A Narrative Approach to a Fiscal DSGE Model

This version: March 20, 2015
First version: February 2014
Thorsten Drautzburg*

Abstract

Structural DSGE models are used both for analyzing policy and the sources of business cycles. Conclusions based on full structural models are, however, potentially affected by misspecification. A competing method is to use partially identified VARs using, for example, narrative shock measures. This paper proposes to use narrative shock measures to aid in the estimation of DSGE models. Here I adapt the existing methods for shock identification with external “narrative” instruments for Bayesian VARs using the standard SUR framework. I show theoretically that this narrative identification recovers the correct policy shocks in DSGE models with Taylor-type policy rules. In my application to fiscal and monetary policy in the US, I find that VAR-identified government spending shocks, tax shocks, and monetary policy shocks line up well with their DSGE model counterparts. Combining the DSGE model with the data, I also find support for the identifying assumption that instruments do not load on non-policy shocks. A standard DSGE with simple fiscal rules fails, however, to capture policy interactions.

Keywords: Fiscal policy; monetary policy; DSGE model; Bayesian estimation; narrative shocks; Bayesian VAR

* Federal Reserve Bank of Philadelphia. Email: tdrautzburg[at]gmail.com. Web: https://sites.google.com/site/tdrautzburg/. I am grateful to Jesus Fernandez-Villaverde, Pablo Guerrero-Quintana, Jim Nason, Barbara Rossi, Frank Schorfheide, Tom Stark, Keith Sill, Harald Uhlig, Mark Watson, Jonathan Wright and audiences at the EEA meetings, the ECB, Emory, Goethe University, the Philadelphia Fed, the SED, and Wharton for comments and suggestions. I would also like to thank Valerie Ramey as well as Karel Mertens and Morten Ravn for making their code and data publicly available. All errors are mine. The views expressed herein are my own views only. They do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or its Board of Governors.
1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are both a popular research and a policy tool. They allow for competitive forecasts and a structural analysis of economic driving forces and policies. Any structural analysis is, however, potentially affected by misspecification of the structural model. This paper formally connects identification schemes relying on non-structural methods to a class of DSGE models and asks whether the different identification methods give consistent answers.

The popularity of quantitative, medium-scale DSGE models such as Christiano et al. (2005), Smets and Wouters (2007), and others is linked to their quantitative success: They can match VAR responses (Christiano et al., 2005) and even beat a VAR in terms of forecasting performance (Smets and Wouters, 2007).\(^1\) Besides forecasting, these models also provide an analysis of the US economy in terms of structural shocks. Other work has exploited the ability of these models to analyze different policy rules, see Faust (2009, p. 52f) for references. The ability of the model to provide such a structural analysis is a by-product of the complete parametrization of the model. Faust (2009) cautions, however, that these models miss out on phenomena typically believed to be important features of reality, such as the long and variable lags of monetary policy. He calls for assessing “where the [DSGE] models reflect and where they contradict common understanding” (p. 63). A comment by Sims (2005, p. 2) similarly cautions that economists may be set back “[i]f Bayesian DSGE’s displace methods that try to get by with weak identification”.\(^2\)

This paper tries to formally bridge the gap between non-structural identification methods and the identification in DSGE models. To that end, I use the recently proposed identification of VARs through external instruments (Stock and Watson, 2012; Mertens and Ravn, 2013). I show that for simple Taylor-type policy rules, the

---

\(^1\)Del Negro et al. (2013) argue that with small modifications, i.e. adding a parsimoniously modeled financial sector and observed inflation expectations, an otherwise standard medium-scale DSGE model also fits well the US experience since the Great Recession. However, an adequate structure is a required maintained assumption in this literature.

\(^2\)There is also implicit or explicit discomfort with structural identification using DSGE models in the wider literature. Mian and Sufi (2012) only use micro data to identify demand shocks and then resort to partial model-specification to assess the macro implications, avoiding the specification of a full DSGE model. Ferreira (2013), in his analysis of risk shocks, also shies away of using an estimated DSGE model to back out shocks and instead uses only model-derived sign restrictions for fear of the DSGE model misspecification. Drautzburg (2013) identifies shocks from cross-sectional data relying only on part of the structural model and then uses them as an input into a dynamic structural model.
identification method proposed in Mertens and Ravn (2013) correctly identifies the
effect of shocks on policy instruments and thereby allows to recover these Taylor
rules in the instrument-identified VAR. In the spirit of the criticism in Faust (2009),
I use established narrative methods of identifying policy shocks as the basis for the
structural identification of the VAR. In ongoing work, I summarize the assessment
of consistency in statistical terms using Bayesian model comparison in an extension
DSGE-VAR in the SUR case complicates the analysis and I use methods proposed in
Chib and Ramamurthy (2010) to sample from the posterior and compute marginal
likelihoods.

In my application, I focus on policy rules for government spending, taxes, and
interest rates using instruments from Ramey (2011), Mertens and Ravn (2013), and
Romer and Romer (2004), respectively. I compare the VAR to an extension of
the standard medium-scale DSGE model from Christiano et al. (2005) and Smets
and Wouters (2007), with linear fiscal policy rules as in Leeper et al. (2010) and
Fernandez-Villaverde et al. (2011) and distortionary taxes as in Drautzburg and
Uhlig (2011). Preliminary results indicate that the implied shock histories of the
VAR and the DSGE model roughly agree for all three shocks, with median R-squares
around 50%. Also the impulse-responses to a tax shock are broadly consistent. How-
ever, the VAR evidence points to fiscal-monetary policy interactions not predicted
by the DSGE model. If robust, this would suggest a role for productive government
spending as in Drautzburg and Uhlig (2011) or consumption complementarities as
in Coenen et al. (2012).3

The application to a fiscal DSGE model with monetary policy is important from
a substantive point of view: With monetary policy constrained by the Zero Lower
Bound, “stimulating” fiscal policy has gained a lot of attention and influential papers
such as Christiano et al. (2009) have used quantitative DSGE models for the analysis
of fiscal policies. Since the fiscal building blocks of DSGE models are less well studied
than, say, the Taylor rule for monetary policy (e.g. Clarida et al., 2000), assessing
the fiscal policy implications of these models is warranted. Indeed, fully structural
and partial identification of fiscal shocks can lead to widely different conclusions

---

3A different interpretation of the differences in the responses and policy interaction is that
narrative shocks reflect shocks other than the ones assumed in the model, violating the identifying
assumption underlying this paper. The DSGE model allows to test whether a given narrative shock
indeed loads up on more than one structural shock. My estimates provide support for the correct
structural interpretation of the narrative shock measures.
about the drivers of business cycles: Rossi and Zubairy (2011) document that when applying a Blanchard and Perotti (2002)-type identification of government spending shocks in a VAR, the fraction of the variance of GDP driven by those shocks rises significantly with the forecast horizon, whereas the DSGE-model based variance decomposition in Smets and Wouters (2007) implies the opposite pattern.\footnote{Table 2 in Rossi and Zubairy (2011) implies that the fraction of the variance of GDP driven by government spending shocks rises with the forecast horizon from below 5\% at four quarters to 35\% at 40 quarters whereas (Smets and Wouters, 2007, Figure 1) implies that the output variance explained by GDP falls from roughly 30\% at one quarter to about 15\% at four quarters and less than 5\% at 40 quarters. Note that throughout this paper, I focus on discretionary fiscal policy, as opposed to the effect of policy rules.}

Methodologically, this paper complements the recent advances VAR-based shock identification using narrative measures as instruments (Stock and Watson, 2012; Mertens and Ravn, 2013) in the frequentist framework by proposing a simple Bayesian analogue based on the textbook SUR model. When only few observations are missing, I show that the proposed estimator with an improper prior has (Bayesian) confidence intervals with an actual size close to the nominal size, unlike a frequentist scheme used in the literature. With little variation in the instrument, however, the proposed procedure with a flat prior is shown to be conservative. This motivates shrinking the model towards some prior. To elicit a prior, I adapt the DSGE-VAR framework in Del Negro and Schorfheide (2004) for the case of the narrative VAR. Estimation of the full DSGE-VAR is in progress, but estimates based on a degenerate (point-mass) prior are discussed.

Having instruments available to identify structural shocks narrows down the well-known VAR identification problem of finding a rotation in an \(m\)-dimensional space, where \(m\) is the number of variables in the VAR, to finding a rotation in an \(m_z < m\) dimensional subspace. I consider two methods to obtain unique rotations, up to sign restrictions. One method is a purely statistical method based on Uhlig (2003) that identifies a shock by choosing it to be the main contributor to the forecast error variance of a certain variable over a given horizon. The other method is the rotation used in Mertens and Ravn (2013).

I show that the rotation proposed by Mertens and Ravn (2013) to solve the remaining identification problem actually recovers the structural impulse-response matrix in the DSGE model with simple Taylor-rules by limiting the contemporaneous interaction between policy instruments. This result is interesting because such rules have been popular in the literature (Leeper et al., 2010; Fernandez-Villaverde
et al., 2011, e.g.). It may also be of interest when estimating DSGE models under partial information by IRF-matching as in Christiano et al. (2005) or Altig et al. (2005). Since no stark zero-restrictions are needed, the proposed identification scheme may provide a useful set of moments which can be more naturally matched by DSGE models.

This paper is structured as follows: Section 2 frames the research question in general and formal terms. The paper proceeds by describing the methods used in the analysis in Section 3. Sections 4 and 5 describe the empirical specification and the empirical results.

2 Framework

To fix ideas, consider the following canonical representation of a state-space representation of a linear DSGE model with the vector $Y$ of observables and vector $X$ of potentially unobserved state variables:

$$Y_t = B^* X_{t-1}^* + A^* \epsilon_t^* \quad (2.1a)$$

$$X_t^* = D^* X_{t-1}^* + C^* \epsilon_t^* \quad (2.1b)$$

where $E[\epsilon_t^* (\epsilon_t^*)'] = I_m$. $m$ is the dimensionality of $Y_t$.

In this paper, I am interested in estimating the following VAR(p) approximation to this state-space model:

$$Y_t = BX_t - 1 + A \epsilon_t \quad (2.2a)$$

$$X_t = \begin{bmatrix} Y_t & Y_{t-1} & \ldots & Y_{t-(p-1)} \end{bmatrix} \quad (2.2b)$$

where the dimension of the state vector is typically different across the VAR and the DSGE model, but the shocks $\epsilon_t$ and $\epsilon_t^*$ are assumed to be of the same dimension.

The typical challenge in VARs is that only $AA'$ is identified from observables, but not $A$. Stock and Watson (2012) and Mertens and Ravn (2013) have shown that narrative or external instruments can be used to identify or partially identify $A$.

Assuming that the VAR(p) is a good approximation to dynamics,\footnote{This assumption is not necessarily satisfied. It can, however, be verified, using the condition in Fernandez-Villaverde et al. (2007), that the popular Smets and Wouters (2007) model with AR(1) shocks does have a VAR(∞) representation. When working with a specific DSGE model, I also} so that $AA' \approx$
A^*(A^*)', this paper asks first whether instruments can theoretically identify the correct \( A = A' \) in a class of DSGE models. Second, this paper uses narrative instruments to partially identify \( A \) and infers fiscal and monetary policy rules in a class of DSGE models based on the partially identified \( A \).

## 3 BVAR estimation with narrative instruments

In this section, I first discuss identification and estimation in the instrument-identified VAR. The second part of this section links identification in the VAR to identification in a DSGE model with Taylor-type policy rules and outlines how to use a DSGE model for prior elicitation and for testing the identifying assumption.

### 3.1 Narrative BVAR

I use the following notation for the statistical model:

\[
y_t = \mu_y + By_{t-1} + v_t \quad \text{(3.1a)}
\]

\[
v_t = A\epsilon_t^{str}, \epsilon_t^{str} \iid \mathcal{N}(0, I_m) \quad \text{(3.1b)}
\]

\[
z_t = \mu_z + Fv_t + \Omega^{-1/2}u_t, u_t \iid \mathcal{N}(0, I_k) \quad \text{(3.1c)}
\]

Here, \( Y_t \) is the observed data, \( B \) is a matrix containing the (possible stacked) lag coefficient matrices of the equivalent VAR(p) model as well as constants and trend terms, \( v_t \) is the \( m \)-dimensional vector of forecast errors, and \( z_t \) contains \( k \) narrative shock measures.

Note that, knowing \( B \), \( v_t \) is data. We can thus also observe \( \text{Var}[v_t] = AA' \equiv \Sigma \). Note that \( A \) is identified only up to an orthonormal rotation: \( \tilde{A}Q(\tilde{A}Q)' = AA' \) for \( \tilde{A} = \text{chol}(\Sigma) \) and \( QQ' = I \).

The observation equation for the narrative shocks (3.1c) can alternatively be written as:

\[
z_t = \begin{bmatrix} G & 0 \end{bmatrix} \epsilon_t^{str} + \Omega^{-1/2}u_t = \begin{bmatrix} G & 0 \end{bmatrix} A^{-1} A\epsilon_t^{str} + \Omega^{-1/2}u_t \quad \text{(3.2)}
\]

By imposing zero restrictions on the structural representation of the covariance matrix, knowledge of \( F \) and \( AA' \) identifies the shocks which are not included in verify this condition in my application.
3.2. However, only the covariance matrix $FAA'$ is needed for identification. It is therefore convenient for inference and identification to introduce a shorthand for the covariance matrix between the instruments and the forecast errors. Formally:

**Assumption 1.** For some invertible square matrix $G$, the covariance matrix $\Gamma$ can be written as:

$$\Gamma \equiv \text{Cov}[z_t, v_t] = FAA' = \begin{bmatrix} G & 0 \end{bmatrix} A'.$$  
(3.3)

The model in (3.1) can then be written compactly as:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} | Y_{t-1} \sim \mathcal{N} \left( \begin{bmatrix} \mu_y + By_{t-1} \\ \mu_z \end{bmatrix}, \begin{bmatrix} AA' & \Gamma' \\ \Gamma & \tilde{\Omega} \end{bmatrix} \right),$$
(3.4)

where $\tilde{\Omega} = \Omega + FAA'F'$ is the covariance matrix of the narrative instruments.

### 3.1.1 Identification given parameters

This section largely follows Mertens and Ravn (2013). It considers the case of as many instruments as shocks to be identified, with $k \leq m$.

Partition $A = [\alpha^{[1]}, \alpha^{[2]}] = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$, $\alpha^{[l]} = [\alpha_{11}' , \alpha_{21}']'$ with both $\alpha_{11}(m_z \times m_z)$ being invertible and $\alpha_{21}((m - m_z) \times m_z)$.

Using the definitions of $\Gamma$ and the forecast errors gives from (3.3):

$$\Gamma \equiv \text{Cov}[z_t, v_t] = \text{Cov}[z_t, A\epsilon_t^{str}] = \begin{bmatrix} G & 0 \end{bmatrix} A' = Go_1' = [Go_{11}', Go_{21}']$$
(3.5)

If the narrative shocks together identify an equal number of structural shocks, $G$ is of full rank. Since also $\alpha_{11}$ is invertible:

$$Go_{11}' = \Gamma_1,$$  
(3.6a)

$$\alpha_{21}' = G^{-1}\Gamma_2 = \alpha_{11}'(\Gamma_1^{-1}\Gamma_2), \equiv \alpha'_{11}\kappa'$$  
(3.6b)

where $\Gamma$ is a known reduced form parameter matrix that can be estimated.
Hence the (structural) impulse-vector to shocks 1, \ldots, m_z satisfies:

\[
\alpha^{[1]} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} I_{m_z} \\ (\Gamma_1^{-1} \Gamma_2)’ \end{bmatrix} \alpha_{11},
\]

where \( \Gamma_2 \) is \( m_z \times (m - m_z) \) and \( \Gamma_1 \) is \( m_z \times m_z \). This \( m \times m_z \) dimensional vector is a known function of the \( m_z^2 \) parameters in \( \alpha_{11} \). It therefore restricts \((m - m_z)m_z\) elements of \( A \).

An alternative way of stating the under-identification of VARs without external instruments or other restrictions on the coefficients is to note that \( A \) is only identified up to multiplication by an orthonormal rotation matrix \( Q' \). Since \( A \) is assumed to be of full rank and, hence, \( \Sigma \) is positive definite, the Choleski factorization of \( \Sigma \) exists: \( \tilde{A} \tilde{A}' = \Sigma \). Since \( A \) and \( \tilde{A} \) are of full rank, there is a unique \( Q \) such that \( A = \tilde{A}Q \). Here, \( Q \) and therefore \( A \) is typically only partially identified. Order the identified columns first, as above. Then:

\[
\begin{bmatrix} \alpha^{[1]}, \alpha^{[2]} \end{bmatrix} = \tilde{A}Q = [\tilde{A}q^{[1]}, \tilde{A}q^{[2]}]
\]

Thus:

\[
\begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11} = \alpha^{[1]} = \tilde{A}q^{[1]} \iff q^{[1]} = \tilde{A}^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11}
\]

\( \alpha_{11} \) is partially pinned down from the requirement that \( Q \) be a rotation matrix.

In particular:

\[
I_{m_z} = (q^{[1]})'q^{[1]} = \alpha_{11}' \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} (\tilde{A}^{-1})' \tilde{A}^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11} = \alpha_{11}' \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \alpha_{11}
\]

\[
(\alpha_{11}')^{-1}(\alpha_{11})^{-1} = \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \Sigma^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \iff \alpha_{11} \alpha_{11}' = \left( \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} I_{m_z} \\ \kappa \end{bmatrix} \right)^{-1}
\]

This requires \( \frac{m_z(m_z - 1)}{2} \) additional restrictions to identify \( \alpha_{11} \) completely.

The following Lemma summarizes the above results:

**Lemma 1.** (Stock and Watson, 2012; Mertens and Ravn, 2013) Under Assump-
tion 1, the impact of shocks with narrative instruments is generally identified up to a \(m_z \times m_z\) scale matrix \(\alpha_{11}\) whose outer product \(\alpha_{11}' \alpha_{11}\) is known, requiring an extra \(\frac{(m_z-1)m_z}{2}\) identifying restrictions and the impulse vector is given by (3.7). Proof: See Appendix A.1.

While the Lemma reduces the dimensionality of the traditional VAR-identification problem with the help of instruments, for \(m_z > 1\) infinitely many rotations are still possible. Here I consider two different statistical approaches for disentangling these shocks. First, following Mertens and Ravn (2013), I consider a Choleski-decomposition in a 2-Stage Least-Squares (2SLS) representation of the above problem. This procedure, as is shown in Section 3.2, recovers the true \(\alpha^{[1]}\) for a certain class of DSGE models. However, to asses the robustness of my empirical findings beyond this class of models I also consider a second scheme: Maximizing the (conditional) forecast error variance, building on Uhlig (2003).

First, consider the Mertens and Ravn (2013) factorization. For given \(\Sigma, \Gamma\), this can be interpreted as population 2SLS. In this 2SLS, the instruments \(z_t\) serve to purge the forecast error variance to the first \(m_z\) variables in \(y_t\) from shocks other than \(\epsilon_t^{[1]}\). Mertens and Ravn (2013) call the resulting residual variance-covariance matrix \(S_1 S_1'\) and propose either an upper or a lower Choleski-decomposition of \(S_1 S_1'\). To see this result mathematically, note that \(A^{-1}v_t\) can be re-written as the following system of simultaneous equations:

\[
\begin{bmatrix}
v_{1,t} \\
v_{2,t}
\end{bmatrix} = \begin{bmatrix} 0 & \eta \\ \kappa & 0 \end{bmatrix} \begin{bmatrix} v_{1,t} \\
v_{2,t}
\end{bmatrix} + \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\
\epsilon_{2,t}
\end{bmatrix},
\]

where \(\eta = \alpha_{12} \alpha_{22}^{-1}\) and \(\kappa = \alpha_{21} \alpha_{11}^{-1}\) are functions of \(\Sigma, \Gamma\), given in Appendix A.2. Using the definition of \(v_t = A \epsilon_t\) and simple substitution allows to re-write this system as:

\[
\begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\
\epsilon_{2,t}
\end{bmatrix} = \begin{bmatrix} v_{1,t} \\
v_{2,t}
\end{bmatrix} = \begin{bmatrix} (I - \eta \kappa)^{-1} \\
(I - \kappa \eta)^{-1}\kappa \end{bmatrix} S_1 \epsilon_{1,t} + \begin{bmatrix} (I - \eta \kappa)^{-1} \eta \\
(I - \kappa \eta)^{-1} \end{bmatrix} S_2 \epsilon_{2,t},
\]

This shows how identifying \(S_1\) identifies \(\alpha^{[1]}\) up to a Choleski factorization as:

\[
\begin{bmatrix} \alpha_{11} \\
\alpha_{21}
\end{bmatrix} = \begin{bmatrix} (I - \eta \kappa)^{-1} \\
(I - \kappa \eta)^{-1}\kappa \end{bmatrix} \text{chol}(S_1 S_1').
\]
The second factorization I consider is a generalization of the Choleski decomposition of $\alpha_{11} \alpha_{11}'$, the contemporaneous forecast error variance attributable to the instrument-identified shocks $\epsilon_t^{[1]}$. The (lower) Choleski decomposition of $\alpha_{11} \alpha_{11}'$ implies that all of the one period ahead conditional forecast error is attributable to the first instrument-identified shock. All of the remaining residual variation in the second shock is then attributed to the second instrument-identified shock and so forth.\(^6\)

The procedure described in Uhlig (2003) and used by Barsky and Sims (2009) generalizes this by identifying the first instrument-identified shock as the shock which explains the most conditional forecast error variance in variable $i$ over some horizon $\{h, \ldots, H\}$. Following the steps in Uhlig (2003), this amounts to solving the following principal components problem:\(^7\)

\[
\max_{\lambda_\ell} S \tilde{q}_\ell^\alpha = \lambda_\ell \tilde{q}_\ell, \quad \ell \in \{1, \ldots, m_z\},
\]

\[
S = \sum_{h=0}^{\bar{h}} (\bar{h} + 1 - \max\{h, \bar{h}\}) \left( B^h \left[ \begin{array}{c} I_{m_z} \\ \kappa \end{array} \right] \tilde{\alpha}_{11} \right)' e_1 e_1' \left( B^{\bar{h}} \left[ \begin{array}{c} I_{m_z} \\ \kappa \end{array} \right] \tilde{\alpha}_{11} \right),
\]

where $\tilde{\alpha}_{11}$ is the Choleski factorization associated with (3.9) and $e_1$ is a selection vector with zeros in all except the first position. Then the desired $\alpha_{11}$ is given by:

\[
\alpha_{11} = \tilde{\alpha}_{11} \tilde{q}_\ell^\alpha,
\]

where the eigenvectors $\tilde{q}_\ell^\alpha$ can be normalized to form an orthogonal matrix because $S$ is symmetric.\(^8\)

With either factorization scheme, identified structural shocks can easily obtained from the data. Recall that $Q = [q^{[1]}_1, q^{[1]}_1 \perp]$ with $q^{[1]}_1 \perp$ such that $(q^{[1]}_1)' q^{[1]}_1 = 0$ and

---

\(^6\)Note that this does not imply that no shock besides the first shock in $\epsilon_t$ has an impact effect on the variable ordered first – shocks in $\epsilon_t^{[2]}$ generally still have effects.

\(^7\)Here I used that the variance whose forecast error variance is of interest is among the first $m_z$ variables. Otherwise, redefine $\alpha_{11}$ to include $\kappa$ or reorder variables as long as $\alpha_{11}$ is guaranteed to be invertible.

\(^8\)Note that this latter scheme is only identified up to sign: I normalize signs such that $\text{diag}(\alpha_{11})$ is positive.
Then use that $\epsilon_1 = A^{-1}u_t = (\tilde{A}Q)^{-1}u_t$ and that $Q^{-1} = Q'$ to get:

$$
\epsilon_t = Q' \tilde{A}^{-1}u_t = \begin{bmatrix} (q_1^{[1]})' \\ (q_\perp^{[1]})' \end{bmatrix}' \tilde{A}^{-1}u_t = \begin{bmatrix} \alpha'_{11} & \alpha'_{1m} & \kappa' \\ \kappa' & (q_\perp^{[1]})' \end{bmatrix} (\tilde{A}'^{-1}u_t)
$$

$$
\Rightarrow \epsilon_t^{[1]} = \alpha'_{11} \left[ I_{m_x} \kappa' \right] \Sigma^{-1}v_t
$$

The next section considers how inference is affected when the covariance matrix underlying the identification scheme has to be estimated.

### 3.1.2 Posterior uncertainty

Here, I consider the case when the posterior over $\Gamma$ is non-degenerate. I abstract from potential weak instruments, which are discussed in, for example, Kleibergen and Zivot (2003) and surveyed in Lopes and Polson (2014).

Inference is analogous to inference in a SUR model (e.g. Rossi et al., 2005, ch. 3.5). In the special case in which the control variables for $Z_t$ coincide with the variables used in the VAR, the SUR model collapses to a standard scheme hierarchical Normal-Wishart posterior. Stack the vectorized model (3.4) as follows:

$$
Y_{SUR} = X_{SUR}\beta_{SUR} + v_{SUR}, \quad v_{SUR} \sim \mathcal{N}(0, V \otimes I_T), \quad (3.13)
$$

using the following definitions:

\[
V = \begin{bmatrix} AA' & \Gamma' \\ \Gamma & \tilde{\Omega} \end{bmatrix}, \quad \beta_{SUR} = \begin{bmatrix} \text{vec}(B) \\ \text{vec}(\mu_z) \end{bmatrix}, \quad Y_{SUR} = \begin{bmatrix} y_{1,1}, \ldots, y_{1,T}, \ldots, y_{m_{x,1}}, \ldots, y_{m_{x,T}}, z_{1,1}, \ldots, z_{1,T}, \ldots, z_{m_{\Delta z}}, \ldots, z_{m_{\Delta z},T} \end{bmatrix}'
\]

\[\text{Note that, by construction, this gives orthogonal historical shocks: } \text{Var}[\epsilon_1^{[1]}] = \alpha'_{11} \left[ I_{m_x} \kappa' \right] \Sigma^{-1} \text{Var}[u_t] \Sigma^{-1} \left[ I_{m_x} \kappa' \right], \text{Var}[u_t] = \alpha'_{11} \left[ I_{m_x} \kappa' \right] \Sigma^{-1} \left[ I_{m_x} \kappa' \right] \alpha'_{11} = I_{m_x} \text{ from (3.8). The responses to other shocks are partially identified, up to } (m - m_x) \times (m - m_x) \text{ matrix } \lambda:
\]

\[
q_\perp^{[1]} = \tilde{A}' \begin{bmatrix} \kappa' \\ -I_{m-m_x} \end{bmatrix} \lambda, \quad \lambda \lambda' = \begin{bmatrix} \kappa' \\ -I_{m-m_x} \end{bmatrix} \Sigma \begin{bmatrix} \kappa' \\ -I_{m-m_x} \end{bmatrix}^{-1},
\]

where the restrictions on $\lambda$ follow from the orthogonality condition $(q_\perp^{[1]})'q_\perp^{[1]} = I_{m-m_x}$. 

9
\[ X_{\text{SUR}} = \begin{bmatrix} I_{m_y} \otimes X_y & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes X_z \end{bmatrix} \]

\[ X_y = \begin{bmatrix} Y_{-1} & \ldots & Y_{-p} & 1_T \end{bmatrix} \]

\[ X_z = \begin{bmatrix} 1_T \end{bmatrix} \]

\[ v_{\text{SUR}} = \left[ [v^y_{1,1}, \ldots, v^y_{1,T}], \ldots, [v^y_{m_c,1}, \ldots, v^y_{m_c,T}], [v^z_{1,1}, \ldots, v^z_{1,T}], \ldots, [v^z_{m_z,1}, \ldots, v^z_{m_z,T}] \right]' \]

The model is transformed to make the transformed errors independently normally distributed, taking advantage of the block-diagonal structure of the covariance matrix: \( \bar{v} = \bar{Y} - \bar{X} \beta \sim \mathcal{N}(0, I) \). Standard conditional Normal-Wishart posterior distributions arise from the transformed model. For the transformation it is convenient to define \( U \) as the Choleski decomposition of \( V \) such that \( U'U = V \):

\[
\begin{align*}
\bar{X} &= ((U^{-1})' \otimes I_T) \begin{bmatrix} I_{m_y} \otimes X_y & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes X_z \end{bmatrix} \\
\bar{Y} &= ((U^{-1})' \otimes I_T) \begin{bmatrix} I_{m_y} \otimes Y & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes Z \end{bmatrix} \\
N_{XX}(V) &= \bar{X}' \bar{X} = \begin{bmatrix} I_{m_y} \otimes X_y' & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes 1_T' \end{bmatrix} (V^{-1} \otimes I_T) \begin{bmatrix} I_{m_y} \otimes X_y & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes 1_T \end{bmatrix} \\
N_{XY}(V) &= \begin{bmatrix} I_{m_y} \otimes X_y' & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes 1_T' \end{bmatrix} (V^{-1} \otimes I_T) \begin{bmatrix} I_{m_y} \otimes Y & 0_{T(m_y p + 1) \times T m_z} \\ 0_{T m_z \times T m_y} & I_{m_z} \otimes Z \end{bmatrix} \\
S_T(\beta) &= \frac{1}{\nu_0 + T} \begin{bmatrix} (Y - XB)' \\ (Z - 1_T \mu_z)' \end{bmatrix} \begin{bmatrix} (Y - XB) & (Z - 1_T \mu_z') \end{bmatrix} + \frac{\nu_0}{\nu_0 + T} S_0.
\end{align*}
\]

Given the above definitions, the following Lemma holds (Rossi et al., 2005, ch. 3.5):

**Lemma 2.** The conditional likelihoods are, respectively, conditionally conjugate with Normal and Wishart priors. Given independent priors \( \beta \sim \mathcal{N}(\tilde{\beta}_0, N_0) \) and \( V^{-1} \sim \mathcal{W}(\nu_0 S_0^{-1}, \nu_0) \), the conditional posterior distributions are given by:

\[
\begin{align*}
\tilde{\beta}_T(V) &= (N_{XX}(V) + N_0)^{-1}(N_{XY}(V) + N_0 \tilde{\beta}_0) \\
\beta|V, Y^T &\sim \mathcal{N}(\tilde{\beta}_T(V), (N_{XX}(V) + N_0)^{-1}), \quad (3.14a) \\
V^{-1}|\beta, Y^T &\sim \mathcal{W}(S_T(\beta)^{-1}/(\nu_0 + T), \nu_0 + T). \quad (3.14b)
\end{align*}
\]
If $X_z = X_y$, then $\tilde{X} = I_{m_c+m_z} \otimes X_y$ and $N_{XX} = (V^{-1} \otimes X'_y X_y)$ and analogously for $N_{XY}$. In this special case, closed forms are available for the marginal distribution of $V$, allowing to draw directly from the posterior. In general, however, no closed form posterior is available and numerical algorithms have to be used.

The following algorithm implements the Gibbs sampler:

1. Initialize $V^{(0)} = S_T(\tilde{\beta}_T)$.
2. Repeat for $i = 1, \ldots, n_G$:
   (a) Draw $\beta^{(i)}|V^{(i-1)}$ from (3.14a).
   (b) Draw $V^{(i)}|\beta^{(i)}$ from (3.14b).

3.1.3 Inferring the response of additional variables

In the empirical application to tax shocks, I face the challenge that only few (namely 13) non-zero observations on the tax shocks are available. Without additional (prior) information, inference on the effect of tax shocks becomes infeasible when many shocks are present.\(^{10}\) When extending the analysis to include additional variables of interest, I therefore maintain the assumption in (3.4) that the variables in the above VAR are influenced by a number of shocks equal to the number of variables only. Following Uhlig (2003), I call the above VAR the “core” and the extra parameters the “periphery”.

In the language of Jarocinski and Mackowiak (2013), I am imposing Granger-causal priority of the core variables over the peripheral variables. This is akin to the forecasting of variables outside the DSGE model using an estimated DSGE model in Schorfheide et al. (2010). The key premise is that the core VAR allows to identify the structural shocks and the peripheral model then allows to trace the effect of these shocks on additional variables not needed for the shock identification.\(^{11}\) Implicitly, a similar assumption seems to be underlying the rotation of additional variables in and out of a smaller VAR in Ramey (2011).

---

\(^{10}\)Inference becomes infeasible insofar as the Bayesian confidence intervals blow up. This is consistent with the classical properties I find for my Bayesian estimator in the Monte Carlo study of section 5.2: As the number of instruments approaches the number of shocks, the actual size of the confidence intervals becomes increasingly larger than their nominal size.

\(^{11}\)With a proper prior, one could use Bayes factors to test the exclusion restrictions. However, given the SUR-framework, the elegant results for standard VARs outline in Jarocinski and Mackowiak (2013) are not applicable here.
Zha (1999) shows that such a block-recursive system can easily be implemented by including current values of the core variables as (exogenous) regressors to a separate periphery-VAR, for which standard inference applies:

\[
Y_{p,t} = B^p_p Y_{p,t-1} + B^p_c Y_{c,t-1} + B^p_{c,0} Y_{c,t} + A^p_{p,c} p.
\] (3.15)

Conditional on current core variables, the peripheral variables in (3.15) are independent of the shocks in (3.4). Given \( Y^T \), the posterior is therefore independent and parameters are drawn according to the following hierarchical procedure (e.g. Uhlig, 1994):

\[
\begin{align*}
\Sigma_p^{-1} &\equiv (A^p_p (A^p_p)' )^{-1} \sim \mathcal{W}_n(\left((\nu_0 + T) S_{p,T}\right)^{-1}, (\nu_0 + T)), \\
N_{p,T} &\equiv N_{p,0} + X_p' X_p, \\
S_{p,T} &\equiv \frac{1}{\nu_T} \left( \hat{B}_p - \bar{B}_{p,0} \right)' N_{p,0}^{-1} X_p' X_p \left( \hat{B}_p - \bar{B}_{p,0} \right) + \frac{\nu_0}{\nu_T} S_0 + \frac{\nu_T - \nu_0}{\nu_T} \hat{\Sigma}_p \\
B_p | \Sigma_p &\sim \mathcal{N}(B_{p,T}, \Sigma_p \otimes N_{p,T}^{-1}).
\end{align*}
\] (3.16a)

Here, \( \hat{B}_p = (X_p' X_p)^{-1} X_p' Y_p \), \( \bar{B}_{p,T} = N_{p,0}^{-1} N_{p,T}^{-1} X_p' X_p \left( \hat{B}_p - \bar{B}_{p,0} \right) \), \( \hat{\Sigma}_p = T^{-1} X_p' (I - X_p' X_p)^{-1} X_p' Y_p \). In the computations I use the flat prior suggested by Uhlig (1994) with \( \nu_0 = 0, N_0 = \mathbf{0} \).

As in Uhlig (2003), the response of peripheral variables to identified shocks is \( B^p_{c,0} \hat{A}_{q IV} \) on impact. In general, for the core and periphery, the model can be stacked to embody exclusion restrictions and extra lags to yield the response at time \( h \) as

\[
\left[ I_{m + m_p} \quad 0_{(m + m_p) \times (p-1)(m + m_p)} \right] B^h \left[ \begin{matrix} \hat{A}_{q IV} \\ 0_{(p-1)(m + m_p) \times m} \end{matrix} \right].
\] 12

3.2 Narrative DSGE-VAR

Having established identification and estimation in the purely narrative VAR, I show how to recover a class of policy rule coefficients in the DSGE model from the narrative BVAR. Then I adapt the idea of dummy variables (e.g. Del Negro and Schorfheide, 2004) to the above SUR framework for estimating the narrative VAR.

---

12 As suggested in Rossi et al. (2005, ch. 2.12), I use an equivalent representation of the posterior for the numerical implementation: I use that \( \mathcal{N}(\mathbf{0}, (\Sigma \otimes (X'X)^{-1})) \equiv (\text{chol}(\Sigma) \otimes \text{chol}((X'X)^{-1})) \mathcal{N}(\mathbf{0}, I) \) and draw using the identity \((\text{chol}(\Sigma) \otimes \text{chol}((X'X)^{-1})) \text{vec}(Z) = \text{vec}(\text{chol}((X'X)^{-1}))Z \text{chol}(\Sigma)') \), with \( \text{vec}(Z) \sim \mathcal{N}(\mathbf{0}, I) \).
In addition, I propose to test the identifying Assumption 1 within the context of the DSGE model. Note that these results can be extended to also add peripheral variables. In that case, however, the independence of the estimation of the core VAR parameters from the periphery parameters is lost since the posterior over the structural parameters depends on both the core and the periphery posterior density kernels. For simplicity, I abstract from this extension in what follows.

3.2.1 Identification of policy rules using instruments

For more than one instrument, \( m_z > 1 \), the narrative VAR identifies shocks only up to an arbitrary rotation. How to choose among these infinitely many rotations? Mertens and Ravn (2013) argue that in their application, the factorization is of little practical importance since two different Choleski compositions yield almost identical results. Here, I argue that for a class of policy rules, the lower-Choleski decomposition actually recovers the true impact matrix \( \alpha^{[1]} \) in the DSGE model.\(^{13}\)

I call the class of policy rules for which the narrative VAR correctly recovers \( \alpha^{[1]} \) “simple Taylor-type rules”:

**Definition 1.** A simple Taylor-type rule in economy (2.1) for variable \( y_{p,t} \) is of the form:

\[
y_{p,t} = \sum_{i=m_p+1}^{m} \psi_{p,i} y_{i,t} + \lambda_p X_{t-1} + \sigma_p \epsilon_{p,t},
\]

where \( \epsilon_{p,t} \subset \epsilon_{Y,t} \) is iid and \( y_{i,t} \subset Y_t, i = 1, \ldots, n_p \).

Note that the policy rules is defined with respect to the set of observables in the structural model. The canonical Taylor Rule for monetary policy based on current inflation and the output gap is a useful clarifying example: Only when the output gap is constructed based on observables, is it a “simple” policy rule according to Definition 1.

\(^{13}\)Note that showing that two different Choleski factorization do not affect the results substantially does not generally mean that the results are robust to different identifying assumptions (cf. Watson, 1994, fn. 42). If taxes follow what I define below as simple Taylor rules, however, the two Choleski factorizations coincide, rationalizing the exercise in Mertens and Ravn (2013).
Example 1. An interest rate rule with observables only is a simple Taylor rule when output $y_t$ and inflation $\pi_t$ are observed:

$$r_t = (1 - \rho_r)(\gamma_\pi \pi_t + \gamma_y y_t) + \rho_r r_{t-1} + (1 - \rho_r)(-\gamma_y y_{t-1}) + \omega_r \epsilon_t^r.$$  

maps into the above with $\lambda_p = e_r \rho_r + (1 - \rho_r)(-\gamma_y)e_y$ and $\psi_1 = (1 - \rho_r)\gamma_\pi, \psi_2 = (1 - \rho_r)\gamma_y$ and $\sigma_p = \omega_r$.

Example 2. Taylor type rules with a dependence on an unobserved output gap, i.e. $\tilde{y}_t = y_t - y^f_t$ with $y^f_t \not\subset Y_t$ are not simple Taylor rules.

$$r_t = (1 - \rho_r)(\gamma_\pi \pi_t + \gamma_y \tilde{y}_t) + \rho_r r_{t-1} + \omega_r \epsilon_t^r$$  

$$y^f_t = B^*_y X^*_t - 1 + A^*_y \epsilon_t$$

In general, the entire column vector $A^*_y$ is non-zero, violating the exclusion restrictions for a simple Taylor rule.

The following proposition shows that in the DSGE model with simple policy rules, $S_1$ has a special structure that allows to identify it uniquely using $\Gamma, \Sigma$, up to a normalization. Equivalently: When the analogue of Assumption 1 holds in the structural model, the narrative VAR recovers the actual policy rules based on the procedure in (3.11).

Proposition 1. Assume $\Sigma = AA' = A^*(A^*)'$ and order the policy variables such that the $m_p = m_z$ or $m_p = m_z - 1$ simple Taylor rules are ordered first and $\Gamma = [G, 0]A^*$. Then $\alpha^{[1]}$ defined in (3.11) satisfies $\alpha^{[1]} = A^*[I_{m_z}, 0]_{(m-m_z) \times (m-m_z)}'$ up to a normalization of signs on the diagonal if

(a) $m_z$ instruments jointly identify shocks to $m_p = m_z$ simple Taylor rules w.r.t. the economy (2.1), or

(b) $m_z$ instruments jointly identify shocks to $m_p = m_z - 1$ simple Taylor rules w.r.t. the economy (2.1) and $\psi_{p,m_z} = 0, p = 1, \ldots, m_p$.

Proof: See Appendix A.2.

While the proof proceeds by Gauss-Jordan elimination, the intuition in case (a) can be understood using partitioned regression logic: $S_1'S_1'$ is the residual variance of the first $m_p$ forecast errors after accounting for the forecast error variance to the
last \( m - m_p \) observed variables. Including the non-policy variables that enter the Taylor rule directly among the observables controls perfectly for the systematic part of the policy rules and leaves only the variance-covariance matrix induced by policy shocks. Since this variance-covariance matrix is diagonal with simple Taylor rules, the Choleski-decomposition in (3.11) works.

Formally, the Choleski-factorization of \( S_1S_1' \) proposed by Mertens and Ravn (2013) imposes the \( m_z(m_z - 1) \) zero restrictions needed for exact identification. The structure imposed by having simple Taylor rules rationalizes these restrictions in a class of DSGE models. In fact, the mechanics of the proof would carry through if the block of policy rules had a Choleski structure, confirming that what is needed for identification via instrumental variables in the model is precisely the existence of \( m_z(m_z - 1) \) restrictions. More generally, identification requires restrictions on the contemporaneous interaction between policy instruments that need not have the form of simple Taylor rules.

### 3.2.2 Prior elicitation

A natural way to elicit a prior over the parameters of the VAR, tracing back to Theil and Goldberger (1961), is through dummy observations. Del Negro and Schorfheide (2004) use this approach to elicit a prior for the VAR based on a prior over structural parameters of a DSGE model. Here I adapt their approach for use within the SUR framework. Using a dummy variable prior with normally distributed disturbances no longer yields a closed form prior, as in Lemma 2. Using dummy variables, however, still generates conditional posteriors in closed form and conditional priors with the same intuitive interpretation as in Del Negro and Schorfheide (2004).

Since the likelihood function of the SUR model is not conjugate with a closed form joint prior over \((\beta, V^{-1})\), unless it collapses to a standard VAR model, the DSGE model implied prior also fails to generate unconditionally conjugate posteriors. I consider two alternatives: First, a dummy observation approach that preserves the conditional conjugacy, but that does not have a closed form representation for the prior density. Second, I consider an independent prior based. This second approach has the advantage that the prior is available in closed form, but the con-
ditional prior variance of $\beta$ is necessarily independent of $V^{-1}$, unlike in the SUR model.

Following Del Negro and Schorfheide (2004), I generate the prior for $B, V^{-1}$ by integrating out the disturbances to avoid extra sampling error in the prior. That is, the prior is centered at:

$$
\begin{align*}
\bar{B}_0^B(\theta) &= \mathbb{E}^{DSGE}[X_0X_0'|\theta]^{-1}\mathbb{E}^{DSGE}[X_0Y_0'|\theta] \quad \Leftrightarrow \quad \bar{\beta}_0^B(\theta) = \text{vec}(\bar{B}_0^B(\theta)) \quad (3.17a)
\end{align*}
$$

$$
\begin{align*}
\bar{V}_0(\theta)^{-1} &= \begin{bmatrix} A(\theta)^*(A(\theta)^*)' & (A(\theta)^*)[I_{m_z}, 0]' \\
[I_{m_z}, 0](A(\theta)^*)' & \omega_z^2\text{diag}(A_1(\theta)^*(A_1(\theta)^*)') + A_1(\theta)^*(A_1(\theta)^*)'
\end{bmatrix}, \\
\end{align*}
$$

(3.17b)

where $\mathbb{E}^{DSGE}[\cdot|\theta]$ denotes the unconditional expectation based on the linear DSGE model (2.1) when the coefficient matrices are generated by the structural parameters $\theta$. $\bar{V}_0(\theta)$ satisfies Assumption 1 with $G = I_{m_z}$ and assuming that the measurement error across instruments is independent and equal to $\omega_z^2$ times the univariate shock variance.

To implement the prior, use the following dummy observations and the corresponding likelihood: The corresponding dummy observations are:

$$
\begin{align*}
[Y_0(\theta), Z_0(\theta)] &= \left[ \sqrt{T_0^B} \left[ \text{chol}_u(\mathbb{E}^{DSGE}[X'_yX_y|\theta]) \right] \left( \text{chol}_u(\mathbb{E}^{DSGE}[X'_yX_y|\theta])\right)_z \bar{\beta}_0(\theta) \right] \\
X_0(\theta) &= \left[ \sqrt{T_0^B} \left[ \text{chol}_u(\mathbb{E}^{DSGE}[X'_yX_y|\theta]) \right] \left( \text{chol}_u(\mathbb{E}^{DSGE}[X'_yX_y|\theta])\right)_z \right] \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\
\end{align*}
$$

(3.18a)

(3.18b)

where $(\text{chol}_u(V))_z$ denotes column $z$ of $\text{chol}_u(V)$. Note that this decomposition ensures that when only the dummy observations are used and $V = \bar{V}(\theta)$ that when treated as in the VAR, that the following target matrix is recovered:

$$
\tilde{X}_{0, SUR}(\theta)' \tilde{X}_{0, SUR}(\theta) = \text{chol}(\mathbb{E}^{DSGE}[X'_YX_YV^{-1}(\theta)\otimes I_{p(m+m_z)}]X_{SUR|\theta})).
$$

The prior incorporates the Normal likelihood over the dummy observations and Jeffrey’s prior over $V^{-1}$ along with scale factors chosen to make the prior information
equivalent to \( T_0^B \) observations about \( B \) and \( T_0^V \) observations on \( V^{-1} \). Lemma 2 implies that, absent the \( |V^{-1}|^{-n_y/2} \) factor, the corresponding marginal priors are given by:

\[
\beta | V^{-1}, \theta \sim N(\beta_0(\theta), (T_0^B \times \hat{X}_0'(V^{-1} \otimes I)\hat{X}_0')^{-1}),
\]

\[
\hat{X}_0 = \text{diag}([X_0^\theta, \ldots, X_0^\theta, X_0^\theta, \ldots, X_0^\theta]),
\]

\[
V^{-1} | \beta, \theta \sim W_{m+m_\theta}(SSR_0(\beta, \theta)^{-1}, n_y + T_0^V)
\]

\[
SSR_0(\beta, \theta) = T_0^V \times \hat{V}_0(\theta) + T_0^B([Y_0(\theta), Z_0(\theta)] - X_0B(\beta))([Y_0(\theta), Z_0(\theta)] - X_0B(\beta))'.
\]

As shown in Appendix A.3, in the version with the adjustment factor, the degrees of freedom of the conditional prior for \( V^{-1} \) are simply \( T_0^V \).

For fixed \( \theta \), e.g. \( \theta \) fixed at its prior mean \( \bar{\theta}_0 \), inference proceeds according to Lemma 2 with a flat prior, treating the dummy observations as regular observations, but setting the prior degrees of freedom to \( \nu_0 = -n_y \).

**Hierarchical prior: Structural parameters**

If \( \theta \) has a non-degenerate prior distribution, an additional sampling step is necessary to sample from the conditional posterior \( \theta | \beta, V^{-1} \) that I now discuss. Allowing for a non-degenerate distribution then also yields estimates of the structural parameters of the DSGE model as a by-product of the DSGE-VAR estimation.

The prior and data density are given by:

\[
\theta \sim \pi(\theta)
\]

(3.19a)

\[
\pi(B, V^{-1} | \theta) \propto |V^{-1}|^{-n_y/2} \ell(B, V^{-1} | Y_0(\theta), Z_0(\theta))
\]

(3.19b)

\[
\tilde{f}(Y, Z | B, V^{-1}, \theta) = f(Y, Z | B, V^{-1}).
\]

(3.19c)

The conditional posterior for \( B, V^{-1} | \theta \) is as characterized before. The conditional posterior for \( \theta \) can be written as:

\[
\pi(\theta | B, V^{-1}, Y, Z) = \frac{f(Y, Z | B, V^{-1})\pi(B, V^{-1} | \theta)\pi(\theta)}{\int f(Y, Z | B, V^{-1})\pi(B, V^{-1} | \theta)\pi(\theta)d\theta} = \frac{p(B, V^{-1} | \theta)\pi(\theta)}{\int p(B, V^{-1} | \theta)\pi(\theta)d\theta}
\]

(3.20)

\[
\propto \pi(B, V^{-1} | \theta)\pi(\theta),
\]

as in the example in Geweke (2005, p. 77).
The following Gibbs Sampler with Metropolis algorithm implements the hierarchical prior:

1. Initialize VAR parameters: Set $B_{(0)}, V_{(0)}^{-1}$ to OLS estimates.

2. Initialize structural parameters: $\theta_0 = \int_\Theta \theta \pi(\theta) d\theta$.

3. Find high density point: $\theta_h = \text{arg max}_\theta \pi(B_{(0)}, V_{(0)}^{-1}|\theta)\pi(\theta)$.

4. Metropolis-Hastings within Gibbs:
   
   (a) With probability $p_h$, set $i = h$, else set $i = rw$.
   
   (b) Draw a candidate $\theta_c$ from pdf $f_i(\cdot|\theta_{(d-1)})$. If $i = h$, the distribution is $\theta_c \sim N(\theta_h, \text{diag}(\sigma_i^2, \text{prior})/s)\theta_{(d-1)}$. Else, draw $\theta_c$ from $\theta_c \sim N(\theta_{(d-1)}), \Sigma)$.
   
   (c) With probability $\alpha_{d-1,i}(\theta_c)$, set $\theta_{(d)} = \theta_c$, otherwise set $\theta_{(d)} = \theta_{(d-1)}$.

   \[
   \alpha_{d-1,i}(\theta_c) = \min\left\{1, \frac{\pi(B_{(d-1)}, V_{(d-1)}^{-1}|\theta_c)\pi(\theta_c) f_i(\theta_{(d-1)}|\theta_c)}{\pi(B_{(d-1)}, V_{(d-1)}^{-1}|\theta_{(d-1)})\pi(\theta_{(d-1)}) f_i(\theta_c|\theta_{(d-1)})}\right\},
   \]

   (d) Draw $B_{(d)}|\theta_{(d)}, V_{(d-1)}^{-1}$ given by (3.14a) and including dummy-observations.

   (e) Draw $V_{(d)}^{-1}|B_{(d)}, \theta_{(d)}$ given by (3.14b) and including dummy-observations.

   (f) If $d < D$, increase $d$ by one and go back to (a), else exit.

Note that when $i = rw$, a symmetric proposal density is used so that the $f_{rw}$ term drops out of the acceptance probability $\alpha_{d-1,rw}(\cdot)$.

In practice I use a random-blocking Metropolis-Hastings with random walk proposal density with t-distributed increments with 15 degrees of freedom. The variance-covariance matrix of the draws is calibrated during a first burn-in phase, while a second burn-in phase is used to calibrate the jump size such that the average acceptance rate across parameters and draws is about 30%. The order of the parameters is uniformly randomly permuted and a new block is started with probability 0.15 after each parameter. This Metropolis-Hastings step is essentially a simplified version of the algorithm proposed in Chib and Ramamurthy (2010). Similar to their application to the Smets and Wouters (2007) model I otherwise obtain a small
effective sample size due to the high autocorrelation of draws when using a plain random-walk Metropolis-Hastings step.\textsuperscript{15}

\section*{Structural and reduced form trends}

The DSGE model implies that the detrended and demeaned variables $\tilde{Y}_t$ are governed by endogenous dynamics, whereas the VAR models the sum of deterministic and stochastic elements. The DSGE model trends and steady state parameters map into the deterministic VAR terms as follows:

\begin{equation}
\tilde{Y}_t = \sum_{l=1}^{k} A_l \tilde{Y}_{t-l} + \epsilon_t, \quad \tilde{Y}_t \equiv Y_t - \tilde{D}_y t - \tilde{\mu}_y.
\end{equation}

\(\Leftrightarrow\)

\begin{equation}
Y_t = \left( \sum_{l=1}^{k} A_l Y_{t-l} \right) + \left( I - \sum_{l=1}^{k} A_l \right) \tilde{D}_y t + \left( \tilde{\mu}_y - \sum_{l=1}^{k} A_l (\tilde{\mu}_y - \tilde{D}_y l) \right) + \epsilon_t
\end{equation}

Equation (3.23) maps the reduced-form trends and means into those of the DSGE model: $\tilde{D}_y = \left( I - \sum_{l=1}^{k} A_l \right)^{-1} D_y$, and $\tilde{\mu}_y = \left( I - \sum_{l=1}^{k} A_l \right)^{-1} \left( \mu_y - \sum_{l=1}^{k} A_l \tilde{D}_y l \right)$ (see also Villani (2009)).

To compute variances and covariances as required to implement the dummy variable prior, I first compute the population moments of $\tilde{Y}_t$ in (3.22), using that $E[\tilde{Y}_t] = 0$. In a second step, I then compute the implied moments for $Y_t$ itself, using finite $T$ deterministic moments for the deterministic mean and trend terms.\textsuperscript{16} Since the detrended endogenous variables and the deterministic trends are independent, this also allows to first back out the prior VAR polynomial and then the implied prior trend and intercepts.

\subsection*{3.2.3 Instrument validity}

Tests of instrument validity typically require overidentification. Conditional on a parametric model such as (2.1), the identifying Assumption 1 can, however, also

\textsuperscript{15}See also Herbst and Schorfheide (forthcoming).

\textsuperscript{16}In particular, use that $T^{-1} \sum_{t=1}^{T} t = \frac{T(T+1)}{2}$, $T^{-1} \sum_{t=1}^{T} t^2 = \frac{(T+1)(2T+1)}{6}$, and $T^{-1} \sum_{t=1}^{T} t^3 = \frac{T(T+1)^2}{4}$. 

20
be tested from the model perspective by assessing its fit of the data in terms of marginal densities.

To be precise, to test the validity of instrument \( Z_i \), I include the following observation equation in (2.1a):

\[
z_{i,t} = c_i \epsilon_{i,t} + \sum_{j \neq i} c_j \epsilon_{j,t} + u_{i,t}, \quad u_{i,t} \sim \text{iid } N(0, 1). \quad (3.24)
\]

Under Assumption 1 applied to the univariate case, \( c_i = G \) and \( c_j = 0 \forall j \neq i \). To test this assumption I compare the marginal data densities of the model (2.1) with and without imposing \( c_j = 0 \forall j \neq i \). When not imposing \( c_j = 0 \forall j \neq i \), I use a prior a Normal prior for \( c_j \) centered at zero. In particular: \( c_j \sim N(0, \frac{1}{2}) \). For \( c_i \) I consider both \( c_i \sim N(1, \frac{1}{2}) \) as well as \( c_i = 1 \) after scaling the instrument to be measured in units of \( \epsilon_{i,t} \).

4 Empirical specification

4.1 Data and sample period

I use an updated version of the Smets and Wouters (2007) dataset for the estimation of both the DSGE and VAR model. I follow their variable definitions with two exceptions: I include consumer durables among the investment goods and use total hours worked from Francis and Ramey (2009) rather than private hours. They show that government hours differ in the early post-war period systematically from private hours worked. This can be particularly relevant when examining the effect of fiscal spending on the economy.\(^{17}\) I also extend the sample period to 1947:Q1 to 2007:Q4, stopping before the ZLB became binding.\(^{18}\)

Narrative shock measures are taken from an update of Romer and Romer (2004) for monetary policy shocks,\(^{19}\) the Survey of Professional Forecasters’ real defense spending forecast errors from Ramey (2011) for government spending, and Mertens and Ravn (2013) for tax policy shocks. These are, broadly speaking, the subset of

\(^{17}\)I normalize all measures by the civilian non-institutional population as in Smets and Wouters (2007), but unlike Francis and Ramey (2009).

\(^{18}\)A significant downside of this strategy is that one period of significant fiscal changes during peacetime is omitted.

\(^{19}\)The underlying Greenbook data involved in the update are preliminary and based on ongoing work by the Real Time Data Research Center at the Federal Reserve Bank of Philadelphia. The data will become publicly available on the RTDRC website no later than June 30, 2015.
the shocks in Romer and Romer (2010) which are not considered to be motivated by economic conditions. \(^{20}\)

While it is arguably important to include a long sample covering WWII and the Korean War to have important instances of exogenous variation in government spending (cf. Ramey, 2011), I consider subsamples of the data before 1980 and after 1982 in robustness exercises. \(^{21}\)

I follow Leeper et al. (2010) in constructing time series for taxes, except for including state and local tax revenue in the calculation of revenue, tax, and debt data, similar to Fernandez-Villaverde et al. (2011). The significant share of state and local governments in both government debt and spending motivates this broader definition: Using Flow of Fund (FoF) data to initialize the time series of government debt at par value, I calculate debt levels using the cumulative net borrowing of all three levels of government. Municipal debt as a share of total government debt has increased from little over 5% in 1945 to around 35% in 1980 before falling to about 20% in 2012 according to FoF data. \(^{22}\) The share of state and local governments in government consumption and investment has been above 50% since 1972. Details are given in Appendix A.6.

### 4.2 DSGE model specification

In this section I outline the empirical specification of the generic DSGE model in (2.1). The model is based on the standard medium-scale New-Keynesian model as exemplified in Christiano et al. (2005) and follows closely Smets and Wouters (2007). There is monopolistic competition in intermediate goods markets and the labor market and Calvo frictions to price and wage adjustment, partial price and wage indexation, and real frictions such as investment adjustment cost and habit

---

\(^{20}\) Since instrument data is scarce, I follow Mertens and Ravn (2013) and set missing observations for instruments to zero. While modeling missing observations is, in principle, possible, having only five tax instrument observations overlap with the spending instrument limits the scope for inference.

\(^{21}\) Bohn (1991) uses both annual data from 1792 to 1988 and quarterly data from 1954 to 1988. He argues that long samples are preferable because debt is often slow moving and because it contains more episodes such as wars which might otherwise appear special. However, he finds that the results for the shorter quarterly sample are qualitatively unchanged.

formation. I add labor, capital, and consumption taxes as in Drautzburg and Uhlig (2011) and fiscal rules as in Leeper et al. (2010) and Fernandez-Villaverde et al. (2011). Here I only discuss the specification of fiscal and monetary policy. The remaining model equations are detailed in Appendix A.5. Priors, when available, follow Smets and Wouters (2007). The exception is the labor supply elasticity, which I fix at unity prior to estimation.\footnote{Otherwise the estimation would imply an implausibly low labor supply elasticity.}

The monetary authority sets interest rates according to the following standard Taylor rule:

\[
\hat{r}_t = \rho_r \hat{r}_{t-1} + (1 - \rho_r) \left( \psi_{r,\pi} \hat{\pi}_t + \psi_{r,y} \bar{\Delta} \bar{y}_t + \psi_{r,\Delta} \Delta \bar{y}_t + \epsilon_t^r \right),
\]

(4.1)

where \(\rho_r\) controls the degree of interest rate smoothing and \(\psi_{r,x}\) denotes the reaction of the interest rate to deviations of variable \(x\) from its trend. \(\bar{y}\) denotes the output gap, i.e. the deviation of output from output in a frictionless world.\footnote{Money supply is assumed to adjust to implement the interest rate and fiscal transfers are adjusted to accommodate monetary policy.}

The fiscal rules are:

\[
\hat{g}_t = -\psi_{g,y} \bar{y}_t - \psi_{g,b} \bar{b} \gamma \hat{b}_t + \xi_t^g
\]

(4.2a)

\[
\hat{s}_t = -\psi_{s,y} \bar{y}_t - \psi_{s,b} \bar{b} \gamma \hat{b}_t + \xi_t^s
\]

(4.2b)

\[
\bar{w} \bar{n} d\bar{y}_t = \psi_{\tau,y} \bar{y}_t + \psi_{\tau,b} \bar{b} \gamma \hat{b}_t + \xi_t^{\tau,n}
\]

(4.2c)

\[
\bar{r}^k - \delta \bar{k} d\bar{y}_t = \psi_{\tau^k,y} \bar{y}_t + \psi_{\tau^k,b} \bar{b} \gamma \hat{b}_t + \xi_t^{\tau,k}
\]

(4.2d)

The disturbances \(\xi_t\) follow exogenous AR(1) processes: \(\xi_t^0 = \rho_s \xi_{t-1}^0 + \epsilon_t^0\).\footnote{Note that Leeper et al. (2010) assume there is no lag in the right hand side variables, while Fernandez-Villaverde et al. (2011) use a one quarter lag.} Note that the sign of the coefficients in the expenditure components \(g_t, s_t\) are flipped so that positive estimates always imply consolidation in good times (\(\psi_{\tau,y} > 0\)) or when debt is high (\(\psi_{\tau,b} > 0\)).

\footnote{Not only the fiscal policy shocks, but all shocks in my specification follow univariate AR(1) processes, unlike Smets and Wouters (2007) who allow some shocks to follow ARMA(1,1) processes. Ruling out MA(1) components helps to guarantee that a VAR can approximate the DSGE model dynamics as discussed by Fernandez-Villaverde et al. (2007).}
The consolidated government budget constraint is:

$$\frac{\bar{b}}{\bar{y}}(\hat{b}_t - \hat{r}) + \frac{\bar{w}\bar{n}}{\bar{y}}(d_{\pi}^n + \pi^n(\hat{\bar{w}}_t + \hat{n}_t)) + \frac{\bar{e}}{\bar{y}}\bar{r}_t\bar{c}_t + \frac{(\bar{r}^k - \delta)\bar{k}}{\bar{y}}(d_{\tau}^k + \pi^k(\hat{\bar{r}}_t^k \bar{r}_t^k - \delta + \hat{k}_t^p))$$

$$= \hat{g}_t + \hat{s}_t + \frac{\bar{b}}{\gamma \bar{y}}(\hat{b}_{t-1} - \hat{\pi}_t)$$  \hspace{1cm} (4.3)

The debate about variable selection in singular DSGE models is still ongoing, cf. Canova et al. (2013) and the comment by Iskrev (2014). When taking the model to the data, I consider as many structural shocks as observables in the observation equation of the DSGE model (2.1a) and the VAR (2.2a). Including the policy variables naturally suggests to include the corresponding policy shocks. When including GDP, I add a TFP shock. The discount factor shock is included when consumption is included, the investment efficiency shock is included when investment is included, the respective markup shocks are included when inflation or wage growth (or hours) are included, and a shock to transfers is included when debt is included in the estimation.28 I use Dynare (Adjemian et al., 2011) to solve the model and to estimate the pure DSGE model.

### 4.3 VAR model specification

Given the sparse data on the fiscal instruments, I only consider a relative small-scale VAR in the three policy instruments of government spending, personal income taxes, and the Federal Funds Rate, as well as output, inflation, and the ratio of government debt to GDP. Consumption, investment, hours, the real wage, and revenue are included in the periphery.

Is the VAR approximation to the underlying DSGE model appropriate? In the baseline specification, I set the lag length $p$ to four. While $p = 4$ falls short on an infinite lag approximation, I implement the ABCD invertibility condition in

---

27Seigniorage revenue for the government enters negatively in the lump-sum transfer to households $\hat{s}_t$.

28When including variables in the periphery and generating a DSGE-model based prior for the block-recursive Core-Periphery VAR, I append measurement error to the peripheral variables.

29With a flat prior, I cannot formally evaluate the exclusion restrictions implied by the Core-Periphery structure. The forecast performance is, however, indicative of whether the the exclusions restrictions are at odds with the data. Table 2 in the appendix shows that, with the exception of inflation and hours, the peripheral variables improve forecasts very little: the R-square barely changes when they are included. For inflation, adding five more variables would increase the forecasting power from a little less than 74% to little less then 79%. Hours are hardest to predict without the peripheral variables but it is a peripheral variable itself.
Fernandez-Villaverde et al. (2007) and verify that, at the prior mean, the DSGE model has a \( \text{VAR}(\infty) \) representation. Extending the lag-length from four to six quarters had no noticeable effect relative to the baseline specification.

In line with the benchmark DSGE model, I allow for a linear, deterministic trend in all variables (without imposing a common trend, however). The results appear to be robust to using a quadratic trend and no trend, as in Ramey (2011) and Mertens and Ravn (2013), respectively.

To calibrate the Gibbs sampler, I choose a 1,000 draw burn-in period and keep every third draw thereafter, until accumulating 2,000 draws. In the case of a flat prior, this generates comparable distributions of model summary statistics across the first and second half of the Markov chain, as well as negligible autocorrelations. See Appendix A.7.2.

5 Results

This section presents estimation results for the VAR with a flat prior. One finding is that the posterior confidence bands surrounding are often wide, which motivates a Monte Carlo study of its sampling properties. With enough in-sample variation, I find that the sampling properties of the proposed estimator are good. With limited in-sample information, however, the confidence bands tend to be conservative, implying a role for prior information. I therefore move on to discuss the preliminary findings of the DSGE-VAR estimates, focusing on the policy rule estimates.

5.1 VAR estimates

Shocks identified with narrative instruments have roughly the effects one would expect a priori, as Figure 1 shows: Increased government spending raises output, whereas tax increases raise revenue, but depress output, as does a monetary tightening. The results are, however, imprecisely estimated. The output effects of taxes and government spending are only marginally significant at the 68% level.

Narrative shocks only identify the space jointly spanned by all shocks and the choice of factorization can matter. While Figure 1 is based on the lower Choleski factorization of \( S_1S'_1 \), Figure 2 identifies the government spending shock instead as the shock which explains the most of its conditional FEVD over the first 0–19 quarters.

\[ \text{Note that the SUR framework here can easily accommodate cross-equation restrictions such as a common trend, as highlighted in the exposition in Geweke (2005).} \]
Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals. Results based on lower Choleski factorization of $S_1S'_1$.

Figure 1: Effects of policy shocks on policy and output: Impulse-response-functions
The results change little for the tax shock and become slightly more expansionary for the government spending shock and less contractionary for the monetary policy shock. In what follows, I therefore focus on the Choleski factorization of $S_1S'_1$ since this factorization has the added advantage of connecting to DSGE models with Taylor-style policy rules.

Figure 2: Effects of policy shocks on output: Impulse-response-functions based on maximum medium-run government spending FEVD

Despite the imprecise estimates, there is clear evidence for contemporaneous interactions between policy instruments: As Figure 3 shows, taxes are estimated to increase in response to an increase in government spending and the federal funds rate, whereas government spending is estimated to fall in response to monetary tightening. Together, these results imply that contractionary movements in fiscal policy instruments go hand in hand with contractionary monetary policy shocks.

Is the estimated fiscal-monetary policy interaction plausible? A recent paper by Romer and Romer (2014) provides qualitative evidence that the Federal Reserve has indeed considered fiscal policy in its monetary policy decisions. Romer and Romer (2014) document staff presentations to the Federal Open Market Committee (FOMC) suggesting monetary accommodation of the 1964 and 1972 tax cuts (p. 38f) as well as monetary easing in response to the 1990 budget agreement. The tax-inflation nexus is also reflected in staff presentations and in the comments by at least one FOMC: According to Romer and Romer (2014) the FRB staff saw social security tax increases in 1966, 1973, and 1981 as exerting inflationary pressure (p. 40). Romer and Romer (2014) conclude that monetary policy did not counteract expansionary tax policy – but may have tightened policy in response to fiscal transfer.
$G$ to $G$ shock

$G$ to Tax

$G$ to FFR shock

Tax to $G$ shock

Tax to Tax

Tax to FFR shock

Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals. Results based on lower Choleski factorization of $S_1S'_1$.

Figure 3: Policy interactions: Impulse-response-functions
programs (p. 41f.).

In the context of my application, the fiscal-monetary policy interaction may explain what appears to be a manifestation of the “price puzzle” (Hanson, 2004, cf.): Inflation actually rises in response to a monetary policy shock. At the same time, however, taxes increase in response to the monetary tightening and tax shocks by themselves tend to be inflationary, see Figure 4.\footnote{An alternative explanation can be the presence of cosh-push shocks within the New Keynesian paradigm, see Rabanal (2007).}

![Figure 4: Price puzzle? Inflation responses to monetary policy shocks and tax shocks](image)

Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals. Results based on lower Choleski factorization of $S_1S'_1$.

These results are not due to omission of commodity prices in the VAR or an artifact of using relatively low-frequency data, rather than monthly data as in Romer and Romer (2004). Appendix A.7.1 presents results for a monthly sample, with and without commodity prices, using either the Romer and Romer (2004) or the Kuttner (2001) monetary policy shock as instruments. In univariate Romer and Romer (2004) style regressions, the monetary policy shock is found to be inflationary initially when it is allowed to have a contemporaneous effect, independent of whether current and lagged commodity prices are controlled for. The same inflationary effect is found in monthly VARs with or without commodity prices when the shock is identified using the Romer and Romer (2004) shocks. In contrast, with or without including commodity prices in the VAR, when shocks are identified using the Kuttner (2001) shocks, which cover a sample period beginning in 1990, no price puzzle is present. Interestingly, across all specifications, the monetary policy shock results in an instant drop in civilian government payroll, in line with the fiscal-monetary policy interaction in the baseline model.
5.2 Simulated VAR sampling properties

The estimated confidence intervals are wide – and wider than they would be if uncertainty about $\Gamma$ were ignored. Rather than drawing $\Gamma$ as part of $V$ from the posterior, one can compute $\hat{\Gamma} = \frac{1}{T}(Y - X_yB_y)'(Z - X_zB_z)$ conditional on the estimated coefficients. Figure 5 compares the posterior uncertainty over the output response to different shocks for both sampling schemes.

![Graphs showing output responses to different shocks](image)

Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals of the SUR-sampling scheme (black and gray) and when the point-estimate $\hat{\Gamma} = \frac{1}{T}(Y - X_yB_y)'(Z - X_zB_z)$ is used for $\Gamma$ (red). Results based on lower Choleski factorization of $S_1S_1'$.

Figure 5: Effects of policy shocks on output: Comparing confidence intervals for IRFs

Given the different results for the posterior uncertainty, which sampling scheme should be used? One criterion to judge the proposed prior and Bayesian inference scheme is to investigate its frequentist properties in a Monte Carlo study. I simulate $N_{MC} = 200$ datasets based on the actual point estimates for the reduced-form VAR as well as the instrument-based inference about a structural impulse-response vector. I initialize the VAR at the zero vector and drop the first 500 observations and keep the last $T = 236$ observations as the basis for estimating the reduced-form VAR. Different scenarios for instrument availability are considered for the structural inference: The fraction of zero observations for the instrument varies from about 5% to 90% of the observations.

The data generating process is given by the OLS point estimate of a VAR in the core variables and its covariance as specified in (3.1). Only one instrument is used: the Romer and Romer (2004) monetary policy shocks. I choose their shocks for the analysis because they are available for about half (47%) of the sample. As in the data, I modify (3.1) by setting randomly chosen observations to zero. The section
of observations set to zero varies from 5% to 90%.

For each dataset \( m \), I compute the pointwise posterior confidence interval using my Bayesian procedure, with and without conditioning on the observed covariance between instruments and VAR forecast errors. Similarly, I use the wild bootstrap proposed in Mertens and Ravn (2013) to conduct frequentist inference.\(^{32}\) Each procedure yields an estimate of the true IRF \( \{ I_{h,m} \}^H_{h=0} \) for each point OLS estimate and a pointwise confidence interval for horizons up to \( H \): \( \{ \hat{C}(\alpha)^j_{h,m} \}^H_{h=0} = \{ [\underline{c}(\alpha)^j_{h,m}, \overline{c}(\alpha)^j_{h,m}] \}^H_{h=0} \). The superscript \( j \) indexes the different methods. \( 1 - \alpha \) is the nominal size of the confidence interval: A fraction \( \alpha \) of draws from the posterior or the bootstrapped distribution for \( I_{h,m} \) should lie under \( \underline{c}(\alpha)^j_{h,m} \) or above \( \overline{c}(\alpha)^j_{h,m} \).

To assess the actual level of the confidence intervals of method \( j \), I compute for each \( (h, m) \) whether the truth at horizon \( h \), \( I_{h,m} \) lies inside the pointwise confidence interval: \( 1 \{ I_{h,m} \in \hat{C}(\alpha)^j_{h,m} \} \). The actual coverage probability is then estimated as:

\[
\hat{\alpha}^j_h = \frac{1}{N_{MC}} \sum_{m=1}^{N_{MC}} 1 \{ I_{h,m} \in \hat{C}(\alpha)^j_{h,m} \}
\]

Figure 6 plots the deviation of the actual coverage probability for model \( j \) at horizon \( h \) from the nominal coverage probability: \( \hat{\alpha}^j_h - \alpha \) for \( \alpha = 0.68 \). The procedures that ignore “first stage” uncertainty about the covariance between the forecast errors and the external instruments understate the size of the confidence intervals substantially at short horizons (“Bayes – certain”, BaC and “Bootstrap – certain”, BoC), while the Bayesian procedure which allows for uncertainty about the covariance matrix errs on the conservative side (“Bayes–uncertain”, BaU). For the latter scheme, the actual level is typically only zero to five percentage points above the nominal level when half of the instruments are nonzero.\(^{33}\) However, when only one in 20 observations for the instrument is non-zero, the actual level exceeds the nominal level by around ten percentage points.

With more than one instrument at a time, the coverage probability of the responses to the different shocks depends on the specific rotation of the shock. Overall, the Monte Carlo study suggests that the proposed Bayesian procedure accounts properly for the uncertain covariance between instruments and forecast errors in the context of the present application, but may be conservative from a frequentist

\(^{32}\)Appendix A.4 describes the algorithm.

\(^{33}\)The actual level depends slightly on the variable under consideration.
point of view both when there is little variation in the instruments or over longer forecast horizons. This motivates using available prior information to improve on the estimates.

5.3 DSGE-VAR results

To compare the DSGE model and the VAR, I begin by comparing separately estimated models in terms of their impulse-responses and the implied histories of historical shocks. I proceed by examining implied policy rules when imposing a DSGE prior over the variance-covariance matrix $V$. Finally, I test the instrument validity in the DSGE model. Ongoing work focuses on estimating the full DSGE-VAR which then also allows for formal model comparisons in terms of Bayes factors, as in Del Negro and Schorfheide (2004).

5.3.1 Impulse-response functions

Consider the response to government spending shocks first: Figure 7 compares the responses from the VAR (in gray) with those from the DSGE model (in red). A
first order difference is the bigger uncertainty in the VAR estimates, stemming from the data-driven identification. For output, interest rates, and government spending itself the two models broadly agree in their implications: Government spending increases past its initial increase, highlighting the importance of allowing for endogenous dynamics rather than imposing an exogenous AR(1) process for government spending. The fiscal expansion causes output to rise, but not the federal funds rate. The models disagree about the timing of the tax increase, the response of inflation, and whether government spending raises private consumption, as in the VAR, or lowers it, as in the DSGE model.

![Graphs showing responses to government spending shock: VAR and DSGE model](image1)

The blue lines represent the median and 60% pointwise Bayesian confidence intervals when shocks are identified using narrative instruments. The red lines represent the median and 60% pointwise Bayesian confidence interval for the estimated DSGE model.

Figure 7: Responses to government spending shock: VAR and DSGE model

The estimated responses to tax shocks are largely consistent across the DSGE model and the VAR: see Figure 8. Tax shocks cause revenue to rise persistently, cause a small increase in inflation, lower output and consumption, and have small effects on government spending. The estimates only disagree about whether the monetary authority increases its interest rate in response to a tax shock.
The blue lines represent the median and 60% pointwise Bayesian confidence intervals when shocks are identified using narrative instruments. The red lines represent the median and 60% pointwise Bayesian confidence interval for the estimated DSGE model.

Figure 8: Responses to tax shock: VAR and DSGE model
The responses to a monetary tightening in Figure 9 show large discrepancies between the different model estimates: While the shock is eventually contractionary, the DSGE model implies no contemporaneous fiscal tightening as estimated in the VAR. The response of output and inflation in the DSGE model are, moreover, typical responses in a New Keynesian model under the Taylor Principle, but opposite the estimated initial increase in output and decrease in inflation in the VAR. The VAR conforms therefore much more with the traditional view of lagged effects of monetary policy echoed in Faust (2009).  

The blue lines represent the median and 60% pointwise Bayesian confidence intervals when shocks are identified using narrative instruments. The red lines represent the median and 60% pointwise Bayesian confidence interval for the estimated DSGE model.

Figure 9: Responses to monetary policy shock: VAR and DSGE model

5.3.2 Policy rule estimates

Motivated by the result in Proposition 1, I use the fact that the narrative identification scheme recovers the actual impact matrix for policy shocks and, as a by-product,
also identifies policy rules under the assumptions in Proposition 1.

Figure 10: Policy rule estimates

Figure 10 presents the estimates for all three policy rules. The upper panel shows results using the flat prior, while the bottom panel augments the flat prior with a DSGE model prior equivalent to another sample worth of observations for the variance-covariance matrix while maintaining a flat prior for the coefficient matrix. The estimates clearly reflect the identifying assumption that neither government spending nor taxes react to monetary policy and that government spending additionally does not react to taxes. Besides these imposed zeros, few reaction coefficients are significantly different from zero without using prior information. With prior information based on the prior mean, however, the results show significant responses. For example, in the case of the monetary policy rule, the rule implies significant monetary tightening in response to higher inflation and government debt. Monetary policy is loosened when taxes fall or government spending rises, consistent
with some degree of monetary accommodation of fiscal policies.

The fact that the estimated monetary policy rule loads on variables other than inflation and output is indicative of the monetary authority putting weight on inflation and the deviation from potential output, which responds to tax and government spending changes, or of other forms of policy interaction.

The fiscal Taylor rules imply reaction coefficients which are not straightforward to interpret, except for further evidence of policy interactions: Taxes are being lowered in response to increases in government spending, potentially to support expansionary policies. Taxes do not react significantly to either debt or output, but are estimated to fall when inflation is high. The reaction of government spending to output is insignificantly different from zero and the significant coefficients imply a rise in government spending in response to higher debt or inflation. The positive reaction of government spending and the negative reaction of taxes to inflation might reflect that, with given nominal rates and persistence in inflation the government financing cost falls with the implied real rate.

5.3.3 Historical shocks

Overall, both models have similar implications for the historical fiscal and monetary policy shocks. Figure 11 compares the time series of identified fiscal shocks and the monetary shock in the VAR to the identified shocks in the estimated DSGE model. The median of the time series is highly correlated, with correlation of 0.68 for tax shocks, 0.69 for government spending shocks and 0.76 for monetary policy shock.\(^{35}\)

Besides a high statistical association of identified shocks across the two identification methods, the recovered shocks line up with historic tax reforms. For example, the median shocks reflect the 1964 tax cuts and the 1969 tax increases as well as the Bush-era tax cuts in the early 2000s.\(^{36}\)

A high correlation between the implied structural shocks is a necessary condition for consistency of the VAR results and DSGE results. However, implicit in the VAR analysis has been the assumption that the instruments are not related to non-policy shocks. I know assess this assumption.

---

\(^{35}\)Note that since the impulse-response-vectors are only identified up to sign, the same is true for the identified shocks. The focus is therefore on the strength, but not the sign, of the correlations.

\(^{36}\)For a list of tax reforms, see [http://www.taxpolicycenter.org/legislation/].
Government spending shock – correlation of medians = 0.69 (0.26,0.74)

Tax shock – correlation of medians = 0.68 (0.18,0.70)

Monetary policy shock – correlation of medians = 0.76 (0.65,0.78)

Note: Shown in gray is the posterior 90% pointwise confidence interval of the shocks from the VAR with flat prior, along with the median values from the separately estimated DSGE model and VAR.

Figure 11: Identified shock comparison
5.3.4 Instrument validity

For most of this paper, I have worked under the assumption that narrative shocks are noisy measures of a particular structural shock. This has allowed inference about the response to these structural shocks. However, what if narrative shock measures are better interpreted as reflecting several structural shocks?

I now test this assumption by comparing the model fit under different assumptions about the narrative instruments, as described in Section 3.2.3: Including the narrative shocks in the equation one at a time, I compute the Bayes factor in favor of the restricted model implicit in the VAR analysis vs. the unrestricted model. Table 1 shows the results. For the narrative policy instruments used in the VAR, the data provides more support for the restricted model consistent with the VAR assumptions than for the unrestricted model. In particular, based on the modified harmonic mean approximation to the posterior data density, the Bayes factor is between +2.8 and +6.3 log points in favor of specifications in which the DSGE model only loads on the intended shock.

<table>
<thead>
<tr>
<th>Narrative shock</th>
<th>log-Bayes factor relative to best model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All shocks</td>
</tr>
<tr>
<td>Money shock, R&amp;R</td>
<td>-5.7</td>
</tr>
<tr>
<td>Money shock, Kuttner</td>
<td>-6.2</td>
</tr>
<tr>
<td>Defense spending ( G )</td>
<td>-2.8</td>
</tr>
<tr>
<td>Personal income tax</td>
<td>-6.9</td>
</tr>
</tbody>
</table>

Table 1: Assessing the restriction of narrative instruments to a single shock

6 Conclusion

A key question for academics and practitioners using quantitative DSGE models is whether these models are consistent with results derived from identification schemes with weaker assumptions. This paper takes a step towards formally assessing this structural fit. The paper makes two methodological contributions: First, it shows that the VAR identification based on narrative instruments correctly identifies policy rules in DSGE models when these are simple Taylor rules. Second, it shows how to estimate a narrative DSGE-VAR using the standard SUR framework. Using a Monte Carlo study I show that, with a flat prior, the proposed Bayesian estimator also captures the uncertainty about the instrument covariance matrix and has, with enough data, accurate frequentist sampling properties in my application.
In terms of the substance, I find that a standard medium-scale DSGE model such as Christiano et al. (2005) and Smets and Wouters (2007) augmented with fiscal Taylor rules matches the narrative VAR well in terms of the path of historical shocks and most impulse-responses. The DSGE model also provides support in favor of the identifying assumption of instrument validity. A caveat is, however, that the VAR results are estimated only imprecisely, so that the rough consistency of the two models may be due to the noise in the narrative instruments. The extended DSGE fails to capture policy interactions and the response of private consumption to government spending.

The estimation of the DSGE-VAR model with a non-degenerate prior over the DSGE model parameters is ongoing work. Given the substantial posterior uncertainty, however, there is scope for using prior information, as is customary in the estimation of DSGE models, as well as in reduced form VARs. As a proof of concept, I show that adding moderate amounts of prior information allows to recover sensible estimates of Taylor rule coefficients even though the results with a flat prior are mostly insignificant.
References


**A Appendix**

**A.1 Narrative Shock identification**

Note that (3.6a) does not restrict $\alpha_{11}$ for any value of $G$, assuming that $\Gamma$ is of maximal column rank $m_z \leq m$. The LHS depends on $m_z^2$ different parameters of $\alpha_{11}$, and $\Gamma_1$ has also $m_z^2$ elements.

$\Sigma$ provides an extra $m \frac{m+1}{2}$ equations, which can be used to solve for the remaining $m_z \leq m$ elements in $\alpha_{11}$ plus the $m \times (m - m_z)$ elements in $\alpha_2$. Thus, in general,
Given that \( \text{Nullspace of the instrument-identified shocks } \mathbf{v} \text{ for } \mathbf{Z} \) Note that: any vector in the nullspace of \( \mathbf{I}_{m_z} \) satisfies the orthogonality condition.}

Define

\[ \kappa = (\Gamma_1^{-1} \Gamma_2)', \quad (A.1) \]

so that \( \alpha_{21} = \kappa \alpha_{11} \). Then:

\[ \Sigma = \begin{bmatrix} \alpha_{11} \alpha_{11}' + \alpha_{12} \alpha_{12}' & \alpha_{11} \alpha_{11}' \kappa' + \alpha_{12} \alpha_{22}' \\ \kappa \alpha_{11} \alpha_{11}' + \alpha_{22} \alpha_{12}' & \kappa \alpha_{11} \alpha_{11}' \kappa' + \alpha_{22} \alpha_{22}' \end{bmatrix} \quad (A.2) \]

The covariance restriction identifies the impulse response (or component of the forecast error) up to a \( m_z \times m_z \) square scale matrix \( \alpha_{11} \):

\[ u_t = A \epsilon_t = [\alpha^{[1]} \alpha^{[2]}] \epsilon_t = [\alpha^{[1]} \epsilon_t^{[1]}] + [\alpha^{[2]} \epsilon_t^{[2]}] = \begin{bmatrix} \mathbf{I}_{m_z} \\ \kappa \end{bmatrix} \alpha_{11} \epsilon_t^{[1]} + \begin{bmatrix} \alpha_{12} \\ \alpha_{22} \end{bmatrix} \epsilon_t^{[2]} \]

Given that \( \epsilon_t^{[1]} \perp \epsilon_t^{[2]} \) it follows that:

\[ \text{Var}[u_t | \epsilon_t^{[1]}] = \alpha^{[2]}(\alpha^{[2]})' = \begin{bmatrix} \alpha_{12} \alpha_{12}' & \alpha_{12} \alpha_{22}' \\ \alpha_{22} \alpha_{12}' & \alpha_{22} \alpha_{22}' \end{bmatrix} \]

\[ \text{Var}[u_t | \epsilon_t^{[2]}] = \alpha^{[1]}(\alpha^{[1]})' = \begin{bmatrix} \alpha_{11} \alpha_{11}' & \alpha_{11} \alpha_{11}' \kappa' \\ \kappa \alpha_{11} \alpha_{11}' & \kappa \alpha_{11} \alpha_{11}' \kappa' \end{bmatrix} \]

\[ \Sigma = \text{Var}[u_t] = \text{Var}[u_t | \epsilon_t^{[2]}] + \text{Var}[u_t | \epsilon_t^{[1]}] = \begin{bmatrix} \Sigma_{12} \Sigma_{12}' & \Sigma_{12} \Sigma_{22}' \\ \Sigma_{22} \Sigma_{12}' & \Sigma_{22} \Sigma_{22}' \end{bmatrix} \]

Note that:

\[ u_t^{\text{res}} = u_t - \mathbb{E}[u_t | \epsilon_t^{[1]}] \perp \mathbb{E}[u_t | \epsilon_t^{[1]}] = \begin{bmatrix} \mathbf{I}_{m_z} \\ \kappa \end{bmatrix} \alpha_{11} \epsilon_t^{[1]} \]

Any vector in the nullspace of \( \begin{bmatrix} \mathbf{I}_{m_z} \\ \kappa' \end{bmatrix} \) satisfies the orthogonality condition. Note that \( \begin{bmatrix} \mathbf{I}_{m_z} \\ \kappa' \end{bmatrix} \) is an orthogonal basis for \( \mathbb{R}^m \).

Define

\[ Z \equiv [Z^{[1]} \ Z^{[2]}] \equiv \begin{bmatrix} \mathbf{I}_{m_z} \\ \kappa \end{bmatrix} \begin{bmatrix} \kappa' \\ -\mathbf{I}_{m-z} \end{bmatrix} \quad (A.3) \]

Note that \( Z^{[2]} \) spans the Nullspace of \( \alpha^{[1]} \). Hence, \( (Z^{[2]})' v_t \) projects \( v_t \) into the Nullspace of the instrument-identified shocks \( \epsilon_t^{[1]} \).

\[ (Z^{[2]})' v_t = (Z^{[2]})' A \epsilon_t = (Z^{[2]})' [\alpha^{[1]} \alpha^{[2]}] \epsilon_t \]

\[ \text{for } m \geq 1, m \geq m_z \geq 0. \]

\[ \frac{m(m-1)}{2} - m_z(m-m_z) = \frac{1}{2}((m - (m_z + 0.5))^2 + m_z^2 - (0.25 + m_z)) = \frac{1}{2}m(m - 2k - 1) + m_z^2 \geq 0 \]
\[ (Z^{[2]})' [Z^{[1]} || \tilde{Z} || \alpha^{11} \alpha^{[2]}] \epsilon_t = [0 (Z^{[2]})' \alpha^{[2]}] \epsilon_t \]
\[ = 0 \times \epsilon_t^{[1]} + (Z^{[2]})' \alpha^{[2]} \epsilon_t \perp \epsilon^{[1]} \]

Note that \((Z^{[2]})' \alpha^{[2]}\) is of full rank and I can therefore equivalently consider \(\epsilon_t^{[2]}\) or \((Z^{[2]})' v_t\). Thus, the expectation of \(v_t\) given \(\epsilon_t^{[2]}\) is given by:

\[ \mathbb{E}[v_t | \epsilon_t^{[2]}] = \text{Cov}[v_t, (Z^{[2]})' v_t] \text{Var}((Z^{[2]})' v_t)^{-1} (Z^{[2]})' v_t, \]
\[ v_t - \mathbb{E}[v_t | \epsilon_t^{[2]}] = (I - \text{Cov}[v_t, (Z^{[2]})' v_t] \text{Var}((Z^{[2]})' v_t)^{-1} (Z^{[2]})' v_t, \]
\[ \text{Cov}[v_t, (Z^{[2]})' v_t] = \Sigma Z^{[2]} = \Sigma \left[ -\kappa' \right] \]
\[ \text{Var}[v_t | \epsilon_t^{[2]}] = \mathbb{E}[(I - \text{Cov}[v_t, (Z^{[2]})' v_t] \text{Var}((Z^{[2]})' v_t)^{-1} (Z^{[2]})' v_t) v_t'] \]
\[ = \mathbb{E}[v_t v_t'] - \text{Cov}[v_t, (Z^{[2]})' v_t] \text{Var}((Z^{[2]})' v_t)^{-1} \mathbb{E}[(Z^{[2]})' v_t v_t'] \]
\[ = \Sigma - \text{Cov}[v_t, (Z^{[2]})' v_t] \text{Var}((Z^{[2]})' v_t)^{-1} \text{Cov}[v_t, (Z^{[2]})' v_t] \]
\[ = \Sigma - \Sigma \left[ -\kappa' \right] \left[ \kappa - I_{m-m_z} \right] \Sigma \left[ -\kappa' \right] \Sigma \]
\[ = \left[ \begin{array}{cc} \alpha_{11} \alpha_1' \\
\kappa \alpha_1 \alpha_1' \end{array} \right] \]
\[ \kappa \alpha_1 \alpha_1' \]

This gives a solution for \(\alpha_{11} \alpha_1'\) in terms of observables: \(\Sigma\) and \(\kappa = \Gamma_1^{-1} \times \Gamma_2\). For future reference, note that this also implies that:

\[ \text{Var}[v_t | \epsilon_t^{[1]}] = \Sigma - \text{Var}[v_t | \epsilon_t^{[2]}] \]
\[ = \Sigma \left[ -\kappa' \right] \left[ \kappa - I_{m-m_z} \right] \Sigma \left[ -\kappa' \right] \Sigma \]
\[ = \left[ \begin{array}{cc} I_{m_z} \\
\kappa \end{array} \right] \alpha_{11} \]

In general, \(\alpha_{11}\) itself is unidentified: An additional \(\frac{(m_z-1)m_z}{2}\) restrictions are needed to pin down its \(m_z^2\) elements from the \(\frac{(m_z+1)m_z}{2}\) independent elements in \(\alpha_{11} \alpha_1'\). Given \(\alpha_{11}\), the impact-response to a unit shock is given by:

\[ \left[ \begin{array}{c} I_{m_z} \\
\kappa \end{array} \right] \alpha_{11} \]

Note that this leaves \(m(m-m_z)\) elements in \(\alpha^{[2]}\) unrestricted, for which there exist \(\frac{m(m+1)}{2}\) equations in \(\Sigma - \text{Var}[v_t | \epsilon_t^{[1]}]\), requiring an extra \(\frac{m}{2}(m - 1 + 2k)\) restrictions. (Note: if \(m_z \geq \frac{1}{2}(m - 1)\), this implies over-identification.)

Intuitively: \(\alpha^{[1]}\) has \(km\) parameters. There are \((m - m_z) \times m_z\) covariances with other structural shocks \((\Sigma_{12})\) which help to identify the model, leaving \(m_z^2\) parameters undetermined. An extra \(\frac{m_z(m_z+1)}{2}\) restrictions comes from the Riccatti
equation via the residual variance, leaving \( \frac{m_x(m_x-1)}{2} \) parameters to be determined. (This should also work when identifying one shock at a time with instruments.) The problem with multiple instrumented shocks is that even if I know that each instrument is only relevant for one specific structural shock, they are silent on the covariance between shocks – the covariance between shocks could be contained in the first stage residual.

When each instruments exactly identifies one shock, then each instrument delivers \((n - 1)\) identifying restrictions: Each instrument then identifies one column of \( A \) up to scale. To see this, consider (3.5) for the case of \( G = \text{diag}([g_1, \ldots, g_{m_x}, 0, \ldots, 0]) \) and let \( \alpha^{[i]} \) now denote the \( i \)’th column of \( A \):

\[
\text{Cov}[z^{[i]}_t, v_t] = e_t G A' = g_i e_t e'_t A' = g_i (\alpha^{[i]})'.
\]

(A.6)

Since the LHS is observable, (??) identifies \( \alpha^{[i]} \) up to scale, imposing \( n - 1 \) restrictions on \( A \) for each instrument. In general, this leads to overidentifying restrictions on \( \Sigma \).

The (normalized) identified shocks are given by \( Fu \), where \( F = \Gamma \Sigma^{-1} \).

A.2 Narrative Policy Rule Identification

To show that the lower Choleski factorization proposed in Mertens and Ravn (2013) identifies Taylor-type policy rules when ordered first, I start by deriving the representation of the identification problem as the simultaneous equation system (3.10). Recall the definition of forecast errors \( v_t \) in terms of structural shocks \( \epsilon_t \):

\[
v_t = A \epsilon_t = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}^{-1} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} = v_t \quad (A.7)
\]

Note that:

\[
\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} & -\alpha_{11}^{-1} \alpha_{12} (\alpha_{22} - \alpha_{12} \alpha_{21}^{-1} \alpha_{11} \alpha_{12}^{-1})^{-1} \\ -\alpha_{21}^{-1} (\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} & (\alpha_{22} - \alpha_{21}^{-1} \alpha_{11} \alpha_{12})^{-1} \end{bmatrix}
\]

\[
= \begin{bmatrix} (\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} & -\alpha_{11}^{-1} (\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} (\alpha_{22} - \alpha_{21}^{-1} \alpha_{12}^{-1})^{-1} \\ (\alpha_{22} - \alpha_{21}^{-1} \alpha_{12})^{-1} & (\alpha_{22} - \alpha_{21}^{-1} \alpha_{12})^{-1} \end{bmatrix}
\]

Note that:

\[
(\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} = \alpha_{11}^{-1} ((\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21}) \alpha_{11}^{-1})^{-1} = \alpha_{11}^{-1} (I - \alpha_{12} \alpha_{22}^{-1} \alpha_{21} \alpha_{11}^{-1})^{-1}
\]

and define:

\[
S_1 \equiv (I - \alpha_{12} \alpha_{22}^{-1} \alpha_{21} \alpha_{11}^{-1}) \alpha_{11} \quad S_2 \equiv (I - \alpha_{21} \alpha_{11}^{-1} \alpha_{12} \alpha_{22}^{-1}) \alpha_{22}
\]

(A.8)

so that:

\[
(\alpha_{11} - \alpha_{12} \alpha_{22}^{-1} \alpha_{21})^{-1} = S_1^{-1} \quad (\alpha_{22} - \alpha_{21} \alpha_{11}^{-1} \alpha_{12})^{-1} = S_2^{-1}
\]
Then:
\[
\begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix}^{-1} v_t = \begin{bmatrix}
S_1^{-1} & -\alpha_{12}A_{22}^{-1} \\
-\alpha_{21}A_{11}^{-1} & S_2^{-1}
\end{bmatrix} v_t = \begin{bmatrix}
S_1^{-1} & 0 \\
0 & S_2^{-1}
\end{bmatrix} \begin{bmatrix} I \\ -\alpha_{21}A_{11}^{-1} I \end{bmatrix} v_t = \begin{bmatrix}
\epsilon_{1,t} \\
\epsilon_{2,t}
\end{bmatrix}
\]
and equivalently:
\[
\begin{bmatrix}
I & -\eta \\
-\kappa & I
\end{bmatrix} v_t = \begin{bmatrix}
S_1 \\
0
\end{bmatrix} \begin{bmatrix}
\epsilon_{1,t} \\
\epsilon_{2,t}
\end{bmatrix}
\]

(A.9)
defining \( \eta \equiv \alpha_{12}A_{22}^{-1} \) and \( \kappa \equiv \alpha_{21}A_{11}^{-1} \). Equation (3.10) follows immediately.

**Lemma 3** (Mertens and Ravn (2013)). Let \( \Sigma = AA' \) and \( \Gamma = \begin{bmatrix} G & 0 \end{bmatrix} A \), where \( G \) is an \( m_z \times m_z \) invertible matrix and \( A \) is of full rank. Then \( \alpha^{[1]} \) is identified up to a factorization of \( S_1S_1' \) with \( S_1 \) defined in (A.8).

**Proof.** Since \( A \) is of full rank, it is invertible and (A.9) holds for any such \( A \). Given \( \eta, \kappa \), (A.9) implies (3.11), which I reproduce here for convenience:

\[
\alpha^{[1]} = \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} (I - \eta\kappa)^{-1} \\ (I - \kappa\eta)^{-1}\kappa \end{bmatrix} \text{chol}(S_1S_1').
\]

(3.11)

If \( \Sigma \) and \( \Gamma \) pin down \( \eta, \kappa \) uniquely, \( \alpha^{[1]} \) is uniquely identified except for a factorization of \( S_1S_1' \).

To show that \( \Sigma \) and \( \Gamma \) pin down \( \eta, \kappa \) uniquely, consider \( \kappa \) first. Since \( \Gamma = \begin{bmatrix} G & 0 \end{bmatrix} A \) and \( G \) is an \( m_z \times m_z \) invertible matrix, it follows that Assumption 1 holds. It then follows from (3.6b) that \( \kappa = \alpha_{21}A_{11}^{-1} = \Gamma^2\Gamma_1^{-1}. \)

To compute \( \eta \), more algebra is needed. Partition \( \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\
\Sigma'_{12} & \Sigma_{22}\end{bmatrix} \), where \( \Sigma_{11} \) is \( m_z \times m_z \), \( \Sigma_{12} \) is \( m_z \times (m - m_z) \) and \( \Sigma_{22} \) is \( (m - m_z) \times (m - m_z) \). Define

\[
\alpha_{22}a_{22}' = \Sigma_{22} - \kappa\alpha_{11}a_{11}'\kappa' = \Sigma_{22} - \kappa(\Sigma_{11} - \alpha_{12}a_{12}')\kappa',
\]

using (A.2) twice. Using the upper left element of (A.5), it follows that

\[
\alpha_{12}a_{12}' = (\Sigma_{12}' - \kappa\Sigma_{11})(ZZ'\Sigma_{12}' - \kappa\Sigma_{11})
\]

with

\[
ZZ' = \kappa\Sigma_{11} - (\Sigma_{12}' + \kappa\Sigma_{12}) + \Sigma_{22} = [\kappa - I_{m - m_z}] \Sigma \begin{bmatrix} -\kappa' \\ -I_{m - m_z} \end{bmatrix}.
\]

The coefficient matrix of interest, \( \eta \), is then defined as:

\[
\eta = \alpha_{12}a_{22}^{-1} = \alpha_{12}a_{22}'(\alpha_{22}a_{22}')^{-1} = (\Sigma_{12}' - \kappa\alpha_{11}a_{11}')'(\alpha_{22}a_{22}')^{-1} = (\Sigma_{12}' - \kappa\Sigma_{11} + \kappa\alpha_{12}a_{12}')'(\alpha_{22}a_{22}')^{-1}.
\]
Thus, $\eta$ and $\kappa$ are uniquely identified given $\Sigma, \Gamma$.

The above derivations link $S_1$ to $A^{-1}$. I now compute $S_1$ for a class of models.

**Proposition 1.** Assume $\Sigma = AA' = A^*(A^*)'$ and order the policy variables such that the $m_p = m_z$ or $m_p = m_z - 1$ simple Taylor rules are ordered first and $\Gamma = [G, 0]A^*$. Then $\alpha^{[1]}$ defined in (3.11) satisfies $\alpha^{[1]} = A^*[I_{m_z}, 0_{(m-m_z)\times(m-m_z)}]'$ up to a normalization of signs on the diagonal if

(a) $m_z$ instruments jointly identify shocks to $m_p = m_z$ simple Taylor rules w.r.t. the economy (2.1), or

(b) $m_z$ instruments jointly identify shocks to $m_p = m_z - 1$ simple Taylor rules w.r.t. the economy (2.1) and $\psi_{p,m_z} = 0, p = 1, \ldots, m_p$.

**Proof.** Given Lemma 3, $\alpha^{[1]}$ is identified uniquely if $S_1$ is identified uniquely. In what follows, I establish that under the ordering in the proposition, $S_1$, as defined in (A.8) for arbitrary full rank $A$, is unique up to a normalization. It then follows that $\alpha^{[1]}$ is identified uniquely and hence equal to $A^*[I_{m_z}, 0_{(m-m_z)\times(m-m_z)}]'$.

To proceed, stack the $m_p$ policy rules:

$$y^p_t = \sum_{i=m_p+1}^m \begin{bmatrix} \psi_{1,i} & 0 & \ldots & 0 \\ 0 & \psi_{2,i} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \psi_{m_p,i} \end{bmatrix} y_{i,t} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m_p} \end{bmatrix} X_{t-1} + \begin{bmatrix} \sigma_{11} & 0 & \ldots & 0 \\ \sigma_{21} & \sigma_{22} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{n_p,1} & \sigma_{n_p,2} & \ldots & \sigma_{n_p,m_p} \end{bmatrix} \epsilon^p_t$$

$$= \sum_{i=m_p+1}^m \Psi_i y_{i,t} + \Lambda X_{t-1} + [D_0 \ 0] \epsilon_t,$$

$$= \left( [D_0 \ 0] \epsilon_t + \sum_{i=m_p+1}^m \Psi_i 1A_i' \right) \epsilon_t + \left( \sum_{i=m_p+1}^m \Psi_i 1B_i'X_{t-1} + \Lambda \right) X_{t-1},$$

where $m - m_p \leq \bar{n} \equiv \max_p n_p$. Define $\epsilon_i$ as the selection vector of zeros except for a one at its $i$th position and denote the $i$th row of matrix $A$ by $A_i = (\epsilon_i'A)$ and similarly for $B_i$.

Without loss of generality, order the policy instruments first, before the $m - m_p = \bar{n}$ non-policy variables. Then $A^*$ in the DSGE model observation equation (2.1a) can be written as:

$$\begin{bmatrix} [D_0 + \sum_{i=m_p}^m D_i 1(A_i^*)'] \\ (A_{m_p+1}^*)' \\ \vdots \\ (A_{m_p}^*)' \end{bmatrix},$$

where $D_0$ is a full-rank diagonal matrix and the $D_j$ matrices are $m_p \times m_p$ matrices.
To find \((A^*)^{-1}\), proceed by Gauss-Jordan elimination re-write the system \(A^*X = I_m\), with solution \(X = (A^*)^{-1}\), as \([A^*|I_m]\). Define \(E\) as a conformable matrix such that \([A^*|I_m] \xrightarrow{E} [B|C] = [EA^*|EI_m]\). Then:

\[
[(A^*)|I_m] = \begin{bmatrix}
[D_0 \ 0] + \sum_{i=m_p+1}^m D_i 1(A_i^*)' & I_{mp} & 0 & 0 & \ldots & 0 \\
(A_{mp+1}') & 0' & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
(A_m') & 0' & 0 & 0 & \ldots & 1 \\
\end{bmatrix}
\]

\[
E_1 \xrightarrow{\cdots} \begin{bmatrix}
[D_0 \ 0] + \sum_{i=m_p+2}^m D_i 1(A_i^*)' & I_{mp} & -D_{mp+1} & 1 & 0 & \ldots & 0 \\
(A_{mp+1}') & 0' & 1 & 0 & \ldots & 0 \\
(A_{mp+2}') & 0' & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
(A_m') & 0' & 0 & 0 & \ldots & 1 \\
\end{bmatrix}
\]

\[
E_2 \xrightarrow{\cdots} \begin{bmatrix}
[D_0 \ 0] + \sum_{i=m_p+3}^m D_i 1(A_i^*)' & I_{mp} & -D_{mp+1} & -D_{mp+2} & 1 & \ldots & 0 \\
(A_{mp+1}') & 0' & 1 & 0 & \ldots & 0 \\
(A_{mp+2}') & 0' & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
(A_m') & 0' & 0 & 0 & \ldots & 1 \\
\end{bmatrix}
\]

Thus \(((A^*)^{-1})_{1:m_p,1:m_p} = (E_D E_{m-m_p} \ldots E_1 I_m)_{1:m_p,1:m_p}\).

Now consider cases (a) and (b):

(a) \(m_z = m_p\). From (A.8), \(S_1\) is the upper left corner of \((A^*)^{-1}\):

\[
S_1 \equiv ((A^*)^{-1})_{1:m_p,1:m_p} = D_0^{-1}
\]

and \(S_1\) is a diagonal.

(b) \(m_z = m_p + 1, \psi_{p,m_p+1} = 0, p = 1, \ldots, m_p\). The second condition implies that
\[ D_{m_p+1} = 0_{m_p \times m_p}. \] It follows that \( S_1 \) defined in (A.8) is given by:

\[
S_1 \equiv ( (A^*)^{-1} )_{1:m_p+1,1:m_p+1} = \begin{bmatrix} D_0^{-1} & D_{m_p+1} \\ s_{m_p+1,1:m_p} & s_{m_p+1,m_p+1} \\ \end{bmatrix} = \begin{bmatrix} D_0^{-1} & 0 \\ s_{m_p+1,1:m_p} & s_{m_p+1,m_p+1} \\ \end{bmatrix}
\]

Thus \( S_1 \) is lower triangular.

In both cases, \( S_1 \) is lower triangular. Since the lower Choleski decomposition is unique, a Choleski decomposition of \( S_1 S_1' \) recovers \( S_1 \) if we normalize signs of the diagonal of \( S_1 \) to be positive. Given identification of \( S_1 \), the identification of \( \alpha^{[1]} \) follows from Lemma 3.

### A.3 Priors and posteriors

Let \( u_t \overset{iid}{\sim} \mathcal{N}(0, V) \) and let \( U = [u_1, \ldots, u_T]' \), where \( u_t \) is \( m \times 1 \) and \( U \) is \( T \times m \). Then the likelihood can be written as:

\[
L = (2\pi)^{-mT/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \sum_{t=1}^{T} u_t' V^{-1} u_t \right) = (2\pi)^{-mT/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \text{tr}(u_t' V^{-1} u_t) \right) = (2\pi)^{-mT/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \text{tr}(V^{-1} \sum_{t=1}^{T} u_t u_t') \right) = (2\pi)^{-mT/2} |V|^{-T/2} \exp \left( -\frac{1}{2} \text{vec}(U)' (V^{-1} \otimes I_T) \text{vec}(U) \right)
\]

using that \( \text{tr}(ABC) = \text{vec}(B') (A' \otimes I) \text{vec}(C) \) and that \( V = V' \).

For the SUR model, \( [Y, Z] = [X_y, X_z] \begin{bmatrix} B_y \\ B_z \end{bmatrix} + U \). Consequently, \( Y_{SUR} \equiv \text{vec}([Y, Z]) = X_{SUR} \text{vec} \left( \begin{bmatrix} B_y \\ B_z \end{bmatrix} \right) + \text{vec}(U) \), where

\[
X_{SUR} \equiv \begin{bmatrix} I_m \otimes X_y & 0 \\ 0 & I_{m_z} \otimes X_z \end{bmatrix}.
\]

The likelihood can then also be written as:

\[
L = (2\pi)^{-mT/2} |V|^{-T/2} \exp \left( -\frac{1}{2} (Y_{SUR} - X_{SUR} \beta)' (V^{-1} \otimes I_T) (Y_{SUR} - X_{SUR} \beta) \right)
\]

51
\[
\begin{align*}
&= (2\pi)^{-mT/2}|V|^{-T/2} \exp \left( -\frac{1}{2}(\bar{Y}_{SUR} - \bar{X}_{SUR}\beta)'(\bar{Y}_{SUR} - \bar{X}_{SUR}\beta) \right) \\
&= (2\pi)^{-mT/2}|V|^{-T/2} \exp \left( -\frac{1}{2}(\bar{Y}_{SUR} - \bar{X}_{SUR}\beta)'(\bar{Y}_{SUR} - \bar{X}_{SUR}\beta) \right) \\
&= (2\pi)^{-mT/2}|V|^{-T/2} \exp \left( -\frac{1}{2}(\bar{X}_{SUR}(\beta - \bar{\beta}_{SUR})'(\bar{X}_{SUR}(\beta - \bar{\beta}_{SUR})) \right) \\
&= (2\pi)^{-mT/2}|V|^{-T/2} \exp \left( -\frac{1}{2}(\beta - \bar{\beta}_{SUR})'(\bar{X}_{SUR}'\bar{X}_{SUR})(\beta - \bar{\beta}_{SUR}) \right),
\end{align*}
\]

where \(\bar{\beta}_{SUR} \equiv (\bar{X}_{SUR}'\bar{X}_{SUR})^{-1}\bar{X}_{SUR}'\bar{Y}_{SUR}\) and the second to last equality follows from the normal equations.

Note that expression (A.11) for the likelihood is proportional to a conditional Wishart distribution for \(\beta|V^{-1} \sim \mathcal{N}(\bar{\beta}_{SUR}, (\bar{X}_{SUR}'\bar{X}_{SUR})^{-1}) \equiv \mathcal{N}(\bar{\beta}_{SUR}, (X_{SUR}'V^{-1} \otimes I)X_{SUR})^{-1}\). Alternatively, expression (A.10) for the likelihood is proportional to a conditional Wishart distribution for \(V^{-1}: V^{-1}|\beta \sim \mathcal{W}_{m_a}((U(\beta)'U(\beta))^{-1}, T + m_a + 1)\). Pre-multiplying with a Jeffreys’s prior over \(V\), transformed to \(V^{-1}\) is equivalent to premultiplying by \(\pi(V^{-1}) \equiv |V^{-1}|^{-\frac{m_a+1}{2}}\) and yields:

\[
\pi(V^{-1}) \times = |V^{-1}|^{-\frac{m_a+1}{2}} \times (2\pi)^{-mT/2}|V|^{-T/2} \exp \left( -\frac{1}{2} \text{tr}(V^{-1}U'U) \right)
\]

\[
= (2\pi)^{-mT/2}|V^{-1}|^{(T-m_a-1)/2} \exp \left( -\frac{1}{2} \text{tr}(V^{-1}U'U) \right),
\]

which is \(V^{-1}|\beta \sim \mathcal{W}_{m_a}((SSR(\beta))^{-1}, T)\), with

\[
SSR(\beta) \equiv U(\beta)'U(\beta) = [Y - X_yB_y(\beta), Z - X_zB_z(\beta)]'[Y - X_yB_y(\beta), Z - X_zB_z(\beta)]
\]

\[
= \sum_{t=1}^{T} [y_t - x_{y,t}B_y(\beta), z_t - x_{z,t}B_z(\beta)]'[y_t - x_{y,t}B_y(\beta), z_t - x_{z,t}B_z(\beta)].
\]

Consider the case of the dummy variable prior:

\[
SSR(\beta) = \left( \sqrt{T_0^B} \begin{bmatrix} B_y(\theta) \\ B_z(\theta) \end{bmatrix} \right)' \left( \sqrt{T_0^B} \begin{bmatrix} I \\ 0 \end{bmatrix} \right) \begin{bmatrix} B_y \\ B_z \end{bmatrix}
\]

\[
\times \left( \sqrt{T_0^B} \begin{bmatrix} B_y(\theta) \\ B_z(\theta) \end{bmatrix} \right)' \left( \sqrt{T_0^B} \begin{bmatrix} I \\ 0 \end{bmatrix} \right) \begin{bmatrix} B_y \\ B_z \end{bmatrix}
\]

\[
= T_0^B \begin{bmatrix} B_y(\theta) - B_y \\ B_z(\theta) - B_z \end{bmatrix}' \begin{bmatrix} B_y(\theta) - B_y \\ B_z(\theta) - B_z \end{bmatrix} + T_0^V V_0
\]

52
A.4 Frequentist inference

Following Mertens and Ravn (2013), the bootstrap procedure I consider is characterized as follows:

1. For \( t = 1, \ldots, T \), draw \( \{e^b_1, \ldots, e^b_T\} \), where \( e^b_t \sim iid \) with \( \Pr\{e^b_t = 1\} = \Pr\{e^b_t = -1\} = 0.5 \).

2. Construct the artificial data for \( Y^b_t \). In a VAR of lag length \( p \), build \( Y^b_t, t > p \) as:
   - For \( t = 1, \ldots, p \) set \( Y^b_t = Y_t \).
   - For \( t = p + 1, \ldots, T \) construct recursively \( Y^b_t = \sum_{j=1}^{p} \hat{B}_j Y^b_{t-j} + e^b_1 u_t \).

3. Construct the artificial data for the narrative instrument:
   - For \( t = 1, \ldots, T \) construct recursively \( z^b_t = e^b_t z_t \).

A.5 DSGE model equations

A.5.1 Households

The law of motion for capital:

\[
\hat{k}^p_t = (1 - \bar{x} \hat{k}^p_{t-1}) + \bar{x} (\hat{x}_t + \hat{q}_{t+s}) \tag{A.14}
\]

Household wage setting:

\[
\hat{w}_t = \frac{\hat{w}_{t-1}}{1 + \beta} + \frac{\beta \gamma \mathbb{E}_t[\hat{w}_{t+1}]}{1 + \beta} + \frac{(1 - \beta \zeta w)(1 - \zeta w)}{(1 + \beta \gamma) \zeta w} \hat{A}_w \left( \hat{c}_t - (h/\gamma) \hat{c}_{t-1} \right) + \nu \hat{n}_t - \hat{w}_t + \frac{d \tau^n}{1 - \tau^n} + \frac{d \tau^c}{1 + \tau^c}
\]

\[
= 1 + \beta \mu_w \hat{\pi}_t + \frac{\hat{\pi}_t}{1 + \beta \gamma} \hat{\pi}_{t-1} + \frac{\beta \gamma}{1 + \beta \gamma} \mathbb{E}_t[\hat{\pi}_{t+1}] + \frac{\hat{c}_t^{\lambda w}}{1 + \beta \gamma} \tag{A.15}
\]

Household consumption Euler equation:

\[
\mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] + \mathbb{E}_t[d \tau^{c}_{t+1} - d \tau^c] = \frac{1}{1 - h/\gamma} \mathbb{E}_t[(\sigma - 1) \frac{1 - \tau^n}{1 + \lambda w} \hat{n}_{t+1} - \hat{n}_t] - \sigma \left( \hat{c}_{t+1} - \left( 1 + \frac{h}{\gamma} \right) c_t + \frac{h}{\gamma} \hat{c}_{t+1} \right), \tag{A.16}
\]
Other FOC (before rescaling of $\hat{q}_t^b$):

\[
\mathbb{E}_t[\hat{\xi}_{t+1} - \hat{\xi}_t] = -\hat{q}_t^b - \hat{R}_t + \mathbb{E}_t[\hat{\pi}_{t+1}],
\]

(A.17)

\[
\hat{Q}_t = -\hat{q}_t^b - (\hat{R}_t - \mathbb{E}_t[\pi_{t+1}]) + \frac{1}{\bar{r}_k(1 - \tau^k) + \delta\tau^k + 1 - \delta} \times \left[ (\tau^k (1 - \tau^k) + \delta\tau^k)\hat{q}_t - (\tau^k - \delta)d\tau^k_{t+1} + \right.
\]

\[
+ \tau^k (1 - \tau^k)\mathbb{E}_t(r_k^k + (1 - \delta)\mathbb{E}_t(\hat{Q}_{t+1}) \right],
\]

(A.18)

\[
\hat{x}_t = \frac{1}{1 + \beta\gamma} \left[ \hat{x}_{t-1} + \beta\gamma\mathbb{E}_t(\hat{x}_{t+1}) + \frac{1}{\gamma^2 S''(\gamma)}[\hat{Q}_t + \hat{q}_t^x] \right],
\]

(A.19)

\[
\hat{u}_t = \frac{a'(1)}{a''(1)}\hat{r}_t^k \equiv \frac{1 - \psi_u}{\psi_u}\hat{r}_t^k.
\]

(A.20)

### A.5.2 Production side and price setting

The linearized aggregate production function is:

\[
\hat{y}_t = \bar{y} + \phi(\hat{\epsilon}_t^s + \hat{\zeta}k^g_{t-1} + \alpha(1 - \zeta)\hat{k}_t + (1 - \alpha)(1 - \zeta)\hat{n}_t),
\]

(A.22)

where $\phi$ are fixed costs. Fixed costs, in steady state, equal the profits made by intermediate producers.

The capital-labor ratio:

\[
\hat{k}_t = \hat{n}_t + \hat{w}_t - \hat{r}_t^k.
\]

(A.23)

Price setting:

\[
\hat{\pi}_t = \frac{1}{1 + t_p\beta\gamma} \hat{\pi}_{t-1} + \frac{1 - \zeta_p\beta\gamma}{1 + t_p\beta\gamma} A_p(\hat{m}_c^t + \hat{\epsilon}_t^p) + \frac{\beta\gamma}{1 + t_p\beta\gamma} \mathbb{E}_t\hat{\pi}_{t+1}.
\]

(A.24)

### A.5.3 Market clearing

Goods market clearing requires:

\[
\hat{y}_t = \frac{\bar{y}}{\bar{y}}\hat{c}_t + \frac{x}{\bar{y}}\tilde{\pi}_t + \frac{\bar{x}^g}{\bar{y}}\hat{\pi}_t + \hat{g}_t + \frac{\bar{r}^k}{\bar{y}}\hat{u}_t.
\]

(A.25)

### A.5.4 Observation equations

The observation equations are given by (??) as well as the following seven observation equations from Smets and Wouters (2007) and three additional equations
By allowing for different trends in the non-stationary observables I treat the data symmetrically in the VAR and the DSGE model.

I use the deviation of debt to GDP and revenue to GDP, detrended prior to the estimation, as observables:

\[
\begin{align*}
\hat{b}_{\text{obs}}^t &= \bar{b} \left( \hat{b} - \bar{y} \right) + \bar{b}_{\text{obs}}^t, \\
\hat{\text{rev}}^n_{\text{obs}}^t &= \bar{\tau} n \frac{\bar{\tau}}{\bar{y}} \left( \frac{d\tau^n_t}{\bar{\tau}} + \hat{\tau}_t + \tilde{\bar{\tau}}_t - \bar{y}_t \right) + \text{rev}^n_{\text{obs}}^t, \\
\hat{\text{rev}}^k_{\text{obs}}^t &= \bar{\tau} k \left( \frac{\bar{\tau}}{\bar{y}} \left( i^k_\delta - \delta \right) \left( \frac{d\tau^k_t}{\bar{\tau}} + \frac{\bar{\tau}}{\bar{\tau}_k - \bar{\tau}} i_\delta^k + \hat{i}_{t-1}^p - \bar{y}_t \right) + \text{rev}^k_{\text{obs}}^t.
\end{align*}
\]

### A.6 Data construction

I follow Smets and Wouters (2007) in constructing the variables of the baseline model, except for allocating durable consumption goods to investment rather than consumption expenditure. Specifically:

\[
\begin{align*}
y_t &= \left( \text{nominal GDP: NIPA Table 1.1.5Q, Line 1} \right)_t \\
&\quad \times \left( \text{Population above 16: FRED CNP16OV} \right)_t \\
&\quad \times \left( \text{GDP deflator: NIPA Table 1.1.9Q, Line 1} \right)_t, \\
c_t &= \left( \text{nominal PCE on nondurables and services: NIPA Table 1.1.5Q, Lines 5+6} \right)_t \\
&\quad \times \left( \text{Population above 16: FRED CNP16OV} \right)_t \\
&\quad \times \left( \text{GDP deflator: NIPA Table 1.1.9Q, Line 1} \right)_t, \\
i_t &= \left( \text{Durables PCE and fixed investment: NIPA Table 1.1.5Q, Lines 4 + 8} \right)_t \\
&\quad \times \left( \text{Population above 16: FRED CNP16OV} \right)_t \\
&\quad \times \left( \text{GDP deflator: NIPA Table 1.1.9Q, Line 1} \right)_t, \\
\pi_t &= \Delta \ln(\text{GDP deflator: NIPA Table 1.1.9Q, Line 1})_t, \\
r_t &= \begin{cases} \\
\frac{1}{4}(\text{Effective Federal Funds Rate: FRED FEDFUNDS})_t & t \geq (1954:Q3) \\
\frac{1}{4}(3\text{-Month Treasury Bill: FRED TB3MS})_t & \text{(else.)}
\end{cases}
\end{align*}
\]
\[ n_t = \frac{\text{(Nonfarm business hours worked: BLS PRS85006033)}_t}{\text{(Population above 16: FRED CNP16OV)}_t} \]
\[ w_t = \frac{\text{(Nonfarm business hourly compensation: BLS PRS85006103)}_t}{\text{(GDP deflator: NIPA Table 1.1.9Q, Line 1)}_t} \]

When using an alternative definition of hours worked from Francis and Ramey (2009), I compute:
\[ n_{t}^{FR} = \frac{\text{(Total hours worked: Francis and Ramey (2009))}_t}{\text{(Population above 16: FRED CNP16OV)}_t} \]

Fiscal data is computed following Leeper et al. (2010), except for adding state and local governments (superscript “s&l”) to the federal government account (superscript “f”), similar to Fernandez-Villaverde et al. (2011). Since in the real world

\[ \tau_c^e = \frac{\text{(production & imports taxes: Table 3.2Q, Line 4)}_t^f + \text{(Sales taxes) s&l}_t}{\text{((Durables PCE) + c_t) \times (GDP deflator)_t} - \text{(production & imports taxes) s&l}_t} \]
\[ \tau_p^p = \frac{\text{(Proprietors’ income) + (wage income) + (wage supplements) + (capital income) }_t}{\text{(Personal current taxes) }_t} \]
\[ \tau_p^n = \frac{\tau_p^p \left( \frac{1}{2} \text{(Proprietors’ income) }_t + \text{(wage income) }_t + \text{(wage supplements) }_t + \text{(wage taxes) }_t \right)}{\text{(wage income) }_t + \text{(wage supplements) }_t + \text{(wage taxes) }_t + \text{(wage taxes) }_t + \frac{1}{2} \text{(Proprietors’ income) }_t} \]
\[ \tau_k^k = \frac{\tau_p^p \left( \text{(capital income) + (corporate taxes) }_t + \text{(corporate taxes) s&l}_t \right)}{\text{(Capital income) }_t + \text{(Property taxes) s&l}_t} \]

where the following NIPA sources were used:

- (Federal) production & imports taxes: Table 3.2Q, Line 4
- (State and local) sales taxes: Table 3.3Q, Line 7
- (Federal) personal current taxes: Table 3.2Q, Line 3
- (State and local) personal current taxes: Table 3.3Q, Line 3
- (Federal) taxes on corporate income minus profits of Federal Reserve banks: Table 3.2Q, Line 7 - Line 8.
- (State and local) taxes on corporate income: Table 3.3Q, Line 10.
- (Federal) wage tax (employer contributions for government social insurance): Table 1.12Q, Line 8.
- Proprietors’ income: Table 1.12Q, Line 9
- Wage income (wages and salaries): Table 1.12Q, Line 3.
• Wage supplements (employer contributions for employee pension and insurance): Table 1.12Q, Line 7.

• Capital income = sum of rental income of persons with CCAdj (Line 12), corporate profits (Line 13), net interest and miscellaneous payments (Line 18, all Table 1.12Q)

Note that the tax base for consumption taxes includes consumer durables, but to be consistent with the tax base in the model, the tax revenue is computed with the narrower tax base excluding consumer durables.

\[
(\text{rev})^c_t = \tau^c_t \times (c_t - (\text{Taxes on production and imports})^f_t - (\text{Sales taxes})^{s&k&l}_t - (\text{Population above 16})_t \times (\text{GDP deflator})_t
\]

\[
(\text{rev})^n_t = \tau^n_t \times ((\text{wage income})_t + (\text{wage supplements})_t + (\text{wage taxes})^f_t + \frac{1}{2}(\text{Proprietors’ income})_t)
\]

\[
(\text{rev})^k_t = \tau^k_t \times ((\text{Capital income})_t + (\text{Property taxes})^{s&k&l}_t)
\]

I construct government debt as the cumulative net borrowing of the consolidated NIPA government sector and adjust the level of debt to match the value of consolidated government FoF debt at par value in 1950:Q1. A minor complication arises as federal net purchases of nonproduced assets (NIPA Table 3.2Q, Line 43) is missing prior to 1959Q3. Since these purchases typically amount to less than 1% of federal government expenditures with a minimum of -1.1%, a maximum of 0.76%, and a median of 0.4% from 1959:Q3 to 1969:Q3, two alternative treatments of the missing data leads to virtually unchanged implications for government debt. First, I impute the data by imposing that the ratio of net purchases of nonproduced assets to the remaining federal expenditure is the same for all quarters from 1959:Q3 to 1969:Q4. Second, I treat the missing data as zero.

In 2012 the FoF data on long term municipal debt was revised up. The revision covers all quarters since 2004, but not before, implying a jump in the debt time series.\(^{38}\) I splice together a new smooth series from the data before and after 2004 by imposing that the growth of municipal debt from 2003:Q4 to 2004:Q1 was the same before and after the revision. This shifts up the municipal and consolidated debt levels prior to 2004. The revision in 2004 amounts to $840bn, or 6.8% of GDP.

The above data is combined with data from the web appendices of Romer and Romer (2004), Fernald (2012), Ramey (2011), and Mertens and Ravn (2013) on narrative shock measures. I standardize the different narrative shock measures to have unit standard deviation.

Figure 12 shows that there are significant differences in the total hours measure from Francis and Ramey (2009) and used in Ramey (2011), arising both from the

---

\(^{38}\)http://www.bondbuyer.com/issues/121_84/holders-municipal-debt-1039214-1.html

“Data Show Changes in Muni Buying Patterns” by Robert Slavin, 05/01/2012 (retrieved 01/24/2014).
Figure 12: Comparison of hours worked measures

hours time series itself (blue solid line vs. red dotted line) and from the difference between population measures (dashed vs. solid lines).

A.7 Additional Figures

<table>
<thead>
<tr>
<th></th>
<th>Core</th>
<th>Periphery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core + Periphery</td>
<td>99.86</td>
<td>99.76</td>
</tr>
<tr>
<td>Core only</td>
<td>99.81</td>
<td>99.58</td>
</tr>
</tbody>
</table>

Table 2: R-squared of core model vs full model
A.7.1 Monthly results

The following time series are used for the monthly analysis: Civilian government payroll (Fred code: USGOVT), industrial production (INDPRO), total private sector hours worked (AWHNONAG × USPRIV), money at zero maturity (MZMSL), the CPI (CPIAUCSL), and the Federal Funds Rate (FEDFUNDS). Commodity prices are measured using the PPI index for all commodities (PPIACO).

\[ G \text{ without } P_{\text{comm}}, \text{ lag} \quad P_{\text{CPI}} \text{ without } P_{\text{comm}}, \text{ lag} \quad P_{\text{comm}} \text{ without } P_{\text{comm}}, \text{ lag} \]

\[ G \text{ without } P_{\text{comm}} \quad P_{\text{CPI}} \text{ without } P_{\text{comm}} \quad P_{\text{comm}} \text{ without } P_{\text{comm}} \]

\[ G \text{ with } P_{\text{comm}} \quad P_{\text{CPI}} \text{ with } P_{\text{comm}} \quad P_{\text{comm}} \text{ with } P_{\text{comm}} \]

Note: Specification following Romer and Romer (2004). Controls include monthly dummies, and one to 24 lags of the dependent variable. Frequentist confidence bands of ± standard error based on the Delta method are shown. Specifications “with \(P_{\text{comm}}\)” also include lags of commodity prices.

Figure 13: Romer & Romer (2004) instrument: Univariate regressions
Note: Specification following Romer and Romer (2004). Controls include monthly dummies, and one to 24 lags of the dependent variable. Frequentist confidence bands of ± standard error based on the Delta method are shown. Specifications “with $P_{comm}$” also include lags of commodity prices.

Figure 14: Kuttner (2001) instrument: Univariate regressions
Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals. Specifications “with $P_{\text{comm}}$” also include commodity prices in the VAR.

Figure 15: Romer and Romer (2004) instrument: Monthly BVAR
Note: Shown are the pointwise median and 68% and 90% posterior confidence intervals. Specifications “with $P_{\text{comm}}$” also include commodity prices in the VAR.

Figure 16: Kuttner (2001) instrument: Monthly BVAR
A.7.2 Gibbs sampler

To calibrate the Gibbs sampler, I examine the autocorrelation functions and the split-sample distribution of univariate summary statistics of the model. In particular, I compare the log-likelihood across the first and second half of the parameter draws. If the distributions differ visibly, I increase the number of draws. Similarly, I compute the autocorrelation of the maximum eigenvalue of the stacked VAR(1) representation of (2.2) as well as of the Frobenius norm of $V$ and the log-likelihood. Figure 17 and Figure 18 in the Appendix show the corresponding plots. With a flat prior, discarding the first 20,000 draws and keeping every tenth draw with a total accept sample of 10,000 produces results consistent with convergence of the sampler: The two split-sample distributions line up closely and the autocorrelation is insignificant.

For computational purposes, I avoid computing the inverse in (3.14a) directly by using a QR-factorization. Since I only consider priors implemented using dummy observations and an otherwise flat prior: $N_0 = 0$ and $N_T(V) = N_{XX}(V)$. Then factor $\tilde{X} = Q_X R_X$, where $Q_X$ is an orthogonal matrix. It follows that $N_{XX}(V) = R'_X Q'_X Q_X R_X = R'_X R_X$ and therefore $N_{XX}(V)^{-1} = (R'_X R_X)^{-1} = (R_X)^{-1}(R'_X)^{-1}$, so that $R_X^{-1} \mathcal{N}((R_X^{-1})' N_{XY}(V), I)$ gives a draw from the posterior.

A summary measure of the multivariate posterior distribution is consistent with convergence of the Gibbs sampler: The plot shows the density estimate of $\frac{\ell(\theta^{(i)}) - \mu^G_1}{\sigma^G_1}$, where $\ell$ is the log-Likelihood of the data and $\theta^{(i)}$ a draw from the Gibbs sampler. $\sigma^G_1$ denotes the standard deviation of $\ell(\theta^{(i)})$ in the 1st half of the draws from the Gibbs sampler and $\mu^G_1$ the corresponding mean.

Figure 17: Gibbs sampler in baseline model: Normalized Distribution of Log-Likelihood
Note: Autocorrelations are reported based on both the Pearson and the Spearman correlation measure. Asymptotic classical 90% confidence intervals for the Pearson coefficient, computed under the assumption of zero correlation, are included around the horizontal axis.

Figure 18: Gibbs-Sampler of baseline model: Autocorrelation of univariate summary statistics