the posterior distribution and has a thinner tail than the posterior kernel, \( P(Y_T | \theta, Z_T, M_k) P(\theta | M_k) \), to ensure convergence of the Monte Carlo average. Geweke (1999) proposes to use a truncated normal distribution, while Sims et al. (2008) constructs a more sophisticated choice for \( f(\theta) \).

\(^8\theta = (\text{vec}(A), \text{vec}(B), \text{vec}(C), \text{vec}(\text{diag}(D)))^\top\)
Appendix B describes those two choices in details.

3 Application: Three instruments for monetary shock

This section applies the above method to investigate the relevance and validity of three instruments for monetary shocks. It also shows how my method can combine information from a relevant and valid instrument with sign restrictions to improve dynamic causal inference in SVAR. I first describe the data then present the analysis from both frequentist and Bayesian perspectives to illustrate the benefits of my approach.

3.1 Data Description

The data are publicly available and are described in detail in Appendix C. Macroeconomic time series are downloaded from FRED, while monetary policy instruments are collected from Mark Watson and Yuriy Gorodnichenko’s website. The sample period is 1954Q3 to 2008Q4. Four quarterly time series are used: real GDP, real potential GDP, personal consumption expenditures deflator (PCE deflator), and effective federal funds rate. The output gap is calculated as the log difference between real and potential GDP, and the inflation rate is computed as the Y/Y change of the PCE deflator. The instruments are similar to those in Stock and Watson (2012a); they come from Romer and Romer (2004)’s narrative method, Smets and Wouters (2007)’s DSGE model, and Sims and Zha (2006)’s Markov-Switching SVAR. Missing values of the instruments are replaced with zeros as done in Romer and Romer (2004).

Figure 1 plots the macroeconomic aggregates and the instruments. It shows that both macroeconomic variables and monetary policy instruments are less volatile after the 1980s, which is consistent with Stock and Watson (2002)’s documentation of the Great Moderation. A dampening in the volatility of monetary policy instruments is also consistent with Ramey (2016)’s observation that monetary policy has been conducted in a more systematic manner after 1980, and thus true monetary policy surprises are hard to identify from the data.

Table 1 shows the summary statistics of the data, and Table 2 displays the pairwise cross-correlation of the instruments. Table 2 paints a similar picture to that in Stock and Watson (2012a), namely monetary policy instruments are not all highly correlated with each other. The Romer-Romer instrument appears to be highly correlated with Sims-Zha instrument and less correlated with that of Smets-Wouters. Although the correlations between instruments are not the same as the correlations between their predictive shocks, their lack of correlation is suggestive evidence that they might not identify the same shock. One possible explanation is that different instruments capture different dimensions of monetary policy shocks, and another explanation is that some of the instruments are contaminated by other contemporaneous shocks and hence invalid. In the following investigation, I will assume that all three monetary policy instruments identify the

---

9 As a robustness check, I also consider two other updated Romer-Romer monetary policy instruments from Wieland and Yang (2016) The results are qualitatively the same.
same monetary policy shock and use a sign-restricted SVAR model to study the latter explanation.

3.2 Proxy SVAR and tests of overidentifying restrictions

This section uses frequentist methods to analyze instrument relevance and validity. To implement the Proxy SVAR, Stock and Watson (2018) recommend the following IV regression

\[
\begin{align*}
    y_t &= \alpha_0 + \alpha_1 r_t + b' x_{t-1} + \varepsilon_t \\
    r_t &= \gamma_0 + c' z_t + d' x_{t-1} + \zeta_t
\end{align*}
\]

where \(y_t\) is the output gap, \(r_t\) is the effective fed funds rate, \(z_t = (z_{1t}, z_{2t}, \ldots, z_{qt})'\) is a \((q \times 1)\) vector of monetary shocks instruments, and \(x_{t-1} = (y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p})'\) is a \((k \times 1)\) vector of lag variables in the VAR. Let \(u_t = (u^s_t, u^d_t, u^m_t)\) denote the supply shock, the demand shock, and the monetary policy shocks respectively. Stock and Watson (2018) show that we can consistently estimate \(\alpha_1\) and use it to study the dynamic causal effect of monetary policy shock on the output gap under three assumptions

1. Instrument relevance: \(\mathbb{E}(u^m_t z_t) \neq 0\)
2. Instrument validity: \(\mathbb{E}(u^s_t z_t) = 0\) and \(\mathbb{E}(u^d_t z_t) = 0\)
3. Invertibility: \(v_t = Qu_t\) where \(v_t\) is the innovations in the reduced-form VAR and \(Q\) is a \((3 \times 3)\) rotation matrix

Assumption (1) and (2) are standard assumptions in IV regressions to achieve consistency of the estimator, while assumption (3) is required in the SVAR setting to ensure that \(\alpha_1\) can be used to infer the dynamic effect of monetary policy shock. Intuitively, assumption (3) means that monetary policy shocks can be recovered from the current and past information of the output gap, inflation, and fed funds rate. Although Stock and Watson (2018) caution that invertibility might fail when the variables in the VAR are not sufficient to account for the Fed information set or when economic agents are forward-looking, Wolf (2018) shows that monetary policy shocks are nearly invertible even for a standard three-variable VAR. In this application, I will only focus on (1) and (2).

To analyze instrument relevance, I first estimate the above IV regression and let each instrument enter one by one. Table 3 reports the partial R-squared, F-statistics, heteroskedasticity-robust F-statistics, and their corresponding p-values for all three instruments.\(^{10}\) Both the normal and robust F-statistics are significantly higher than 10, and their p-values are indistinguishable from zero. Thus, the empirical evidence suggests that all three instruments are strong. I note that the first-stage F-statistics are higher than those found in Stock and Watson (2012a). One potential explanation is that their study employs a Factor-SVAR model with more than 200 macroeconomic time series whereas my model only has three time series.

\(^{10}\)The partial R-squared measures how much the variation in the dependent variable is explained by the variation of the instruments excluded other exogenous variables.
Next, I use tests of overidentifying restrictions to investigate the validity of the instruments. I conduct two separate exercises. The first uses all three instruments at once, and the second uses two instruments at a time. I use three different tests, which are developed by Sargan (1958), Basmann (1960), and Wooldridge (1995). The first two tests assume that the error-terms are i.i.d., whereas the last one controls for heteroskedastic errors. The null hypothesis is that all three instruments are valid. Table 4 reports the test statistics and their corresponding p-values for the first exercise. The Sargan’s score and the Basmann’s test deliver p-values that are close to zero, while the Wooldrige’s score has a p-value around 6%. Hence, the overall results suggest that at least one instrument is invalid.

Table 5 summarizes the results from the second exercise and sheds new light on the first results. All three tests fail to reject the null hypothesis when the Romer-Romer and Sims-Zha instruments are used together, however, they strongly reject for the other two cases. A failure to reject in overidentifying-restriction tests does not imply that the instruments are valid but only implies that the IV estimates from those instruments are similar to each other. Thus, the results suggest that the Romer–Romer and Sims-Zha instruments deliver similar IV estimates, which are different from those estimated by the Smets-Wouters instrument.

In summary, frequentist evidence reveals that all three instruments are strong but at least one of them is not valid. However, they cannot say which instrument is valid and which one is not because the null hypothesis that only one instrument is valid is untestable under the frequentist paradigm.

### 3.3 A Bayesian proxy SVAR of monetary policy

This section applies my Bayesian approach to investigate instrument relevance and validity. I first introduce the trivariate system of monetary policy in Baumeister and Hamilton (2018) which consists of the output gap, inflation, and nominal fed funds rate. Let $y_t$ be the output gap, $\pi_t$ be the inflation rate, $r_t$ be the effective fed funds rate, $z_t$ be the monetary policy instrument (i.e. Romer-Romer instrument, Sims-Zha instrument, or Smets-Wouters instrument), and $x_{t-1}$ be the vector of lag variables in the SVAR, the model is characterized by the following structural equations

1) The Phillips curve:
\[ y_t = k^s + \alpha^s \pi_t + [b^s]'x_{t-1} + u^s_t \]  \hspace{1cm} (33)

2) The aggregate demand equation:
\[ y_t = k^d + \beta^d \pi_t + \gamma^d r_t + [b^d]'x_{t-1} + u^d_t \]  \hspace{1cm} (34)

3) The Monetary Policy Rule:
\[ r_t = k^m + (1 - \rho) \psi^y y_t + (1 - \rho) \psi^\pi \pi_t + + \rho r_{t-1} + [b^m]'x_{t-1} + u^m_t \]  \hspace{1cm} (35)
I impose four sign restrictions on the structural parameters: (1) the Phillips curve is downward sloping \( (\alpha_s > 0) \), (2) raising interest rate will not stimulate aggregate demand \( (\gamma^d < 0) \), (3) the Fed will raise interest rate when the inflation rate is higher than its target \( (\psi^y > 0) \) or the output gap is higher than its potential \( (\psi^\pi > 0) \), and (4) the Fed wants to increase its interest rate smoothly over time \( (0 < \rho < 1) \). I do not use any prior on the impact effect of monetary policy shocks on those three variables because these effects should be identified by the instruments. Together, equation (33), (34), and (35) constitute the baseline model \( M_0 \) where the instrument is irrelevant.

If one has an instrument for monetary policy shock, one can model the monetary policy shock as

4) Monetary Policy shock:

\[
\mathbf{u}_t^m = \chi^m \mathbf{z}_t + \omega_t^m
\]  

(36)

where \( \chi^m \) measures the effect of the instrument on the monetary policy shock, and \( \omega^m \) is the part of the monetary policy shock that is orthogonal to the instrument as well as other shocks in the system. Equation (33), (34), (35), and (36) constitute the first alternative model \( M_1 \) where the instrument is assumed to be valid.

Furthermore, if one has doubt about the validity of the instrument, one can investigate that assumption by augmenting model \( M_1 \) with two additional equations

5) Supply shock:

\[
\mathbf{u}_t^s = \chi^s \mathbf{z}_t + \omega_t^s
\]  

(37)

6) Demand shock:

\[
\mathbf{u}_t^d = \chi^d \mathbf{z}_t + \omega_t^d
\]  

(38)

where \( \chi^s \) and \( \chi^d \) measure the effects of the instruments on supply and demand shocks respectively. If either \( \chi^s \) or \( \chi^d \) is different from zero, that will be evidence against instrument validity. The shocks \( u^m \), \( u^s \), and \( u^d \) are assumed to be orthogonal to each other as in the general formulation. Thus, the model where the instrument is allowed to be invalid, \( M_2 \), is summarized by six equations (33), (34), (35), (36), (37), and (38).

### 3.3.1 Priors for contemporaneous coefficients and shocks

This section discusses the priors for the contemporaneous parameters and the coefficients in the monetary, supply, and demand shock equations. I assume that the structural parameters in \( \mathbf{A}, \mathbf{C} \) are independent of each other. In particular, the priors are

1. Baseline model where the instruments are restricted to be irrelevant \( (M_0) \):

\[
P(\mathbf{A}, \mathbf{C}) = p(\alpha^s) p(\beta^d) p(\gamma^d) p(\psi^y) p(\psi^\pi) p(\rho)
\]
2. Model where the instrument is assumed to be valid and generally relevant ($M_1$):

$$P (A, C) = p(\alpha^s) p(\beta^d) p(\gamma^d) p(\psi^y) p(\psi^\pi) p(\rho) p(\chi^m)$$

3. Model where the instrument is allowed to be invalid ($M_2$):

$$P (A, C) = p(\alpha^s) p(\beta^d) p(\gamma^d) p(\psi^y) p(\psi^\pi) p(\rho) p(\chi^s) p(\chi^d) p(\chi^m)$$

I follow Baumeister and Hamilton (2018) in using additional prior information about elements of $A$ and do not repeat their motivations in this paper. My main innovation here is the specification of the matrix $C$. I set the priors for all elements of $C$ to be centered at zero with a small variance. I use symmetric priors because it is hard to tell how the instruments will affect the structural shocks before seeing the data. Also, because the main interest is whether those parameters are different from zero, a tight prior centered at zero will help avoid the Bartlett-Jeffreys-Lindley’s paradox (Jeffreys (1967), Lindley (1957), Bartlett (1957)). The phenomenon occurs when a sharp null hypothesis is rejected by frequentist methods but nonetheless received a high posterior odds based on a Bayesian analysis with small prior probability for the null and diffuse prior for the alternative. To avoid this apparent disagreement between frequentist and Bayesian approaches, Kass and Raftery (1995) recommend using a proper prior with small scale for the tested parameter, essentially ruling out the use of improper priors. In this application, I set the prior belief for $\chi^m$ to be a student $t$ distribution centered at 0 with scale 0.2 and 3 degrees of freedom in model $M_1$. This prior is proper, has a small variance, and allows for the possibility of a weak instrument. Similarly, I use that same prior for all three parameters $\chi^s$, $\chi^d$, and $\chi^m$ in model $M_2$. Thus, these proper priors allow for the possibility of an invalid instrument.

### 3.3.2 Priors for the covariance matrix

This section describes the prior for $B, D$ conditional on $A, C$. They are similar for all three models. As in the general formulation, I set the prior belief for elements of the covariance matrix to be independent from each other

$$P (D | A, C) = \prod_{i=1}^{3} p(d_{ii}^{-1} | A, C)$$

where each of the element $d_{ii}$ follows an inverse-Gamma distribution

$$p(d_{ii}^{-1} | A, C) = \begin{cases} \frac{\tau_{ii}^{\kappa_i} (d_{ii}^{-1})^{\kappa_i - 1} \exp \left( -\tau_i d_{ii}^{-1} \right)}{\Gamma(\kappa_i)} & \text{for } d_{ii}^{-1} \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$
In this specification, $\kappa_i / \tau_i$ is the prior mean for $d_{ii}^{-1}$, and $\kappa_i / \tau_i^2$ is the prior variance. Since the prior of the variances should reflect the scale of the data, I first fit three univariate AR(4) model to the data

$$ y_t = \beta_{10} + \sum_{i=1}^{4} \beta_{1i} y_{t-i} + e_{1t} $$

$$ \pi_t = \beta_{20} + \sum_{i=1}^{4} \beta_{2i} \pi_{t-i} + e_{2t} $$

$$ r_t = \beta_{30} + \sum_{i=1}^{4} \beta_{3i} r_{t-i} + e_{3t} $$

Then, I calculate the fitted residuals from those regressions: $\hat{e}_t = [\hat{e}_{1t}, \hat{e}_{2t}, \hat{e}_{3t}]'$, and estimate the sample covariance matrix:

$$ \hat{S} = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t' $$

Finally, I set $\kappa_i = 2$ for all $i$, which gives my prior a weight that equals to four observations of the data in the posterior. Next, I set the prior mean for $d_{ii}^{-1}$ to be the reciprocal of the $i^{th}$ diagonal element of $A\hat{S}A$.

In sum, the prior density for elements of the covariance matrix will be

$$ p \left( d_{ii}^{-1} \mid A, C \right) \sim \Gamma \left( 2, 2a_i' \hat{S}a_i \right) $$

### 3.3.3 Priors for the lag coefficients

I set the priors for the structural lag coefficients in such a way that the reduced-form lag coefficients are consistent with a Minnesota prior. Specifically, the priors for lag coefficients are assumed to be independent across equations

$$ P \left( B \mid A, C, D \right) = \prod_{i=1}^{3} P \left( b_i \mid A, C, D \right) $$

where $P \left( b_i \mid A, C, D \right) \sim N(\mathbf{m}_i, d_{ii} \mathbf{M}_i)$. For the mean, I set $\mathbf{m}_i(\alpha) = \eta' \mathbf{a}_i$, where $\eta = \begin{bmatrix} 1 & 0_{3 \times 10} \end{bmatrix}$.

And for the covariance matrix $\mathbf{M}_i$, let $\sqrt{s_{ii}}$ be the estimated standard deviation of the AR(4) that fits to variable $i$. I define:

$$ \mathbf{v}_1' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 12\lambda_1 & 2\lambda_1 & 3\lambda_1 & 4\lambda_1 \end{bmatrix} $$

$$ \mathbf{v}_2' = \begin{bmatrix} 1 \\ 2 \lambda_1 \\ 3 \lambda_1 \\ 4 \lambda_1 \end{bmatrix} $$

Then, $\mathbf{M}_i$ will be the diagonal matrix whose row $r$ column $r$ element is the $r^{th}$ element of $\mathbf{v}_3$: $M_{i,rr} = v_{3r}$.

Intuitively, we are letting coefficients on higher lags to shrink to zero by setting decreasing values for diagonal elements of $\mathbf{M}_i$. The hyper-parameter $\lambda_0$ captures how much confidence we have in the prior. A higher value

---

\(^{11}\)Note that if the reduced-form lag coefficients ($\Phi$) follow the Minnesota prior (meaning $E(\Phi) = \eta$). Then, $E(B \mid A) = E(A\Phi \mid A) = A E(\Phi \mid A) = \eta A$. Thus, if the prior for the reduced-form lag coefficients are the Minnesota prior, then the prior for the lag coefficients in the structural model are normal with mean $\mathbf{m}_i(\alpha) = \eta' \mathbf{a}_i$. 


implies a higher variance and less confidence. The hyper-parameter $\lambda_1$ governs how quickly the coefficients shrink to zero. And the hyper-parameter $\lambda_3$ describes the confidence in the prior for the constant, the higher the value, the less confidence we have. I set $\lambda_0 = 0.2$, $\lambda_1 = 1$, and $\lambda_3 = 100$.

In addition, the third element of the monetary policy rule equation is expected to be close to $\rho$. This prior information is captured using a prior for the third element of $b_3$ as follows

$$\rho = I_{13}^{(3)} b_3 + v_3$$

where $v_3 \sim N(0, d_{13} V_3)$ and $I_{13}^{(3)}$ is the third row of $I_{13}$. The variance parameter, $V_3$, reflects the strength of the belief in the sense that the smaller $V_3$ is, the more likely that parameter is close to $\rho$. In the application, I set $V_3 = 0.1$. To estimate the model with this information, the new $\tilde{Y}_3$ and $\tilde{X}_3$ are modified as

$$\tilde{Y}_3 = \left[ a_3' y_1 - \chi^m z_1 \ldots a_3' y_T - \chi^m z_T \right]' \frac{\rho}{\sqrt{V_3}}$$

$$\tilde{X}_3 = \left[ x_0 \ldots x_{T-1} \right]' P_3 \frac{I_{13}^{(3)}}{\sqrt{V_3}}$$

Tables 6, 7, 8 provide complete summaries for the priors of model $M_0$, $M_1$, and $M_2$ respectively.

### 3.3.4 Estimation results for the baseline model

Combining the prior and the likelihood, the baseline model $M_0$ is estimated by the modified Baumeister-Hamilton algorithm, which is summarized in the proposition 1 above. Figure 2 displays the prior and posterior distributions of contemporaneous parameters of the matrix $A$. For each panel, the red curves represent the prior, and the blue histograms represent the posterior. The prior and posterior distributions of $\beta^d$ and $\psi^y$ are significantly different, while they are quite similar for the rest of the parameters, suggesting a lack of identification. Thus, the data are informative about the effect of inflation on aggregate demand ($\beta^d$) and the effect of the output gap on the fed funds rates ($\psi^y$), but they are uninformative about other structural parameters. In particular, the data suggest that higher inflation will reduce aggregate demand and the Fed has a stronger reaction to a positive output gap than we initially believe. The estimation results are mostly similar to those in Baumeister and Hamilton (2018), however, they found a smaller effect of inflation rate on aggregate demand and a smoother policy reaction function. The discrepancy can be explained by my different sample period that runs from 1954Q3 to 2008Q4 and includes the Volcker era of aggressive monetary policy and high inflation, while Baumeister and Hamilton (2018) restrict their sample to the Great Moderation period.

---

12 In all of the applications, I simulate 2,000,000 draws by the Metropolis-Hasting algorithm and discard the first 1,000,000 draws in order to ensure convergence of the Markov Chain.
### 3.3.5 Bayesian investigation of instrument relevance

To formally investigate instrument relevance, I estimate model $M_1$ for each instrument and compute the Bayes factor with respect to $M_0$. Because the only difference between $M_0$ and $M_1$ is the use of the instrument in the latter model, the odds ratio is the indication for instrument relevance. Figures 3a, 3b, and 3c report the prior and posterior distributions of contemporaneous structural parameters together with the one that captures instrument relevance ($\chi^m$) for the Romer-Romer, Sims-Zha, and Smets-Wouters instruments, respectively. In all those figures, the posterior distributions of the parameter $\chi^m$ are far away from zero and more concentrated than the prior, confirming that all three instruments are highly correlated with the monetary policy shock in the baseline model $M_0$.

Having a valid and relevant instrument also provides more identifying power to the models, and in this case, the instruments provide more information about the effect of the interest rate on aggregate demand ($\gamma^d$) and the inertia of the fed funds rate ($\rho$). In particular, the posterior distributions of $\gamma^d$ are more concentrated near zero, suggesting a negative but small effect of the interest rate on aggregate demand. Also, the posterior distributions of $\rho$ are more concentrated around higher values, implying that the Fed does have a desire to implement gradual change in its interest rate policy. This implication is different from that of the baseline model because most of the interest rate volatility in the Volcker era is now captured by the instruments.

Table 9 reports the marginal likelihood of the baseline model $M_0$, the alternative model $M_1$, and the test statistics. For each instrument, I calculate the marginal likelihood using four different methods. In the first two methods, I use Geweke (1999)’s proposal, and in the remaining two, I use Sims et al. (2008)’s. Since both methods estimate the marginal likelihood from the truncated posterior distributions, I follow Herbst and Schorfheide (2015) and use two different levels of truncation, as represented by the value of the parameter $\tau$. A value for $\tau$ of 0.5 means that we use 50% of the posterior draws and a value for $\tau$ of 0.9 means that we use 90% of the posterior draws. Table 9 shows that the estimates of the marginal likelihood and test statistics are robust to different methods.

How high does the Bayes factor have to be before we decide that model $k$ is a better description of the data than model $k'$? Jeffreys (1967) emphasizes that the Bayes factor is a guide to decision making, thus its exact magnitude depends on the application. In some situations, decision makers might want to see a very large number of the Bayes factor before they reject the baseline model.\(^\dagger\) Table 10 reports a modified version of Jeffrey’s guideline as presented by Kass and Raftery (1995). It describes different values of the test statistics, their corresponding Bayes factors, and the suggested interpretation. For ease of computation and comparison with the frequentist likelihood ratios, the statistics are calculated as two times the log transformation of the Bayes factor as shown in the first column of the table.

\(^\dagger\)"... (The Bayes factor) is not a physical magnitude. Its function is to grade the decisiveness of the evidence. It makes little difference to the null hypothesis whether the odds are 10 to 1 or 100 to 1 against it, and in practice no difference at all whether they are $10^4$ or $10^{10}$ to 1 against it. In any case whatever alternative is most strongly supported will be set up as the hypothesis for use until further notice." (Jeffreys, 1967, Appendix B)
According to Jeffreys’ criteria from Table 10, there is very strong evidence that the instruments are relevant. For example, the test statistics for the Romer-Romer instrument are in the range of 44-45. It means that if a researcher’s prior belief is that there is 50% chance the instrument is relevance, after seeing the data, she will revise her odd of the instrument being relevance to more than three billion to one. The test statistics for Sims-Zha’s instrument and Smets-Wouters one have even higher values. These results are qualitatively similar to the frequentist evidence in Table 3, namely all the instruments are strong but the Romer-Romer one is weaker than the others.

3.3.6 Bayesian investigation of instrument validity

To investigate instrument validity, I first estimate model $M_2$, which is the model where both relevance and validity of the instrument are in doubt. Figures 4a, 4b, and 4c report the prior and posterior distributions for the contemporaneous structural parameters and those associated with the Romer-Romer, Sims-Zha, and Smets-Wouters instrument, respectively. Posterior distributions for $\chi^m$ are similar to those of $M_1$, while those for $\chi^s$ and $\chi^d$ are different from their priors, suggesting that the data are also informative about those parameters. Specifically, posterior distributions of $\chi^s$ tend to concentrated around zero, and those of $\chi^d$ appear to have more mass in the positive support. Thus, the analysis suggests that the instruments are likely to be invalid. Consequently, they no longer provide as much identification power as before. Posterior distributions for $\rho$ are still concentrated around the interval 0.4 to 0.8, however, these of $\gamma^d$ are no longer skewed toward zero as much as those seen in the model $M_1$.

The frequentist tests of overidentifying restrictions, previously reported in Table 4 and Table 5, also suggest that at least one instrument is invalid, but they can’t tell which instrument is valid and which one isn’t. The Bayesian approach, on the other hand, allows us to make that determination. Table 11 reports the marginal likelihood for each model. The last column of Table 11 shows the test statistics for different instruments. Here, the tests do have the power to differentiate instruments based on their validity. For instance, the test statistics for the Romer-Romer instrument is between 7 and 9 across all four methods, suggesting that there is strong evidence that the Romer-Romer instrument is not valid. If a researcher has a prior belief that there is 50% chance the instrument is invalid, after seeing the data, she has to revise the odds of the instrument being invalid to about 50 to 1. Similarly, the test statistics for the Sims-Zha instrument is between 4 and 7, and thus there is positive to strong evidence that the instrument is not valid. On the other hand, the test statistics for the Smets-Wouters instrument suggest that there is no significant evidence against the claim that the instrument is valid. Even if we use the most conservative number from table 11, the posterior odd against the claim of validity is only about 3 to 1. Therefore, the Bayesian approach suggests that the Romer-Romer and Sims-Zha instruments are invalid, whereas the Smets-Wouters instrument is valid.
3.4 Combining a relevant and valid instrument with sign restrictions

3.4.1 Replication of Romer and Romer (2004)’s specification

This section replicates the main regression in Romer and Romer (2004) to gain insight into the instruments’ information content. One would expect that if the instruments have relevant information about the response of the output gap in a simple regression, it will be useful for identification in the sign-restricted SVAR. In their study, Romer and Romer (2004) use the following regression with monthly data

$$\Delta y_t = \alpha_0 + \sum_{k=1}^{11} \alpha_k D_k + \sum_{i=1}^{24} \beta_i \Delta y_{t-i} + \sum_{j=1}^{36} c_j S_{t-j} + e_t \quad (40)$$

where $y$ is the log of industrial production, $D$ is monthly dummies, and $S$ is the Romer-Romer monetary policy shocks. In their specification, they assume that monetary policy instrument doesn’t affect industrial production within a month. As pointed out in Hamilton (2017), equation (40) is equivalent to a two-variable VARX(36) with restrictions on the lag parameters. Because I have quarterly data, I will use a VARX(12) with restrictions on the lag parameters as follow

$$\begin{bmatrix} \Delta y_t \\ S_t \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix} + \sum_{k=1}^{8} \begin{bmatrix} \phi_{11}^{(k)} \\ \phi_{12}^{(k)} \end{bmatrix} \begin{bmatrix} \Delta y_{t-k} \\ S_{t-k} \end{bmatrix} + \sum_{k=9}^{12} \begin{bmatrix} 0 \\ \phi_{12}^{(k)} \end{bmatrix} \begin{bmatrix} \Delta y_{t-k} \\ S_{t-k} \end{bmatrix} + \sum_{k=1}^{3} \begin{bmatrix} 1 \\ 0 \end{bmatrix} D_k + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (41)$$

where $y$ is the output gap, $D$ is the quarterly dummies, and $S$ is the monetary policy instrument, which is either that of Romer-Romer, Sims-Zha, or Smets-Wouters. For each instrument, I replace missing values with zero as done in Romer and Romer (2004) and estimate equation (41) with data from 1954Q3 to 2008Q4.

Figure 5 reports the SIRFs of output gap to one unit increase in the instruments together with their corresponding 95% confidence interval.\(^{14}\) The patterns of response for all three instruments are similar to that of Romer and Romer (2004), namely all SIRFs show a U-shape response of output gap to a contractionary monetary policy shock.\(^{15}\) Thus, all three instruments deliver evidence that contractionary monetary policy shocks reduce the output gap.

However, there are still some significant differences in the SIRFs in Figure 5. The SIRF estimated from the Smets-Wouters instrument suggests a faster response of output gap to interest rate shock than those estimated from the Romer-Romer and Sims-Zha instruments. In particular, the Smets-Wouters instrument implies a peak effect about one year after the shock, whereas those implied from Romer-Romer and Sims-Zha happen about two years after the shock. Moreover, although the magnitudes of those SIRFs are quite similar, their confidence intervals are not the same. Specifically, the SIRFs of Smets-Wouters and Sims-Zha instruments appear to be more precisely estimated than that of the Romer-Romer instrument.

\(^{14}\)Generally, one unit increase in the instrument doesn’t translate to one percentage point increase in the fed funds rate. Nakamura and Steinsson (2017) makes a similar observation.

\(^{15}\)Instruments of monetary policy shocks and fed funds rates are positively correlated. Thus, an increase in the measures will be equivalent to a contractionary monetary policy shock.
In summary, although the empirical results from different instruments are qualitatively similar, there are still some significant differences regarding their practical implications. In light of previous evidence about instrument validity, I will only consider the Smets-Wouters instrument in the following section.\textsuperscript{16}

3.4.2 Estimation results for the model with a relevant and valid instrument

In this section, I estimate model $M_1$ with the Smets-Wouters instrument from 1954Q3 to 2008Q4. The SIRFs are then compared with the baseline model $M_0$ where I don’t incorporate the instrument. I report the median SIRFs together with their 95 percent credible set. Baumeister and Hamilton (2018) shows that this is an optimal inference if one assumes an absolute-value loss function regarding the SIRFs. Nonetheless, my credible set is not directly comparable to the above results from Romer-Romer’s regression since their 95 percent confidence interval will tend to be larger than the 95 percent credible set in models that are only set-identified.

Since the main interest is in the effect of a monetary policy shock, I first focus on the results for the SIRFs of a contractionary monetary policy shock. Figure 3c shows the posteriors of the sign-restricted SVAR with the Smets-Wouters instrument. Besides $\chi^m$, $\gamma^d$ is also identified from the data since the posterior is highly concentrated around zero. To see the implication of this result, recall that the contemporaneous matrix is

$$A = \begin{bmatrix} 1 & -\alpha^s & 0 \\ 1 & -\beta^d & -\gamma^d \\ -(1-\rho)\psi^y & -(1-\rho)\psi^\pi & 1 \end{bmatrix}$$

which implies that the matrix for impact effect is

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} [\beta^d + \gamma^d (1-\rho)\psi^\pi] & \alpha^s & \alpha^s\gamma^d \\ \gamma^d (1-\rho)\psi^y - 1 & 1 & \gamma^d \\ -(1-\rho)\left(\psi^\pi + \beta^d\psi^y\right) & (1-\rho)(\psi^\pi + \alpha^s\psi^y) & \alpha^s - \beta^d \end{bmatrix}$$

where $|A| = \alpha^s \left[1 - \gamma^d (1-\rho)\psi^y\right] - [\beta^d + \gamma^d (1-\rho)\psi^\pi]$. When $\gamma^d = 0$, this matrix becomes

$$A^{-1} = \frac{1}{\alpha^s - \beta^d} \begin{bmatrix} -\beta^d & \alpha^s & 0 \\ -1 & 1 & 0 \\ -(1-\rho)\left(\psi^\pi + \beta^d\psi^y\right) & (1-\rho)(\psi^\pi + \alpha^s\psi^y) & \alpha^s - \beta^d \end{bmatrix}$$

In this case, a one unit increase of the monetary policy shock, $u^m$, increases the fed funds rate by one percentage point upon impact and have no contemporaneous effect on output and inflation. This implication

\textsuperscript{16}I also perform a similar analysis for the period of the Great Moderation. The U-shape pattern disappears. The evidence seems to depend mostly on the relationship between monetary policy and output before the Great Moderation period. The plots of the instruments suggest that most of their variations come from the Paul Volcker Era. These findings are similar to those of Coibion (2012), which found that the results of Romer and Romer (2004) are sensitive to the inclusion of the nonborrowed reserve targeting, 1979-1982. As Ramey (2016) points out, the consensus view is that monetary policy has been conducted more systematically in the recent period, thus pure monetary policy shocks are harder to identify and most of what we classify as "shocks" are just information effects from the Fed.
is similar to the "recursive assumptions" in the literature, which states that monetary policy shock doesn’t affect the output gap and inflation contemporaneously.

The above implication is verified from Figure 6, which shows the SIRFs of the output gap, inflation, and fed funds rate to a one unit increase of the monetary policy shock. This figure shows that the estimated SIRFs from the sign-restricted SVAR with external instrument tend to have a tight 95 percent credible sets, especially upon impact. Also, the impact effects on the output gap and inflation are closer to zero than the baseline model as suggested by the above derivation. The median SIRF shows that a contractionary monetary policy shock decreases the output gap for about two years after impact with the trough effect is around -0.3 percent. Then, the effect gets smaller and completely dies out about five years after. Thus, the sign-restricted SVAR with external instruments delivers a SIF that has a similar shape to that of the Romer and Romer (2004)’s regression in Figure 5. This dynamic is also found in previous empirical literature but is different from that of Baumeister and Hamilton (2018)’s specification, which suggests a large, negative effect upon impact and dies out gradually over the next three years.17

Figure 7 shows the full result of all the SIRFs. Besides SIRFs for the monetary policy shock, I also show the SIRFs for the demand and supply shock. However, the results suggest that the instrument is not useful for shrinking down the identified set of these two shocks. The most significant difference is in the SIRF of inflation in response to the demand shock, in which the augmented model suggests a stronger response of inflation relative to that of the baseline model. This result is not surprising since the instrument is designed for monetary policy shock, and the previous identification analysis suggests that this instrument only provides information about the SIRFs of the monetary policy shock.

In summary, the analysis shows that my method can incorporate information from a relevant and valid instrument to improve inference of the dynamic causal effects. In this example, I use the instrument to identify the SIRFs of the monetary policy shocks, and I use the sign restrictions to identify the SIRFs of the supply and demand shocks. My results for the SIRFs of the monetary policy shock are qualitatively similar to those from the frequentist procedure in Romer and Romer (2004).

4 Conclusion

Identifying dynamic causal effects by external instruments is increasingly popular in empirical macroeconomics. However, its key identifying assumption, instrument validity, is either untestable in a just-identified model or fails to hold in some special cases when it is formally tested. Conventionally, researchers often defend that assumption by relying on ad hoc arguments to say that their instruments are "plausibly exogenous", but without a formal method to analyze instrument validity, they are simply advocating to replace

17For example, Ramey (2016) studies the effect of monetary policy shock on industrial production using various specifications and found similar dynamics across all specification when the sample includes earlier observations in the 1970s. Her findings are shown in Figure 3.1, 3.2, and 3.3. She uses Christiano et al. (1999)’s model, Coibion (2012)’s model, Jordà et al. (2005)’s local projection method, proxy SVAR with Romer-Romer instrument, and proxy SVAR with high-frequency identification as in Gertler and Karadi (2015).
incredible identification assumptions with untestable ones. The problem arises because existing econometric
techniques cannot evaluate the validity assumption in a just-identified model nor can they say much about
the validity of any particular instrument in an over-identified model. This paper aims to fill that gap.

I propose to use identifying assumptions from a sign-restricted SVAR as a benchmark to evaluate the
underlying assumptions of the instrument. I first generalize the sign-restricted SVAR model developed by
Baumeister and Hamilton (2015) to incorporate instrumental variables. Then, I show that my framework
enables researchers to formally investigate the instrument relevance and validity assumptions by incorpo-
rating information from both theory and the empirical literature. In addition, my framework can combine
information from relevant and valid instruments to improve inference in sign-restricted SVAR models. This
latter contribution is of interest to practitioners who want to incorporate sign restrictions into a Proxy SVAR.

In the empirical application, I use the sign-restricted SVAR model of Baumeister and Hamilton (2018)
to investigate the relevance and validity assumptions of three popular instruments for monetary shocks,
proposed by Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007). My method
not only replicates results from standard frequentist procedures, namely that the instruments are strong
but all of them might not be valid, but it also differentiates valid from invalid instruments. In particular,
I conclude that the Smets-Wouters instrument is valid, while the other two are not. Moreover, when the
instrument is strong and valid, the framework delivers improved inferences for the SIRFs. I find that the
SIRFs of the output gap follow a U-shape pattern in response to a contractionary monetary policy shock,
and the effect reaches its maximum about one year after the shock then persists in the next four years.
This finding is largely in line with the previous literature using external instruments, suggesting that my
framework can incorporate information efficiently in practice.
References


Bartlett, M. S. (1957), ‘Comment on "a statistical paradox" by dv lindley’, Biometrika 44(1-2), 533–534.


Uhlig, H. (2005), ‘What are the effects of monetary policy on output? results from an agnostic identification procedure’, *Journal of Monetary Economics* 52(2), 381–419.

Watson, M. W. (2019), ‘Comment on "the empirical (ir)relevance of the zero lower bound constraint"’, *NBER Macroeconomics Annual* 34.


A Motivation for using conditional likelihood

A.1 Bayesian estimation

This Appendix shows that we can use conditional likelihood to estimate the model as long as the prior belief about the distribution of the exogenous variables are independent from the belief about the structural parameters. Consider the system of two equations:

\[ Ay_t = Bx_{t-1} + Cz_t + w_t \]

\[ z_t = v_t \]

where \( w_t \sim N(0,D) \) and \( v_t \sim N(0,V) \). \( D \) is a diagonal matrix and \( V \) is allowed to be non-diagonal. We can define \( y^*_t = (y'_t, z'_t)' \), \( A^* = \begin{bmatrix} A & -C \\ 0 & I_r \end{bmatrix}, D^* = \begin{bmatrix} D & 0 \\ 0 & V \end{bmatrix}, B^* = \begin{bmatrix} B \\ 0 \end{bmatrix}, x_{t-1} = (y'_{t-1}, y'_{t-2}, \ldots, y'_{t-m}, 1)' \). Then, the system can be rewritten as

\[ A^*y_t = B^*x_{t-1} + u^*_t \]

Because \( |A^*| = |A| \), the full likelihood function will be

\[ P(Y_T|A^*, B^*, D^*) = (2\pi)^{-T(n+r)/2} |A^*|^T \det |D^*|^{-T/2} \times \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (A^*y^*_t - B^*x_{t-1})' D^{*-1} (A^*y^*_t - B^*x_{t-1}) \right] \]

\[ = (2\pi)^{-Tn/2} |A|^T \det |D|^{-T/2} \times \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (Ay_t - Cz_t - Bx_{t-1})' D^{-1} (Ay_t - Cz_t - Bx_{t-1}) \right] \times (2\pi)^{-Tr/2} \det |V|^{-T/2} \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} z_t' V^{-1} z_t \right] \]

\[ = f(Y_T|A^*, B^*, D^*, Z_T) f(Z_T|V) \]

If the priors on \((A^*, B, D)\) are independent from those on \( V \), then full system Bayesian inference for the elements of \( A^*, B, D \) will be numerically identical to that based on the conditional likelihood

\[ P(Y_T|A^*, D^*, B^*, Z_T) = (2\pi)^{-Tn/2} |A|^T \det |D|^{-T/2} \times \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (Ay_t - Cz_t - Bx_{t-1})' D^{-1} (Ay_t - Cz_t - Bx_{t-1}) \right] \]

Thus, the expression separates into two independent problems, and the posterior for the elements of \( A^*, B, D \) can be simulated by the Baumeister-Hamilton algorithm as described in the main text.

A.2 Bayesian model comparison

Following the above motivation, this section shows why we can safely ignore the exogenous variable, \( z_t \), in the calculation of the marginal likelihood. Consider two model \( k \) and \( k' \), the posterior odd ratio can be written as
\[
\frac{P(M_k|Y_T, Z_T)}{P(M_{k'}|Y_T, Z_T)} = \frac{P(Y_T|Z_T, M_k) \cdot P(Z_T|M_k)}{P(Y_T|Z_T, M_{k'}) \cdot P(Z_T|M_{k'})} \cdot \frac{P(M_k)}{P(M_{k'})}
\]

where

\[
P(Z_T|M_k) = \int P(Z_T|\Psi, M_k) \cdot P(\Psi|M_k) \, d\Psi
\]

\[
P(Z_T|M_{k'}) = \int P(Z_T|\Psi, M_{k'}) \cdot P(\Psi|M_{k'}) \, d\Psi
\]

where \(\Psi\) governs the likelihood of \(Z_T\). Therefore, as long as \(P(Z_T|\Psi, M_k) = P(Z_T|\Psi, M_{k'})\) and \(P(\Psi|M_k) = P(\Psi|M_{k'})\) for every \(\Psi\) in the parameter space, the two marginal likelihood will be the same and the expression simplifies to

\[
\frac{P(M_k|Y_T, Z_T)}{P(M_{k'}|Y_T, Z_T)} = \frac{P(Y_T|Z_T, M_k) \cdot P(M_k)}{P(Y_T|Z_T, M_{k'}) \cdot P(M_{k'})}
\]

### B Marginal Likelihood Estimation

This Appendix describes the Geweke (1999) and Sims and Zha (2006) strategy to estimate the marginal likelihood. As described in the main text, the estimator for the marginal likelihood for model \(M_k\) is

\[
\hat{P}(Y_T|Z_T, M_k) = \left[ \frac{1}{N_0} \sum_{n=1}^{N_0} f(\theta^{n_0}) \right] \cdot \frac{1}{N_0} \sum_{n=1}^{N_0} P(Y_T|\theta^{n_0}, Z_T, M_k) \cdot P(\theta^{n_0}|M_k)
\]

The choice of \(f(\theta)\) is important in numerical calculation. Geweke (1999) and Sims and Zha (2006) propose different choices for \(f(\theta)\). The description below closely follows that of Herbst and Schorfheide (2015).

#### B.1 Geweke (1999)

Geweke (1999) proposes to use a truncated normal distribution as \(f(\theta)\). Particularly, denote \(\bar{\theta}\) and \(\bar{V}_\theta\) be the mean and covariance matrix of \(\theta\). Both of those quantities are numerically computed from the posterior distributions. Then, Geweke’s choice of \(f(\theta)\) is

\[
f_{\text{Geweke}}(\theta) = \tau^{-1} (2\pi)^{-d/2} |\bar{V}_\theta|^{-1/2} \exp \left[ -0.5 (\theta - \bar{\theta})' \bar{V}_\theta^{-1} (\theta - \bar{\theta}) \right] \chi^{-1}_{d}(\tau)
\]

which is simply a truncated normal. \(\chi^{-1}_{d}(\tau)\) denotes the inverse of the chi-squared distribution with \(d\) degrees of freedom, and the degree of truncation is controlled by \(\tau\). The lower value of \(\tau\), the more outliers will be removed from the posterior draws.
B.2 Sims, Waggoner, and Zha (2008)

One shortcoming of the Geweke’s proposal is that the posterior distribution might be very different from
the Gaussian distributions, which will lead to poor estimate of the marginal likelihood. Sims et al. (2008)
proposes a more sophisticated choice for \( f(\theta) \). Denote the mode of the posterior distribution as \( \hat{\theta} \), then we
can calculate an analog of the covariance matrix that is centered at the mode of the distribution

\[
\hat{V}_{\theta} = \frac{1}{N} \sum_{i=1}^{N} (\theta^i - \hat{\theta}) (\theta^i - \hat{\theta})'
\]

Next, define

\[
r(\theta) = \sqrt{\left(\theta - \hat{\theta}\right)' \hat{V}_{\theta}^{-1} \left(\theta - \hat{\theta}\right)}
\]

And let \( r^i = r(\theta^i) \), then we can construct \( f(\theta) \) in four steps

1. Construct a heavy-tailed univariate density \( g(r) \)

\[
g(r) = \begin{cases} \frac{vr^{v-1}}{b^v - a^v} & \text{if } r \in [a, b] \\ 0 & \text{otherwise} \end{cases}
\]

where \( v = \frac{\ln(0.1/0.9)}{\ln(c_{10}/c_{90})} \), \( a = c_1 \), and \( b = \frac{c_{90}}{0.9^{1/v}} \).

\( (c_1, c_{10}, c_{90}) \) are the first, 10th, and 90th percentiles of the empirical distribution of \( \{r^i\}_{i=1}^{N} \)

2. Define a new density \( \tilde{f}(r) \) as

\[
\tilde{f}(r) = \frac{\Gamma(d/2)}{2^{d/2} \pi^{d/2} |\hat{V}_{\theta}|^{1/2} r^{d-1}} g(r)
\]

where \( \Gamma(.) \) is the Gamma function, \( d \) is the dimension of the parameter vector \( \theta \).

3. Define a truncating function

\[
\mathbb{I}(\theta) = \mathbb{I}(\ln P(Y_T|\theta, Z_T) P(\theta) > L_{1-q}) \times \mathbb{I}(r(\theta) \in [a, b])
\]

Then, we can approximate the probability that function equal to 1 by simulation as

\[
\hat{\tau} = \hat{\mathbb{P}}(\mathbb{I}(\theta) = 1) = \frac{1}{J} \sum_{j=1}^{J} \mathbb{I}(\theta^j)
\]

where \( \theta^j \) is i.i.d and \( \theta^j \sim \tilde{f}(\theta) \).
4. Sims, Waggoner, and Zha’s choice of $f(\theta)$ is

$$f_{SWZ}(\theta) = \sqrt{V_{\theta}^{-1}} I(\theta)$$

The shortcoming of this method is that it is quite computationally expensive because we have to estimate $\hat{\tau}$ by simulation.

C Data Description

The data in this study are publicly available. Macroeconomic time series are downloaded from FRED, while monetary policy shocks are collected from Mark Watson and Yuriy Gorodnichenko’s website. Particularly, I use four quarterly time series from FRED.

2. GDPC1: Real Gross Domestic Product, Billions of Chained 2009 Dollars, Quarterly, Seasonally Adjusted Annual Rate
3. DPCERD3Q085SBEA (DPCE, for short): Personal consumption expenditures (implicit price deflator), Index 2009=100, Quarterly, Seasonally Adjusted
4. FEDFUNDS: Effective Federal Funds Rate, Percent, Quarterly Average, Not Seasonally Adjusted

The sample period is from 1954Q3 to 2008Q4 since data for Effective Federal Funds Rate starts later relative to the first three. Denote $y$= output gap, $\pi$= inflation, and $r$= fed funds rate. The variables in the baseline VAR will be calculated as follows

1. $y_t = 100 \times [\ln (GDPC1_t) - \ln (GDPPOT_t)]$
2. $\pi_t = 100 \times [\ln (DPCE_t) - \ln (DPCE_{t-4})]$
3. $r_t = FEDFUNDS_t$

Basically, the output gap is the difference between real and potential output, inflation is $Y/Y$ change in the PCE deflator, and the fed funds rate is just the raw data.

For the choice of monetary policy shocks, I start with the same set of instruments used in Stock and Watson (2012a). Those four shocks are

1. Romer-Romer shock: Described in Romer and Romer (2004) and downloaded from Yuri Gorodnichenko’s website. They are residuals from the regression between shocks constructed by the narrative methods on the Fed’s Greenbook forecasts of output and inflation. The orginal data from their
paper are from 1969Q1 to 1996Q4. In this study, I use the updated version constructed by and used in Coibion et al. (2017b). I use the quarterly data from their spreadsheet which are available from 1969Q1 to 2008Q4.

2. Smets-Wouters’ shocks: Described in Smets and Wouters (2007) and downloaded from Mark Watson’s website. They are interest rate shocks as calculated from Smets and Wouters’ DSGE model. The data from Stock and Watson (2012a) study are calculated by King and Watson (2012). The data are available from 1959Q1 to 2004Q4.

3. Sims-Zha’s shocks: Described in Sims and Zha (2006) and downloaded from Mark Watson’s website. The data are supplied by Tao Zha. They are shocks constructed from the VAR that includes Markov-Switching variances and no time-varying parameters. Following Stock and Watson (2012a), I convert them to quarterly data by taking the monthly average. The data are available from 1960Q1 to 2003Q1.
Table 1: **Summary statistics.** This table shows the summary statistics of the variables in the empirical application. Descriptions of the data and their availability are explained in the text. There are 218 quarterly observations for each variable in the period between 1954Q3 and 2008Q4. The units are all in percentage points.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviations</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output gap</td>
<td>-0.5</td>
<td>2.2</td>
<td>-7.6</td>
<td>5.5</td>
</tr>
<tr>
<td>Inflation rates</td>
<td>3.4</td>
<td>2.3</td>
<td>-0.2</td>
<td>10.9</td>
</tr>
<tr>
<td>Fed Funds rates</td>
<td>5.6</td>
<td>3.3</td>
<td>0.5</td>
<td>17.8</td>
</tr>
<tr>
<td>Romer-Romer monetary instrument</td>
<td>0.0</td>
<td>0.5</td>
<td>-4.1</td>
<td>2.5</td>
</tr>
<tr>
<td>Sims-Zha monetary instrument</td>
<td>0.0</td>
<td>2.4</td>
<td>-15.3</td>
<td>14.9</td>
</tr>
<tr>
<td>Smets-Wouters monetary instrument</td>
<td>-0.2</td>
<td>0.9</td>
<td>-3.6</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Table 2: **Cross-correlations:** This table shows the cross-correlation of the instruments for monetary policy shocks. Descriptions of the data and their availability are explained in the text.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Romer-Romer instrument</th>
<th>Sims-Zha instrument</th>
<th>Smets-Wouters instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romer-Romer instrument</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sims-Zha instrument</td>
<td>0.76</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Smets-Wouters instrument</td>
<td>0.37</td>
<td>0.44</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3: **First-stage statistics in SVAR-IV model.** This table shows the partial R-squared, F-statistics, Robust F-statistics, and their corresponding p-values in the first-stage regression of the SVAR-IV model. The p-values for both F-statistics and Robust F-statistics are both close to 0, thus I only show one column for both.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Partial R-squared</th>
<th>F-statistics</th>
<th>Robust F-statistics</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romer-Romer shocks</td>
<td>0.37</td>
<td>119</td>
<td>19</td>
<td>0.00</td>
</tr>
<tr>
<td>Sims-Zha shocks</td>
<td>0.57</td>
<td>262</td>
<td>34</td>
<td>0.00</td>
</tr>
<tr>
<td>Smets-Wouters shocks</td>
<td>0.50</td>
<td>203</td>
<td>52</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 4: Test of overidentifying restrictions in SVAR-IV model-all three instruments. This table shows the test statistics from three different tests of overidentifying restrictions: Sargan (1958), Basmann (1960), and Wooldridge (1995), together with their corresponding p-values.

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistics</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sargan’s score</td>
<td>10.81</td>
<td>0.0045</td>
</tr>
<tr>
<td>Basmann’s test</td>
<td>10.53</td>
<td>0.0052</td>
</tr>
<tr>
<td>Wooldridge’s score</td>
<td>5.62</td>
<td>0.0601</td>
</tr>
</tbody>
</table>
Table 5: Test of overidentifying restrictions in SVAR-IV model-two instruments. This table shows the test statistics from three different tests of overidentifying restrictions: Sargan (1958), Basmann (1960), and Wooldridge (1995), together with their corresponding p-values.

<table>
<thead>
<tr>
<th>Test</th>
<th>Romer-Romer/Sims-Zha</th>
<th>Romer-Romer/Smets-Wouters</th>
<th>Sims-Zha/Smets-Wouters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sargan’s score</td>
<td>0.11</td>
<td>0.74</td>
<td>7.17</td>
</tr>
<tr>
<td>Basmann’s test</td>
<td>0.11</td>
<td>0.75</td>
<td>6.90</td>
</tr>
<tr>
<td>Wooldrige’s score</td>
<td>0.14</td>
<td>0.71</td>
<td>4.10</td>
</tr>
</tbody>
</table>
Table 6: **Prior distributions of the baseline model** ($M_0$). This table shows the prior distribution of $A$, $D$, and $B$ together with their hyper-parameters. For Student $t$ distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^s$</td>
<td>Effect of $\pi$ on supply</td>
<td>2</td>
<td>0.4</td>
<td>–</td>
<td>$\alpha^s \geq 0$</td>
</tr>
<tr>
<td>$\beta^d$</td>
<td>Effect of $\pi$ on demand</td>
<td>0.75</td>
<td>0.4</td>
<td>–</td>
<td>None</td>
</tr>
<tr>
<td>$\gamma^d$</td>
<td>Effect of $r$ on demand</td>
<td>-1</td>
<td>0.4</td>
<td>–</td>
<td>$\gamma^d \leq 0$</td>
</tr>
<tr>
<td>$\psi^y$</td>
<td>Fed response to $y$</td>
<td>0.5</td>
<td>0.4</td>
<td>–</td>
<td>$\psi^y \geq 0$</td>
</tr>
<tr>
<td>$\psi^\pi$</td>
<td>Fed response to $\pi$</td>
<td>1.5</td>
<td>0.4</td>
<td>–</td>
<td>$\psi^\pi \geq 0$</td>
</tr>
</tbody>
</table>

| | **Priors for contemporaneous coefficients of $A$** |
| | Student $t$ distribution with 3 degrees of freedom |
| $\rho$ | Interest rate smoothing | 0.5 | 0.2 | – | $0 \leq \rho \leq 1$ |

| $d_{ii}^{-1}$ | Reciprocal of variance | $1/\left(\text{a}'(\text{Sa}_i)\right)$ | $1/\left(\sqrt{2\text{a}'(\text{Sa}_i)}\right)$ | – | $d_{ii} > 0$ |

| | **Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$** |

| $b_i$ | Lagged coefficients of equation $i$ | $\eta'a_i$ | $\sqrt{d_{ii}M_i}$ | – | None |
| In addition, | Third element of monetary equation | $\rho$ | $\sqrt{d_{33}/10}$ | – | None |
Table 7: Prior distributions of the model with a relevant and valid instrument ($M_1$). This table shows the prior distribution of $A$, $D$, $B$, and $C$ together with their hyper-parameters. For Student $t$ distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^s$</td>
<td>Effect of $\pi$ on supply</td>
<td>2</td>
<td>0.4</td>
<td>–</td>
<td>$\alpha^s \geq 0$</td>
</tr>
<tr>
<td>$\beta^d$</td>
<td>Effect of $\pi$ on demand</td>
<td>0.75</td>
<td>0.4</td>
<td>–</td>
<td>None</td>
</tr>
<tr>
<td>$\gamma^d$</td>
<td>Effect of $r$ on demand</td>
<td>-1</td>
<td>0.4</td>
<td>–</td>
<td>$\gamma^d \leq 0$</td>
</tr>
<tr>
<td>$\psi^y$</td>
<td>Fed response to $y$</td>
<td>0.5</td>
<td>0.4</td>
<td>–</td>
<td>$\psi^y \geq 0$</td>
</tr>
<tr>
<td>$\psi^\pi$</td>
<td>Fed response to $\pi$</td>
<td>1.5</td>
<td>0.4</td>
<td>–</td>
<td>$\psi^\pi \geq 0$</td>
</tr>
</tbody>
</table>

Student $t$ distribution with 3 degrees of freedom

Beta distribution with $\alpha = 2.6$ and $\beta = 2.6$

Priors for contemporaneous coefficients of $A$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>Interest rate smoothing</td>
<td>0.5</td>
<td>0.2</td>
<td>–</td>
<td>$0 \leq \rho \leq 1$</td>
</tr>
</tbody>
</table>

Student $t$ distribution with 3 degrees of freedom

Priors for coefficients of $C$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda^m$</td>
<td>Effect of $z$ on $u^m$</td>
<td>0</td>
<td>0.2</td>
<td>–</td>
<td>None</td>
</tr>
</tbody>
</table>

Student $t$ distribution with 3 degrees of freedom

Priors for structural variances $D|A, C$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{ii}^{-1}$</td>
<td>Reciprocal of variance</td>
<td>$1/\left(a_i'Sa_i\right)$</td>
<td>$1/\left(\sqrt{2a_i'Sa_i}\right)$</td>
<td>–</td>
<td>$d_{ii} &gt; 0$</td>
</tr>
</tbody>
</table>

Gamma distribution

Priors for lag coefficients $B|A, C, D$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_i$</td>
<td>Lagged coefficients of equation $i$</td>
<td>$\eta'a_i$</td>
<td>$\sqrt{d_{ii}M_i}$</td>
<td>–</td>
<td>None</td>
</tr>
<tr>
<td>$b_{33}$</td>
<td>Third element of monetary equation</td>
<td>$\rho$</td>
<td>$\sqrt{d_{33}/10}$</td>
<td>–</td>
<td>None</td>
</tr>
</tbody>
</table>

In addition,
Table 8: Prior distributions of the model with a relevant but invalid instrument (\(M_2\)). This table shows the prior distribution of A, D, B, and C together with their hyper-parameters. For Student \(t\) distributions, the location parameter refers to the mode. For Beta, Gamma, and Normal distributions, the location parameter is the mean and the scale parameter is the standard deviation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Location</th>
<th>Scale</th>
<th>Skew</th>
<th>Sign restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Priors for contemporaneous coefficients of A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\alpha^s)</td>
<td>Effect of (\pi) on supply</td>
<td>Student (t) distribution with 3 degrees of freedom</td>
<td>2</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>(\beta^d)</td>
<td>Effect of (\pi) on demand</td>
<td></td>
<td>0.75</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>(\gamma^d)</td>
<td>Effect of (r) on demand</td>
<td></td>
<td>-1</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>(\psi^y)</td>
<td>Fed response to (y)</td>
<td></td>
<td>0.5</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>(\psi^\pi)</td>
<td>Fed response to (\pi)</td>
<td></td>
<td>1.5</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td><strong>Beta distribution with (\alpha = 2.6) and (\beta = 2.6)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\rho)</td>
<td>Interest rate smoothing</td>
<td></td>
<td>0.5</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td><strong>Priors for coefficients of C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\chi^s)</td>
<td>Effect of (z) on (u^s)</td>
<td>Student (t) distribution with 3 degrees of freedom</td>
<td>0</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td>(\chi^d)</td>
<td>Effect of (z) on (u^d)</td>
<td></td>
<td>0</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td>(\chi^m)</td>
<td>Effect of (z) on (u^m)</td>
<td></td>
<td>0</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td>**Priors for structural variances (D_i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d_{ii}^{-1})</td>
<td>Reciprocal of variance</td>
<td>Gamma distribution</td>
<td>(1/(\hat{a}_i' \hat{S}_a))</td>
<td>(1/\sqrt{2\hat{a}_i' \hat{S}_a})</td>
<td>–</td>
</tr>
<tr>
<td><strong>Priors for lag coefficients (B_i)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b_i)</td>
<td>Lagged coefficients of equation (i)</td>
<td>Normal distribution</td>
<td>(\eta^i a_i)</td>
<td>(\sqrt{d_{ii} M_i})</td>
<td>–</td>
</tr>
<tr>
<td>In addition,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b_{33})</td>
<td>Third element of monetary equation</td>
<td></td>
<td>(\rho)</td>
<td>(\sqrt{d_{33}/10})</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 9: **Bayesian test for instrument relevance.** This table shows the marginal likelihood for the alternative and the baseline model, together with their difference. The baseline model, $M_0$, is the sign-restricted SVAR model estimated without using any external instrument, and the alternative model, $M_1$, is the sign-restricted SVAR model estimated with the instrument. Each instrument is entered one by one into the model. "Geweke" refers to the method proposed by Geweke (1999), and "SWZ" refers to the one proposed by Sims et al. (2008). The parameter $\tau$ specifies the level of truncation of the posterior distribution. A value for $\tau$ of 0.5 means that 50\% of the posterior draws are used, and a value for $\tau$ of 0.9 means that 90\% of the posterior draws are used.

<table>
<thead>
<tr>
<th>Test of instrument relevance-Bayesian approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{P}(Y_T</td>
</tr>
<tr>
<td><strong>Romer-Romer instrument</strong></td>
</tr>
<tr>
<td>Geweke (tau=0.5)</td>
</tr>
<tr>
<td>Geweke (tau=0.9)</td>
</tr>
<tr>
<td>SWZ (tau=0.5)</td>
</tr>
<tr>
<td>SWZ (tau=0.9)</td>
</tr>
<tr>
<td><strong>Sims-Zha instrument</strong></td>
</tr>
<tr>
<td>Geweke (tau=0.5)</td>
</tr>
<tr>
<td>Geweke (tau=0.9)</td>
</tr>
<tr>
<td>SWZ (tau=0.5)</td>
</tr>
<tr>
<td>SWZ (tau=0.9)</td>
</tr>
<tr>
<td><strong>Smets-Wouters instrument</strong></td>
</tr>
<tr>
<td>Geweke (tau=0.5)</td>
</tr>
<tr>
<td>Geweke (tau=0.9)</td>
</tr>
<tr>
<td>SWZ (tau=0.5)</td>
</tr>
<tr>
<td>SWZ (tau=0.9)</td>
</tr>
</tbody>
</table>
Table 10: Jeffrey’s criteria for model selection, originally proposed by Jeffrey’s (1967). I use the modified version in Kass and Raftery (1995). The Bayes factor, $B_{k,k'}$, is the ratio of the marginal likelihood of the model $M_k$ to the marginal likelihood of the model $M_{k'}$. When the prior belief is that the probabilities of the two models are equal, the Bayes factor is also their odds ratio.

<table>
<thead>
<tr>
<th>2log($B_{k,k'}$)</th>
<th>$B_{k,k'}$</th>
<th>Evidence against $M_{k'}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 2</td>
<td>1 to 3</td>
<td>Not worth more than a bare mentioning</td>
</tr>
<tr>
<td>2 to 6</td>
<td>3 to 20</td>
<td>Positive</td>
</tr>
<tr>
<td>6 to 10</td>
<td>20 to 150</td>
<td>Strong</td>
</tr>
<tr>
<td>$&gt;10$</td>
<td>$&gt;150$</td>
<td>Very strong</td>
</tr>
</tbody>
</table>
Table 11: **Bayesian test for instrument validity**: This table shows the marginal likelihood for the alternative and the baseline model, together with their difference. The baseline model, $M_1$, is the sign-restricted SVAR model estimated with the external instrument entered the third equation, and the alternative model, $M_2$, is the sign-restricted SVAR model estimated with the instrument entered all three equations. The test is performed for one instrument at a time. "Geweke" refers to the method proposed by Geweke (1999), and "SWZ" refers to the one proposed by Sims et al. (2008). The parameter $\tau$ specifies the level of truncation of the posterior distribution. A value for $\tau$ of 0.5 means that 50\% of the posterior draws are used, and a value for $\tau$ of 0.9 means that 90\% of the posterior draws are used.

| Test of instrument validity-Bayesian approach | $P(\mathbf{Y}_T|\mathbf{Z}_T,M_1)$ | $P(\mathbf{Y}_T|\mathbf{Z}_T,M_2)$ | $2 \times (P(\mathbf{Y}_T|\mathbf{Z}_T,M_2) - P(\mathbf{Y}_T|\mathbf{Z}_T,M_1))$ |
|---------------------------------------------|-----------------------------------|-----------------------------------|---------------------------------|
| **Romer-Romer instrument**                  |                                   |                                   |                                 |
| Geweke (tau=0.5)                            | -662                              | -659                              | 7                               |
| Geweke (tau=0.9)                            | -662                              | -658                              | 7                               |
| SWZ (tau=0.5)                               | -664                              | -660                              | 9                               |
| SWZ (tau=0.9)                               | -663                              | -659                              | 8                               |
| **Sims-Zha instrument**                     |                                   |                                   |                                 |
| Geweke (tau=0.5)                            | -650                              | -649                              | 4                               |
| Geweke (tau=0.9)                            | -650                              | -648                              | 4                               |
| SWZ (tau=0.5)                               | -653                              | -650                              | 6                               |
| SWZ (tau=0.9)                               | -653                              | -649                              | 7                               |
| **Smets-Wouters instrument**                |                                   |                                   |                                 |
| Geweke (tau=0.5)                            | -588                              | -588                              | 0                               |
| Geweke (tau=0.9)                            | -588                              | -588                              | 1                               |
| SWZ (tau=0.5)                               | -590                              | -590                              | 0                               |
| SWZ (tau=0.9)                               | -590                              | -589                              | 1                               |
Figure 1: Output gap, inflation, fed funds rates, together with Romer-Romer, Sims-Zha, and Smets-Wouters instruments. Output gap is the log difference between real and potential GDP, multiplied by 100. Inflation rate is the log difference between the Y/Y change of the PCE deflator, multiplied by 100. Fed funds rates is the Effective Federal Funds Rate. Romer and Romer’s instruments are the residuals from the regression between shocks constructed by the narrative methods on the Fed’s Greenbook forecasts of output and inflation. Sims and Zha’s instruments are shocks constructed from the VAR that includes Markov-Switching variances and no time-varying parameters. Smets and Wouters’ instruments are interest rate shocks as calculated from Smets and Wouters’ DSGE model. Shaded area indicates NBER recession periods. More detailed descriptions and data sources are in Appendix C.
Figure 2: Prior and posterior distributions of contemporaneous structural parameters of the baseline model \((M_0)\). This figure shows the prior and posterior distributions of contemporaneous structural parameters (i.e. elements of \(A\) matrix) of the baseline model \((M_0)\). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 3a: **Prior and posterior distributions of structural parameters of the alternative model ($M_1$) with the Romer-Romer instrument.** This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $A$, $C$ matrices) of the alternative model where the instrument is valid ($M_1$). The instrument is the monetary policy shocks originally constructed by Romer and Romer (2004) and updated by Coibion et al. (2017a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 3b: Prior and posterior distributions of structural parameters of the alternative model ($M_1$) with the Sims-Zha instrument. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $A, C$ matrices) of the alternative model where the instrument is not valid ($M_1$). The instruments are the monetary policy shocks estimated from Sims and Zha (2006)’s regime-switching SVAR and used in Stock and Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 3c: **Prior and posterior distributions of structural parameters of the alternative model ($M_1$) with the Smets-Wouters instrument.** This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $A, C$ matrices) of the alternative model where the instrument is not valid ($M_1$). The instruments are interest rate shocks estimated from the medium-scale DSGE model of Smets and Wouters (2007) and used in Stock and Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 4a: **Prior and posterior distributions of structural parameters of the alternative model \( (M_2) \) with the Romer-Romer instrument.** This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of \( A, C \) matrices) of the alternative model where the instrument is not valid \( (M_2) \). The instruments are monetary policy shocks originally constructed by Romer and Romer (2004) and updated by Coibion et al. (2017a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 4b: **Prior and posterior distributions of structural parameters of the alternative model (M2) with the Sims-Zha instrument.** This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of A, C matrices) of the alternative model where the instrument is not valid (M2). The instruments are the monetary policy shocks estimated from Sims and Zha (2006)’s regime-switching SVAR and used in Stock and Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 4c: Prior and posterior distributions of structural parameters of the alternative model ($M_2$) with the Smets-Wouters instrument. This figure shows the prior and posterior distributions of contemporaneous structural parameters and parameters associated with the instrumental variable (i.e. elements of $A, C$ matrices) of the alternative model where the instrument is not valid ($M_2$). The instruments are interest rate shocks estimated from the medium-scale DSGE model of Smets and Wouters (2007) and used in Stock and Watson (2012a). The prior distributions are the red curves, and the posterior distributions are the blue histograms. The sample period is 1954Q3 to 2008Q4. The posterior distributions are approximated by the Metropolis-Hasting algorithm with 2,000,000 draws and 1,000,000 burn-in samples.
Figure 5: Replication of Romer-Romer (2004)’s regression with data from 1954Q3 to 2008Q4. This figure shows the impulse response function (IRF) of the output gap to one unit increase of the monetary policy instrument together with their 95 percent confidence interval. The three different monetary policy instruments are those of Romer and Romer (2004), Sims and Zha (2006), and Smets and Wouters (2007).
Figure 6: **SIRFs of monetary shocks estimated with the Smets-Wouters instrument.** This figure shows the estimated SIRFs of the sign-restricted SVAR augmented with the Smets-Wouters instrument together with that of the baseline model. The sample period is 1954Q3 to 2008Q4. The solid blue line is the median SIRF of the baseline model, and the dashed red line is the median SIRF for the model augmented with the instrument. The shaded blue region is the 95 percent credible set for the baseline model, and the shaded red region is the 95 percent credible set for the model augmented with the instrument.
Figure 7: **SIRFs comparison between the sign-restricted SVAR augmented with Smets-Wouters instrument and the baseline model.** This figure shows the estimated SIRFs of the sign-restricted SVAR augmented with the Smets-Wouters instrument together with those of the baseline model. The sample period is 1954Q3 to 2008Q4. The solid blue line is the median SIRF of the baseline model, and the dashed red line is the median SIRF for the model augmented with the instrument. The shaded blue region is the 95 percent credible set for the baseline model, and the shaded red region is the 95 percent credible set for the model augmented with the instrument.